Molecular characterisation of the *EAS* gene cluster for ergot alkaloid biosynthesis in epichloë endophytes of grasses

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Molecular Genetics

at Massey University, Palmerston North,
New Zealand

Damien James Fleetwood

2007

Abstract

Clavicipitaceous fungal endophytes of the genera *Epichloë* and *Neotyphodium* form symbioses with grasses of the family Pooideae in which they can synthesise an array of bioprotective alkaloids. Some strains produce the ergot alkaloid ergovaline, which is implicated in livestock toxicoses caused by ingestion of endophyte-infected grasses.

Cloning and analysis of a plant-induced non-ribosomal peptide synthetase (NRPS) gene from *Neotyphodium lolii* and analysis of the *E. festucae* E2368 genome sequence revealed a complex gene cluster for ergot alkaloid biosynthesis. The *EAS* cluster contained a single-module NRPS gene, *lpsB*, and other genes orthologous to genes in the ergopeptine gene cluster of *Claviceps purpurea* and the clavine cluster of *Aspergillus fumigatus*. Functional analysis of *lpsB* confirmed its role in ergovaline synthesis and bioassays with the *lpsB* mutant unexpectedly suggested that ergovaline was not required for black beetle (*Heteronychus arator*) feeding deterrence from epichloë-infected grasses. Southern analysis showed the cluster was linked with previously identified ergot alkaloid biosynthetic genes, *dmaW* and *lpsA*, at a subtelomeric location. The ergovaline genes are closely associated with transposon relics, including retrotransposons, autonomous DNA transposons and miniature inverted-repeat transposable elements (MITEs), which are very rare in other fungi.

All genes in the cluster were highly expressed *in planta* but expression was very low or undetectable in mycelia from axenic culture, including under nitrogen-, carbon- or phosphate-limited conditions. Comparative analysis of the *EAS* gene cluster in four different epichloë strains showed marked differences in gene expression and ergot alkaloid synthesis. Gene order is conserved in each strain although evidence for recombination between two MITEs and expansion or reduction of a simple sequence repeat (SSR) at a single intergenic region was observed. Heterologous expression of a candidate regulatory gene, *laeA*, from *Aspergillus nidulans*, which is a global regulator of secondary metabolism in aspergilli, did not affect *eas* gene

expression. This, along with phylogeny and microsynteny analysis, suggests there is not an orthologue of this gene in epichloë.

This work provides a genetic foundation for elucidating biochemical steps in the ergovaline pathway, the ecological role of individual ergot alkaloid compounds, and the regulation of their synthesis *in planta*.

Acknowledgements

Like raising a child, so I am told, it takes a community to complete a PhD. Thank you so much to everyone that had input into this thesis, whether intellectual, experimental or emotional.

Thank you to my supervisors, Richard Johnson and Barry Scott. You made a great team and you were, in different ways, wonderful supervisors: rigorous, supportive and astute. Thank you both.

Thank you to everyone at AgResearch for a fun, stimulating environment. Special thanks to Tina Voisey for your support and many, many laughs. Thank you Greg Bryan for doing all the boss stuff I don't know much about but know to be important. Thanks Mike Christensen for imparting a measure of your immense knowledge and passion for epichloë endophyte biology. Thanks Wayne Simpson for showing me what to do with dirt and other very interesting stuff. Thanks Geoff Lane and Brian Tapper for several enjoyable discussions about chemistry and biology. Thanks Anar Khan for superlative bioinformatics support from afar. Thank you to members of the lab: Shalome, Linda, Charlotte, Steven, Kurtis, Lee, Jasmin, Jen, Anu and Melle, for a supportive, occasionally hilarious, environment. Thank you Helal Ansari, Michelle Thornton and Kim Richardson, for failed, but fun, attempted experiments. Thank you Sue, Julie, Nadia, Maree, Kathy, Dale and everyone else that keeps the world going around without names on papers.

Thanks to the members of ScottBase, Massey University, past and present, for being so inclusive. Special thanks to my science heroes Brendon Monahan and Carolyn Young for many stimulating discussions. Thank you Ruth Wrenn for running around after me from 2 km; and always with a smile.

Thank you to my parents for your support and encouragement from across the sea.

Several people need acknowledgement for technical and/or intellectual input into the data presented in this thesis. Aiko Tanaka provided the ps12 sequence that started the ball rolling. Chris Schardl, University of Kentucky, allowed access to the

pre-release of the *E. festucae* E2368 genome sequence, which added immensely to this work. Geoff Lane, Karl Fraser and Albert Koulman performed or interpreted LC-MS/MS analyses. Andrew Griffiths provided some of the genomic DNA samples and Jennifer Pratt performed some of the diagnostic PCRs described in Table 6.1. Anar Khan prepared the phylogenetic tree in Figure 4.5. Mike Christensen, Wayne Simpson and Anouck De Bonth provided biological material and took special care of my plants. Alison Popay provided expertise and materials for design and interpretation of the insect feeding choice test, Lisa Evans scored the feeding damage and John Koolard performed statistical analysis.

Finally, Lisa Paton. Above all else your love and support has kept me going. Thank you.

This PhD was truly a community effort and I have had a wonderful time taking part in it. To everyone that helped along the way, named and unnamed, family, friends, workmates, thank you.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

°C	degrees celcius
bp	base pair
c	centi
CDD	conserved domain database
dNTP	deoxynucleoside triphosphate
EFT	*Epichloë festucae* transposon
g	gram
g	gravitational force
h	hour
k	kilo
kb	kilobase
L	litre
LC-MS/MS	liquid chromatography tandem mass spectrometry
LTR	long terminal repeat
m	metre or milli
M	molar
min	minute
MITE	miniature inverted-repeat transposable element
n	nano
NRPS	non-ribosomal peptide synthetase
ORF	open reading frame
P	designates promoter
p	pico or designates plasmid
PKS	polyketide synthase
RIP	repeat induced point mutation
rpm	revolutions per minute
s	second
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
SSR	simple sequence repeat
T	designates terminator
UTR	untranslated region
TIR	terminal inverted repeat
TSD	target site duplication
V	volt
v/v	volume to volume
w/v	weight to volume
μ	micro

1. Introduction

1.1 Introduction

Cool season grasses of the sub-family pooideae, which include agriculturally important forages such as perennial ryegrass, are frequently naturally infected with fungal endophytes of the genera *Epichloë* and *Neotyphodium*. The interaction between epichloë endophytes (which signifies both *Epichloë* and *Neotyphodium* genera) and their temperate grass hosts is important for research, because of its value to agriculture and because the range of interactions, a spectrum from antagonistic to mutualistic, represents a unique opportunity to study the evolution of mutualism in fungi (Schardl *et al.*, 2004).

A focus of endophyte research has been the production of plant-protective secondary metabolite alkaloids. The non-ribosomal peptide ergot alkaloid ergovaline has received particular interest as it is implicated in a neurotoxicosis of stock grazing endophyte-infected pasture. Toxicoses associated with livestock grazing pasture grasses colonised by ergovaline-producing epichloë endophytes may have an economic cost of up to US$1 billion per annum in the USA (Panaccione *et al.*, 2001) and up to NZ$100 million in New Zealand (S. Easton, pers. com.). This thesis describes the cloning of a cluster of genes from *N. lolii* that are required for ergovaline biosynthesis and subsequent characterisation in the genetically tractable *E. festucae*.

Current knowledge of the endophyte symbiosis is described below, along with fungal secondary metabolism, with a particular emphasis on non-ribosomal peptide synthesis, and ergot alkaloid biosynthesis.

1.2 The Epichloë Endophyte/Grass Symbiosis

The relationship between temperate grasses and epichloë endophytes is ecologically beneficial to both partners. The fungi receive all of their nutrients from the host plant and the plant receives several ecological benefits as a result of endophyte infection. The fungi produce alkaloids that deter herbivory from both vertebrates and invertebrates - this is the most understood host protection mechanism and is reviewed in more detail below. Other host benefits include resistance to nematodes (Elmi *et al*., 1995), growth enhancements (Latch *et al*., 1985), and improved drought tolerance (Arachevaleta *et al*., 1989).

The epichloë endophyte symbioses can have both positive and negative effects for agriculture: pastures are protected from insect herbivory, however anti-mammalian compounds cause the toxicoses ryegrass staggers (Fletcher and Harvey, 1981) and tall fescue toxicosis (Lyons *et al*., 1986). Much research has gone into selecting endophyte strains that protect against insect herbivory but do not produce anti-mammalian effects (Tapper and Latch, 1999).

For general reviews of the symbiosis see: Schardl, (2001); Scott, (2001); and Schardl *et al*., (2004).

1.2.1 Endophyte Life Cycle

Epichloë endophytes exist as sexual *Epichloë* and asexual *Neotyphodium* species. Both sexual and asexual species spend most of their lifecycle in the apoplast (intercellular space) of the plant host in an asymptomatic relationship and nutrition is derived from the apoplastic fluid (Schardl, 1996). Haustoria, or feeding structures, are not formed although whether acquisition of nutrients from the host is passive or occurs through an active mechanism is unknown. Endophyte mycelia are found only in the above-ground parts of the plant, from the meristem to the leaf sheaths, blades and inflorescences (Fig. 1.1) (Christensen *et al*., 2002). Hyphae grow longitudinally, parallel to the leaf axis, and are seldom branched. Hyphal growth is rapid as the leaves are growing but halts when leaf growth ceases (Christensen *et al*., 2002). A recent finding showed that endophyte hyphae grow by normal

3
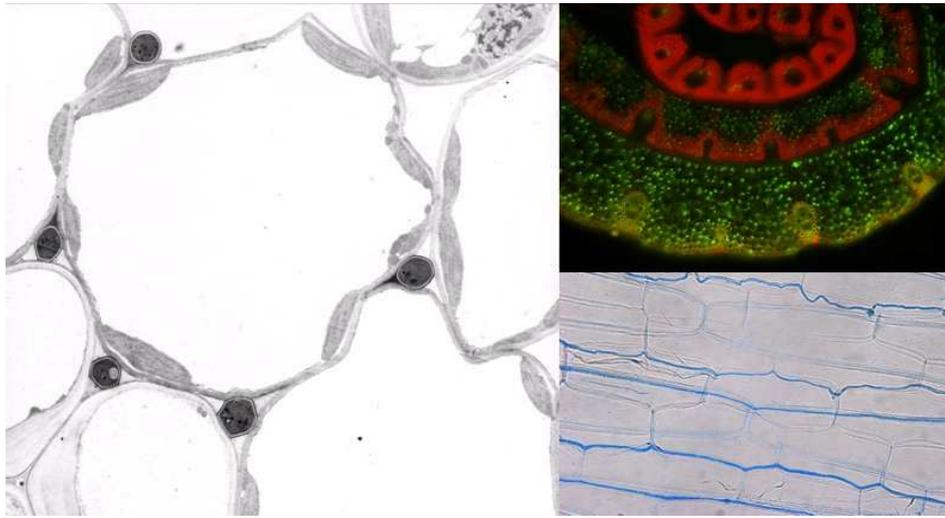
Figure 1.1. Epichloë endophyte growth *in planta*
A. Transmission electron micrograph of a transverse section of a perennial ryegrass leaf sheath showing *N. lolii* hyphae (dense cytoplasm) in aploplastic space between plant cells.
B. Fluorescence micrograph of a transverse section of a perennial ryegrass pseudostem infected with *E. festucae* Fl1 expressing green fluorescent protein (GFP). Red fluorescence is young leaf autofluorescence.
C. Light micrograph of aniline blue-stained longditudinal section of perennial ryegrass infected with *N. lolii* (blue).
Micrographs courtesy M. Christensen.

fungal tip growth in the meristematic zone but then grow by intercalary extension - cell division and extension behind the tip - in the leaf expansion zone, as host plant cells divide at the base rather than the tip (Christensen *et al.*, 2007). This challenges a defining characteristic of fungal growth - that vegetative growth occurs solely at the apical tip.

The sexual *Epichloë* species are capable of entering a pathogenic sexual stage and transmitting horizontally via ascospores (Fig. 1.2). During plant floral development, epiphytic growth of the fungus is triggered and a stroma is formed around the flag sheaf of the developing inflorescence, thereby sterilising the tiller (the unit structure of grass plants) and causing "choke" disease (Schardl, 2001). *Epichloë* species show a spectrum of pathogenicity; some species only transmit horizontally and are highly antagonistic, choking most or all tillers of the host plant. Others form pleiotropic

associations with fungal stromata forming on only some tillers and the others remaining asymptomatic. These pleiotropic associations are considered mutualistic as the benefits of endophyte infection outweigh the cost of losing some reproductive ability (Schardl and Clay, 1997). Molecular insight into an aspect of mutualism was recently gained when it was shown that blockage of endophyte production of reactive oxygen species (ROS) caused a "switch" to antagonism with overgrowth of fungal hyphae, resulting in stunting of the host plant (Takemoto *et al.*, 2006; Tanaka *et al.*, 2006).

The *Epichloë* mating system is heterothallic and requires mediation by an anthomyiid fly of the genus *Botanophila*, a third member in the symbiosis (Bultman *et al.*, 1995). Female flies are attracted to *Epichloe* stromata by chokol K, a volatile antimicrobial compound synthesised by the fungus (Schiestl *et al.*, 2006), and can transfer spermatia, in the form of conidia, from one stroma to another while laying eggs on the stroma surfaces. Ascospores are formed in perithecia following successful cross-fertilisation of conidia of opposite mating type and these are ejected into the air and initiate new infections on neighbouring plants. The mechanism of grass host colonisation via ascospore dispersal may be similar to that of *Claviceps purpurea*, where ascospores germinate to produce conidia and hyphae from germinated conidia invade the ovule (Chung and Schardl 1997).
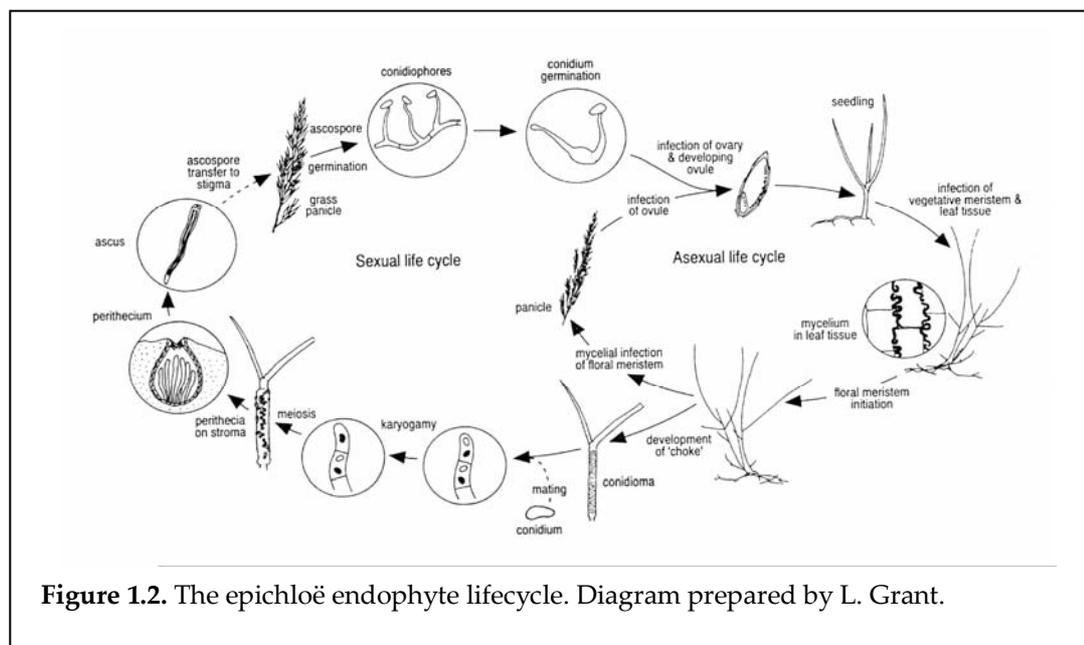
Figure 1.2. The epichloë endophyte lifecycle. Diagram prepared by L. Grant.

In the vertical mode of transmission undergone by both *Epichloë* and asexual *Neotyphodium* species, fungal hyphae grow intercellularly in the floral meristems and in the ovules of the florets (Fig. 1.2) (Philipson and Christey, 1986; White *et al.*, 1991). The seed is then infected and the fungus is transmitted vertically to the next generation of host plant. Vegetative propagation also occurs when developing tillers are infected. Asexual *Neotyphodium* species are generally considered wholly mutualist as no epiphytic growth occurs and plant inflorescences develop normally.

1.2.2 Endophyte Taxonomy

Endophyte taxonomy is complicated by the inability to characterise anamorphs as a species using a traditional biological species concept (Mayr, 1999). Thus the sexual *Epichloë* species are considered separately to the closely related *Neotyphodium* species.

Epichloë species belong to the family Clavicipitaceae, order Hypocreales, phylum Ascomycota (Kuldau *et al.*, 1997). Other members of this family include *Claviceps, Balansia, Cordiceps* and *Metarhizium*. There are ten known species of *Epichloë,* which separate into two clades based on *tub*2 and *tef*1 gene phylogeny (Clay and Schardl, 2002). The most diverse clade, designated the *E. typhina* complex, contains three phlogenetically interreticulated species, *E. typhina, E. clarkii* and *E. sylvatica* (Clay and Schardl, 2002). These species have a very broad host-range, and usually transmit horizontally in antagonistic relationships. The second, larger clade, designated the main clade, contains seven described species: *E. festucae, E. amarillans, E. baconii, E. bromicola, E. elymi, E. glyceriae* and *E. brachyelytri* (Clay and Schardl, 2002). These species belong to individual subclades and are each specific for a single host plant tribe. Cocladogenesis is thought to have occurred with these species and their host plant tribes. Main clade species usually form pleiotropic relationships with host plants, transmitting both horizontally and vertically. *E. festucae,* however, usually forms mutualistic symbioses and very seldom forms stromata (Leuchtmann *et al.*, 1994).

The asexual *Neotyphodium* species are obligate mutualists, incapable of stromata formation. Phylogenetic analyses place these species into 14 named species and several unnamed species placed into taxonomic groupings named after the host plant they were first identified from, for example, FaTG-2 for *Festuca arundinacea* taxonomic grouping number 2 (Christensen *et al.*, 1993). Most of the *Neotyphodium* species are thought to have arisen by hybridisation of two or more species of *Epichloë* (Tsai *et al.*, 1994; Moon *et al.*, 2000; Moon *et al.*, 2004). A study by C. Moon *et al* (2004) identified 32 distinct genotypes of asexual endophytes by *tub*2 and *tef*1 phylogenies and microsatellite length polymorphisms. Twenty of these had apparently arisen by hybridisation, evident by multiple alleles of *tub*2 and *tef*1 in these species. Sexual species tested contained only one copy of each gene. It was possible to infer from gene similarity the sexual ancestors for the asexual species with most being hybrids of two ancestors, for example *N. uncinatum* was shown to be a hybrid of *E. bromicola* and an *E. typhina* complex endophyte. A few species were shown to have arisen from two separate hybridisation events. *N. coenophialum* for example has *E. festucae*, *E. typhina* complex and *E. baconii* ancestry. The remaining twelve species were thought to have simply arisen from the loss of sexuality, *N. lolii* for example being an asexual descendant of *E. festucae*. Hybridisation in endophytes is proposed to arise from anastomosis (hyphal fusion) of two species in an infected tiller, probably following a loss of sexuality. However, anastomosis or sexual mating of two sexual species cannot be ruled out as the means of hybridisation in all cases (Schardl and Craven, 2003; Moon *et al.*, 2004).

1.2.3 Epichloë Bioprotective Alkaloids

The most agriculturally beneficial, and harmful, aspects of the grass-endophyte symbiosis – the conferment of resistance to herbivory and toxicity to grazing stock – are caused by the production of fungal alkaloids. There are four main classes of host-protective endophyte alkaloids: pyrrolopyrazines, 1-aminopyrrolizidines, indole diterpenes and ergot alkaloids. Different species and strains of endophyte produce different combinations of these alkaloids. Endophyte alkaloids are reviewed in Bush *et al.*, (1997) and Scott, (2003a).

1.2.3.1 1-Aminopyrrolizidines

Alkaloids of this class are commonly called lolines. These alkaloids have potent insecticidal and feeding-deterrent properties and can accumulate up to 2% of the dry weight of the infected plant's dry mass, far exceeding the biomass of the fungus and the amounts of other alkaloids (Spiering *et al.*, 2002). Lolines are produced by *E. festucae* and several *Neotyphodium* species (Clay and Schardl, 2002).

Until recently, it was not conclusively known if lolines were produced by the plant, the endophyte or both. Blankenship *et al* (2001) showed that axenic cultures of *N. uncinatum*, a common meadow fescue endophyte, can produce lolines in minimal media, demonstrating that these products can be solely synthesised by the fungus. Other strains of endophyte that produce lolines *in planta* however could not be induced to produce lolines in culture (Blankenship *et al.*, 2001).

Mendelian analysis of genetic crosses of Loline[+] and Loline[-] *E. festucae* showed that loline synthesis segregated as a single genetic locus, designated *LOL,* which is linked to two AFLP markers (Wilkinson *et al.*, 2000). Genes expected to be involved in loline synthesis have since been identified in *N. uncinatum* using suppression subtractive hybridisation (Spiering *et al.*, 2002) and genome walking (Spiering *et al.*, 2005). The *lol* genes are organised as two almost identical (~93% gene nucleotide identity) plant-regulated gene clusters of nine genes each in the *N. uncinatum* genome, the taxonomic distribution of which strictly correlated with loline production. The role of *lolC*, a putative pyridoxal phosphate-containing enzyme, was confirmed by RNA interference (RNAi) (Spiering *et al.*, 2005).

1.2.3.2 Pyrrolopyrazines

The only known pyrrolopyrazine synthesised by endophytes is peramine, an alkaloid synthesised by most species of *Epichloe* and *Neotyphodium*. Peramine is a feeding deterrent to Argentine Stem Weevil (*Listronotus bonariensis*), which would otherwise devastate New Zealand pastures (Rowan, 1993). Peramine is proposed to be synthesised from arginine and either proline (Rowan, 1993) or 1-pyrroline-5-carboxylate (Tanaka *et al.*, 2005). A non-ribosomal peptide synthetase (NRPS) for

peramine synthesis, *perA*, was recently cloned from *N. lolii* using degenerate primers to conserved NRPS domains and may be the only gene required for peramine synthesis (Tanaka *et al.*, 2005). Deletion of this gene blocked peramine production, leaving the symbiotum more susceptible to herbivory by Argentine Stem Weevil. This study provided rare experimental evidence of an adaptive advantage for production of a secondary metabolite by a fungus.

1.2.3.3 Indole Diterpenes

Indole diterpenes are a large group of fungal terpenes with diverse structures, synthesised from an indole moiety and geranylgeranyl diphosphate. Many of the indole diterpenes are mammalian tremorgens (Scott *et al.*, 2004). Lolitrem B, the most studied endophyte indole diterpene, is responsible for the syndrome ryegrass staggers, first described as a result of livestock grazing on *N. lolii*-infected perennial ryegrass in New Zealand by L. Fletcher and I. Harvey in 1981 (Fletcher and Harvey, 1981).

A cluster of genes necessary for paxilline biosynthesis, a related indole diterpene, has been characterised from *Penicillium paxilli* (Young *et al.*, 1998; Young *et al.*, 2001; McMillan *et al.*, 2003). A geranylgeranyl diphosphate synthase gene from this cluster, *pax*G, was used as a probe to find the equivalent gene in *N. lolii* (Young *et al*, 2005). Chromosome walking, combined with the use of ESTs isolated from an SSH library, identified further clustered *pax* gene homologues proposed to be required for epichloë/neotyphodium indole diterpene synthesis (Young *et al.*, 2005; Young *et al.*, 2006). The structure of the *LTM* cluster is interesting in that the *ltm* genes are arranged in three "platforms", interrupted by nested arrays of two degenerate retrotransposons, Tahi and Rua (Young *et al.* 2005, Young 2006), rather than a single contiguous cluster, or multiple distinct clusters as is common for other secondary metabolite gene clusters (Keller and Hohn, 1997). RT-PCR analysis of the *ltm* genes showed that these genes are highly upregulated *in planta* compared to in axenic culture and potential *cis* regulatory elements were identified by sequence comparison between the *LTM* upstream regions (Young *et al.*, 2006) however

whether these sequences are functionally important remains to be experimentally tested.

1.2.3.4 Ergot Alkaloids

These clavine, lysergyl amide and ergopeptine compounds are associated with livestock toxicity in endophyte-infected pastures. Ergot alkaloids are reviewed in Section 1.5.

1.3 Fungal Secondary Metabolism – Natural Products

Secondary metabolism is distinct from primary metabolism, and is classically defined, by being non-essential for growth under nutrient-rich culture conditions. Under this definition, catabolic pathways that use, for example, an alternative or unusual carbon source are called secondary metabolism, however, for the purposes of this review, secondary metabolism describes the anabolic formation of low molecular weight, often bioactive families of metabolites. Functional properties of secondary metabolites are often described based on their effects on human interests and include antibiotics, immunosuppressants and toxins, however the biological function of many, particularly in the "natural" setting, is unknown (Keller *et al*., 2005). Different secondary metabolites are produced by discontinuous taxonomic groups of fungi, plants and bacteria, but not animals, and are typically synthesised, at least in fungi and bacteria, during a particular growth or developmental phase.

Classical definitions of fungal secondary metabolism are challenged by advances in large-scale functional genomics, as more and more secondary metabolite pathways, and some biological functions, are elucidated. The definition of "non-essential" becomes somewhat ambiguous outside the axenic culture context. For example, siderophores, small peptides produced by fungi and bacteria to scavenge iron, are synthesised via non-ribosomal peptide synthetases – typical secondary metabolism enzymes – but are essential for growth and pathogenesis (Haas, 2003; Oide *et al*., 2006). Can they then be considered secondary metabolites? Some secondary metabolites are also produced throughout the producer's lifecycle, rather than the defined developmental phase classically described for secondary metabolism. For example, the plant-protective alkaloids produced by the asexual epichloë endophytes (described in the previous section) are constitutively produced *in planta* (Spiering *et al*., 2005; Tanaka *et al*., 2005; Young *et al*., 2005), the only known habitat for these species in nature. The term "natural product" is thus more frequently

being used to describe fungal products hitherto known as secondary metabolites and the two terms are used interchangeably in this thesis.

1.3.1 Classes of Fungal Natural Products

Fungal secondary metabolites generally fall into three classes: polyketides, terpenes and non-ribosomal peptides.

1.3.1.1 Polyketides

The polyketides are a large and diverse family, the most well studied being aflatoxin/sterigmatocystin (Yu *et al*., 1995; Brown *et al*., 1996; Brown *et al*., 1999), lovastatin/compactin (Kennedy *et al*., 1999; Abe *et al*., 2002), T-toxin (Kodama *et al*., 1999) and the fumonisins (Proctor *et al*., 1999; Seo *et al*., 2001; Proctor *et al*., 2003). They are synthesised from a starter unit of acetyl coenzyme A (CoA) and extender units, usually of malanoyl CoA, leading to a chain of differing lengths. Fungal polyketides are synthesised by multidomain type I polyketide synthase (PKS) enzymes similar to eukaryotic fatty acid synthases, containing a minimal set of three domains required for condensation of starter and extender units, the ketoacyl CoA synthase (KS), acyltransferase (AT), and acyl carrier protein (ACP). Fungal PKSs may also contain different combinations of domains that catalyse reduction of the β-keto group formed on extender group condensation, ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains. All known fungal PKS enzymes are type I, meaning they contain all the domains for at least one round of malanoyl CoA condensation, as opposed to type II PKSs, which contain domains on separate proteins. Fungal PKSs are also iterative – a single module catalyses the addition of repeat units in the polyketide chain. In bacterial systems, multiple PKSs often function as a complex to add different repeat units. Polyketides are usually released from the PKS enzyme via cleavage mediated by a thioesterase (TE) domain at the carboxy-terminal region.

Diversity in polyketide structures results from modifications catalysed by the PKSs and also a host of enzymes that modify the polyketide after release from the PKS. Modifications mediated by the PKS include the number and type of extender units incorporated, the extent of reductive processing of β-keto groups, and cyclisations.

Post-PKS modifying enzymes commonly include cytochrome P450 monooxygenases, oxidoreductases, esterases and methyltransferases. Some polyketide natural products are subject to a great deal of modification subsequent to polyketide synthesis, the aflatoxin pathway for example is proposed to include at least 17 enzymes (Brown *et al.*, 1999). Polyketide synthesis is reviewed in: Rawlings, (1997); Tkacz, (2000); Staunton and Weissman, (2001); Walsh, (2004); and Fischbach and Walsh, (2006)

Polyketides are the most abundant fungal natural products and analysis of fungal genomes suggest ~7 – 25 PKSs per ascomycete genome (Kroken *et al.*, 2003). There are no studied polyketides produced by epichloë, however, analysis in *N. lolii* by degenerate PCR revealed at least 10 PKS genes and work is underway to determine the product and possible function of several of these (C. Voisey, pers. com.).

1.3.1.2 Terpenes

Terpenes are derived from isoprenoid precursors from primary metabolism, which are required for sterol synthesis and prenylation of proteins. Well-studied examples of fungal terpenes include gibberellins (Tudzynski, 2005), indole diterpenes (Young *et al.*, 2001), and tricothecenes (Brown *et al.*, 2003; Brown *et al.*, 2004). Terpenes are subclassified as monoterpenes, which are derived from geranyl diphosphate, sesquiterpenes, derived from farnesyl diphosphate and diterpenes, generated from geranylgeranyl diphosphate. Chain elongation of the isoprenoid precursors is catalysed by a family of isoprenyl diphosphate synthases. Two genes for the geranylgeranyl diphosphate synthase required for diterpene synthesis appear to be found in diterpene-producing fungi, one for primary metabolism and one for diterpene formation (Tudzynski and Holter, 1998; Young *et al.*, 2001; Young *et al.*, 2005). This may be a signature of diterpene-producing organisms (Young *et al.*, 2005). Following formation of isoprenoids, terpenes are formed by cyclisation mediated by terpene cyclises. Considerable diversity among mono-, sesqui- and diterpenes is then caused by a variety of modifying enzymes including cytochrome P450 monooxygenases, prenyl transferases and oxidoreductases. Fungal terpene synthesis is reviewed in Scott *et al.*, (2004).

1.3.1.3 Non-Ribosomal Peptides

Non-ribosomal peptides are a structurally diverse group of peptide natural products formed by extremely large enzymes called non-ribosomal peptide synthetases (NRPSs). Some well studied examples of fungal non-ribosomal peptides are penicillin/cephalosporin (Martin, 1998; Brakhage *et al.*, 2005), HC- and AM-toxins (Johnson *et al.*, 2000; Walton, 2006), enniatin (Haese *et al.*, 1993) and peptaibols (Wiest *et al.*, 2002). Non-ribosomal peptide synthesis is described in detail in Section 1.4.

1.3.2 Biosynthetic Gene Clusters

Genes encoding enzymes required for fungal natural product biosynthesis are usually found tightly linked in contiguous clusters in the genome. This was not originally expected as all eukaryotic genes were thought to be dispersed in the genome, unlike prokaryotes where genes for a particular pathway are often found clustered as operons transcribed on a polycistronic mRNA (Keller and Hohn, 1997). Genes found in fungal gene clusters each have their own promoters and are individually transcribed, however, a feature of these clusters is the co-ordinate expression of each of the genes (Keller and Hohn, 1997; Scott, 2003b). (Regulation of natural product genes is described in Section 1.3.4.) An interesting finding from fungal genome sequencing projects is that many gene clusters appear to be found near to telomeres (Galagan *et al.*, 2005; Nierman *et al.*, 2005; Rehmeyer *et al.*, 2006). This may be important in gene regulation and evolution, although a role in either has not been experimentally determined. A list of biosynthetic gene clusters characterised include those for aflatoxin (Brown *et al.*, 1999), dothistromin (Bradshaw *et al.*, 2002), fumomisin (Proctor *et al.*, 2003), gibberellin (Tudzynski and Holter, 1998), HC toxin (Ahn *et al.*, 2002), paxilline (Young *et al.*, 2001), penicillin (Fierro *et al.*, 1995), sterigmatocystin (Brown *et al.*, 1996) and trichothecenes (Brown *et al.*, 2004), although many more have been discovered, particularly since several fungal genomes have now been sequenced. It should be noted that genes for nutrient utilisation are also often found as gene clusters, indeed, the first fungal genes found to be part of a cluster were those for proline utilisation (Hull *et al.*, 1989).

Recently, clustered genes for a secondary metabolite in oat were described (Qi *et al.*, 2004) and analyses with whole genome sequences have also shown a much larger amount of clustering of genes with non-secondary metabolic roles in eukaryotes than was previously thought (Cohen *et al.*, 2000; Lercher *et al.*, 2002; Hurst *et al.*, 2004). However, while the genes in these studies are clustered with statistical significance, the number of genes can be great, the distance between genes large (possibly with several genes in between the "clustered" genes) and there is ambiguity between different clusters; several stress-induced genes in the *S. cerevisiae* genome, for example, are found in stress-repressed "clusters" and vice versa (Burhans *et al.*, 2006). There is therefore a clear distinction between this generalised clustering of genes into very broad transcriptional blocks in genomes and tightly linked gene clusters for natural product biosynthesis and nutrient utilisation in fungi.

1.3.3 Gene Cluster Evolution

Of great interest is the question of how, and why, fungal gene clusters evolved. Two main theories exist, the first, that gene clusters arose from horizontal transfer of bacterial operons, the second, that gene clusters evolved due to a selectable advantage for genes to be clustered, possibly in order to facilitate co-regulation.

Horizontal transfer is a common mechanism in bacteria, where conjugation or transduction results in the transfer of plasmids or integrative conjugative elements resulting in new, potentially beneficial traits in the recipient cell. Mechanisms for horizontal transfer in fungi also exist, relying on anastomosis (hyphal fusion). Parasexual recombination following formation of heterokaryons and fusion of nuclei is a possible mechanism for incorporation of "new" DNA into a fungal genome (Schoustra *et al.*, 2007). Transfer of a small dispensable chromosome has also been shown in *Colletotrichum gloeosporioides* (He *et al.*, 1998). However, while horizontal transfer of bacterial secondary metabolism genes between fungi and bacteria can be hypothesised to explain the clustering of genes, experimental evidence for inter-kingdom horizontal transfer is generally lacking.

Genes for penicillin and cephalosporin, closely related β-lactam compounds, have frequently been hypothesised to have been horizontally transferred from bacteria to fungi (Carr *et al.*, 1986; Landan *et al.*, 1990; Penalva *et al.*, 1990; Aharonowitz *et al.*, 1992; Brakhage *et al.*, 2005). Two of the three genes for penicillin synthesis and all cephalosporin genes are found in both bacteria and fungi (Brakhage *et al.*, 2005). Aside from the genes being clustered in both kingdoms, additional evidence for transfer is proposed based on the sequence of the genes responsible for the earliest two steps, *acvA* (*pcbAB*) and *ipnA* (*pcbC*), which respectively catalyse the formation of δ(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV) and isopenicillin N (IPN), the shared intermediates in penicillin and cephalosporin synthesis. Neither of these genes have introns and the *ipnA* genes of fungi and bacteria share 60% DNA sequence identity (Brakhage *et al.*, 2005). The *ipnA* genes of *Aspergillus nidulans* and *Penicillium chrysogenum* are also extremely GC rich at the third codon position, more typical of streptomycetes than fungi (Aharonowitz *et al.*, 1992). While this circumstantial evidence appears convincing, there are few known fungal gene clusters with a homologous pathway identified in bacteria and most genes in secondary metabolism contain typical fungal introns. Thus, even should transfer of pathways from bacteria to fungi have occurred, it seems to have been a very rare event and other processes are likely responsible for the majority of biosynthetic gene clustering.

If gene clusters were not transferred "ready made" from bacteria, what selective pressures may exist for genes to become clustered? Walton, (2000) hypothesised, based on the "selfish operon" theory for bacteria of Lawrence and Roth, (1996), that secondary metabolism genes become clustered because of a selective advantage to the cluster itself, rather than the whole organism. As fungal genomes are relatively unstable and unessential genes are under weak selection pressure compared with housekeeping genes, having genes for secondary metabolism clustered allows these to be transferred into other organisms together and the trait they encode then conferred on that organism. The genes are then able to confer a potential selectable advantage to retaining the gene cluster. Single secondary metabolic genes

transferred to a different organism would not confer any selectable trait and would therefore be rapidly selected against. This "selfish cluster" hypothesis relies on two contestable theories: that evolutionary selection pressure is exerted at the level of the gene – the "selfish gene" theory proposed by Dawkins, (1976), and that horizontal transfer among fungi is common. As mentioned above, mechanisms exist for horizontal transfer, however, few events have been conclusively demonstrated and a comprehensive phylogenetic analysis of PKS genes in fungi suggested that the discontinuous distribution of most of these genes was likely to be due to gene duplication, divergence and gene loss, rather than horizontal transfer (Kroken *et al.*, 2003).

Another hypothesis to explain the clustering of secondary metabolite genes is that clustering optimises co-regulation of the genes. While clustering would not seem to be important for regulation by *trans* acting transcription factors, clustering may facilitate regulation at the level of chromatin (Keller and Hohn, 1997; Keller *et al.*, 2005). The importance of this level of regulation for clustering is supported by the generalised clustering phenomenon described above where many genes, including housekeeping genes, are loosely clustered within co-transcribed blocks of chromatin in eukaryotic genomes. The recent finding that natural product gene clusters in aspergillus are regulated by LaeA, a histone methyl transferase (Section 1.3.4) (Bok and Keller, 2004; Bok *et al.*, 2006b; Bok *et al.*, 2006a; Perrin *et al.*, 2007), adds further weight to this hypothesis, as does a significant recent bioinformatics study, which provided the first report of the "birth" of a metabolic gene cluster.

Wong and Wolfe, (2005) compared the genome sequences of ten different yeast species and showed that the *Saccharomyces cerevisiae* and *S. castelli DAL* cluster, which contains six of the eight genes for allantoin utilisation as a nitrogen source, was assembled relatively recently in evolutionary time, by relocation of previously-scattered genes to a subtelomeric location. These genes were relocated rapidly following the loss of the ability to convert urate to allantoin in the purine catabolism pathway and in urate-utilising yeast species the *DAL* genes remain scattered. Two selective features are proposed to have led to the formation of the

gene cluster: epistatic selection, whereby the fitness of one allele relies on another and therefore the two alleles are under pressure to remain linked, and the relocation of the genes to a single subtelomeric region that is subject to chromatin-mediated regulation in response to nitrogen starvation. This study provides strong support for the hypothesis that gene clusters evolve due to a selectable advantage to genes being clustered.

With several filamentous fungal genomes now sequenced and many more imminent, comparative analysis may begin to answer the question of how gene clusters for natural product biosynthesis evolved and the importance of horizontal transfer in fungal evolution.

1.3.4 Regulation of Secondary Metabolite Genes

Regulation of secondary metabolite genes can be complex. Gene expression can be mediated by both pathway-specific and broad-target transcription factors, regulation of several natural product gene clusters is linked with fungal developmental pathways and epigenetic regulation was also recently found to be important. Regulation of fungal secondary metabolism is reviewed in: Keller *et al.*, (2005); Yu and Keller, (2005); and Hoffmeister and Keller, (2007).

1.3.4.1 Narrow-Range Transcription Factors

Many fungal gene clusters contain pathway-specific transcription factors that positively regulate the transcription of other genes in the cluster. These proteins can be Cys_2His_2 zinc-finger proteins, such as Tri6 in tricothecene production (Proctor *et al.*, 1995; Hohn *et al.*, 1999), ankyrin repeat proteins, such as ToxE in HC-toxin production (Ahn and Walton, 1998), and, most commonly, the fungi-specific $Zn(II)_2Cys_6$ zinc binuclear cluster proteins, such as AflR in sterigmatocystin/aflatoxin production (Yu *et al.*, 1996). For the pathway-specific regulators AflR, Tri6 and ToxE, *cis* binding sites in the promoters of genes in the respective clusters have been identified (Fernandes *et al.*, 1998; Hohn *et al.*, 1999; Pedley and Walton, 2001). While many gene clusters contain pathway-specific regulators, others, such as the gibberellin and fumonisin gene clusters (Mihlan *et al.*, 2003; Proctor *et al.*, 2003), do not. In these cases the pathway-specific regulatory

gene may be found outside the cluster or gene expression may be mediated by non-specific transcription factors.

1.3.4.2 Broad-Range Transcription Factors

While narrow-range regulators enable natural product genes to be expressed in response to specific environmental inducers, secondary metabolism can also be regulated in response to general cellular metabolic cues, including carbon and nitrogen levels and pH. The broad-range transcription factors AreA (nitrogen) (Hynes, 1975), CreA (carbon) (Dowzer and Kelly, 1991) and PacC (pH) (Tilburn *et al.*, 1995) positively and negatively mediate the expression of genes in response to these cues and are conserved among eukaryotes. Examples of pathways known to be regulated by these proteins include cephalosporin/penicillin, genes for which are regulated by all three proteins (Haas and Marzluf, 1995; Tilburn *et al.*, 1995; Jekosch and Kuck, 2000) and gibberellin, which is regulated by AreA (Mihlan *et al.*, 2003). A recent study by Teichert *et al* (2006) showed that AreA regulation of gibberellin and bikaverin in *Fusarium fujikuroi* was moderated by phosphorylation from the phosphoinositol kinase Target of Rapamycin (TorA) (Teichert *et al.*, 2006). TorA is a conserved protein that integrates signals for nutrient presence and mediates cell growth via ribosome biogenesis, transcriptional and translational regulation (Cutler *et al.*, 1999). This was the first study to show that secondary metabolism can also be under the control of TorA signalling.

1.3.4.3 Secondary Metabolism and Development

The expression of some natural product genes is linked with fungal development, particularly conidiation (reviewed in Calvo *et al.*, (2002)). This phenomenon is best studied in aspergillus although mutants in other species in genes known to be important in conidiation, such as heterotrimeric G-protein genes in *Botrytis cinerea* (Gronover *et al.*, 2004), *Trichoderma atroviridae* (Reithner *et al.*, 2005) and *Fusarium sporotrichiodes* (Tag *et al.*, 2000), are also affected in secondary metabolite production. The best-studied fungal developmental signal transduction pathway affecting secondary metabolism is the heterotrimeric G-protein network in *A. nidulans*. Mutants deleted in either FadA or SfaD, the α and β subunits of the G-

protein complex respectively, have either increased or premature conidiation and early sterigmatocystin production (Rosen *et al.*, 1999; Tag *et al.*, 2000). The same phenotype is observed for a strain deleted in protein kinase A (PkaA), the effector protein of the downstream cyclic AMP (cAMP) signalling network (Shimizu and Keller, 2001). The G-protein/cAMP signal transduction cascade is thus a negative regulator of both conidiation and sterigmatocystin production. This occurs via PkaA-mediated phosphorylation of AflR, the sterigmatocystin transcription factor and BrlA, a transcription factor required for expression of conidiation genes (Shimizu and Keller, 2001). PkaA also mediates transcriptional repression of *aflR* via the LaeA protein (Bok and Keller, 2004), this mode of regulation is described below.

1.3.4.4 LaeA and Epigenetic Regulation

One of the most significant recent findings with regard to the regulation of natural product genes was the identification of the *laeA* gene in *A. nidulans*. Deletion of *laeA* specifically abolishes the expression of all secondary metabolite genes examined, while overexpression of *laeA* increases expression of the same genes (Bok and Keller, 2004). Detailed examination of the sterigmatocystin cluster showed that LaeA affects expression of only the genes within the cluster; bordering genes were not affected. LaeA localises to the nucleus but does not appear to be a DNA-binding protein, rather it appears to be a histone methyltransferase and is thus likely to exert control over its target genes via chromatin remodelling. That chromatin context is important for regulation is supported by the finding that placement of the primary metabolic gene *argB* within the sterigmatocystin cluster resulted in its silencing in the *laeA* deletion mutant background, while moving the *aflR* regulatory gene outside the cluster in this background resulted in remediation of cluster gene expression (Bok *et al.*, 2006a). Given this evidence, it seems likely that secondary metabolic gene clusters are usually found in repressing heterochromatin (condensed chromatin structure) environments and that LaeA mediates chromatin remodelling via histone methylation, although the histone and amino acid residue remain to be determined (Keller *et al.*, 2005).

Although it has not yet been shown whether *laeA* orthologues exist in other fungi, it seems likely that a similar mechanism at least would be ubiquitous. This provides the strongest clue so far as to why natural product genes exist as clusters, that is, to enable each gene to be regulated via chromatin context.

1.4 Non-Ribosomal Peptide Synthetases

Non-ribosomal peptide synthetases (NRPSs) are enzymes found in bacteria and fungi that catalyse the formation of small, secondary metabolite peptides. These peptides have highly diverse structures and activities, often being very bioactive, some examples being vancomycin, cyclosporin, siderophores and the known endophyte non-ribsosomal peptides ergovaline and peramine. The usual size of non-ribosomal peptides is between 2 and 15 amino acids although some are larger (Wiest *et al.*, 2002). The upper size of the non-ribosomal peptides is possibly imposed by the very large size of the NRPS enzymes (Mootz *et al.*, 2002). The structures of non-ribosomal peptides are extremely diverse: products can be cyclic, linear or branched and often incorporate non-proteinogenic amino acids, as well as α-hydroxy and carboxylic acids. Peptide products can be further modified by the NRPSs – amino acids may be *N*-methylated, acylated and glycosylated, stereoisomerisation can be performed and heterocylic ring formation can occur. The huge structural diversity confers the potential for these products to bind to many biological targets. (Mootz *et al.*, 2002).

Despite the heterogeneity of the products, the mode of non-ribosomal peptide synthesis is highly conserved. NRPSs are large modular enzymes where each module in the enzyme is semi-autonomous, with the ability to activate a specific substrate, tether it covalently, and condense a peptide bond with the substrate bound to a downstream module, resulting in a product with defined sequence. This mode of synthesis has been called the multiple carrier thiotemplate mechanism (Stein *et al.*, 1996; Marahiel *et al.*, 1997). In bacterial NRPS systems, separate polypeptides usually encode one or more modules that are components of larger NRPS complexes, while in fungal systems modules are found on single, long polypeptides. The one known exception to this is in the *Claviceps purpurea* ergopeptine NRPS, comprised of two enzymes, LpsB and LpsA, with one and three modules respectively (Riederer *et al.*, 1996); a homologous system is also examined in this thesis.

Non-ribosomal peptide synthetases are comprehensively reviewed in several papers, a selection include: Marahiel *et al.*, (1997); von Döhren *et al.*, (1997); Mootz *et al.*, (2002); and Finking and Marahiel, (2004).

1.4.1 Domain Structure of NRPSs

Other than initiation modules (Fig. 1.3), each NRPS module consists of, at least, an adenylation domain (A domain), thiolation domain (T domain) and a condensation domain (C domain) (Marahiel *et al.*, 1997). These three domains constitute the fewest domains required for activation, covalent tethering and condensation of a substrate with a growing upstream peptide chain. The final module of a NRPS must also contain a release domain that allows the covalently bound non-ribosomal peptide to be untethered from the enzyme. Domains that modify substrates are also found in NRPSs, including *N*-methylation, epimerisation and reductase domains. These modifications greatly increase the diversity of non-ribosomal peptide products. The relationship of NRPS domains, modules and genes is shown in Figure 1.3. The adenylation, thiolation, condensation and release domains are discussed in more detail below.

1.4.1.1 Adenylation Domains

The A domain is responsible for the recognition and chemical activation of a specific substrate, hence these domains are ultimately responsible for the structure of the final NRP product and have been the focus of investigation into NRPS structure. A domains recognise their substrate by a specific binding pocket where the substrate is non-covalently bound (Conti *et al.*, 1997). There, the substrate is activated as an aminoacyl-adenylate – as in ribosomal synthesis – in the presence of

Figure 1.3. Relationship of gene, protein modules and domains. NRPS gene products are subdivided into modules specific for a single substrate. Modules can be further subdivided into domains. Module 1 is an initiation module, module 2 is an elongation module and module 3 is a release module.
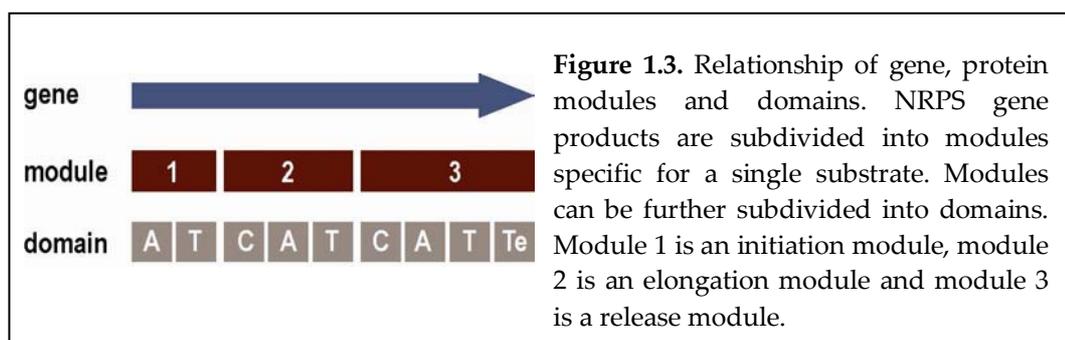
Mg^{2+}-ATP (Pavela-Vrancic *et al.*, 2004). Activated substrate is then covalently bound to the carrier domain, the T domain (described below).

A domains are about 550 amino acids in length and contain 10 highly conserved motifs proposed to be involved in activation and binding of substrate (Marahiel *et al.*, 1997). A domains are highly similar between NRPSs and share significant homology with the family of acyl-CoA synthases and luciferases (Conti *et al.*, 1996).

The crystal structure of the A domain of GrsA (designated PheA), which inserts phenylalanine into the decapeptide antibiotic gramicidin S, has been obtained and its three-dimensional structure solved (Conti *et al.*, 1997). Eight amino acids found between core motifs A3 and A6 have been shown to be involved in the specificity of the binding pocket of PheA for phenylalanine (Conti *et al.*, 1997). Using sequence alignments of other A domains with PheA, the equivalent 8 residues can be found in other A domains and these can be used to determine the specificity of domains with unknown specificity using a BLAST algorithm (Stachelhaus *et al.*, 1999; Challis *et al.*, 2000). This approach has been useful in several systems, however where there is no experimentally determined substrate for a certain 8-residue signature, this approach for specificity determination cannot be used (Walton *et al.*, 2004). Also, the 8 residues do not appear to be representative in fungal A domains, possibly fungal A domains are slightly different in their three dimensional structure or the slightly more relaxed substrate specificity in fungi may mean a sequence-based approach to substrate prediction is not possible (Walton *et al.*, 2004).

1.4.1.2 Thiolation Domains

The thiolation domain (T domain), also called the peptidyl carrier protein (PCP) domain, is approximately 80 amino acids long and is responsible for covalently tethering the activated aminoacyl substrate and shuttling the substrate to the next module (Stachelhaus *et al.*, 1996). The T domain requires the posttranslational addition of a 4'phosphopantetheinyl (Ppant) group from coenzyme A to the side-chain of the conserved serine residue in the T domain signature sequence (Stachelhaus *et al.*, 1996). Activated aminoacyl adenylate substrates are covalently

attached as a thioester to the thiol moiety of the Ppant which acts as a "swinging arm" that shuttles the substrate from one cognate module to the next where a peptide bond is formed, the growing peptidyl chain is then shuttled to the next module via the next Ppant and so on until release from the final module (Mootz *et al.*, 2002).

The posttranslational attachment of the Ppant group to the T domain serine is catalysed by a 4' phosphopantetheinyl transferase (Pptase) (Lambalot *et al.*, 1996). In bacterial systems these enzymes are often specific to individual NRPSs, although some have broad substrate-specificity, for example, the *Bacillus subtilis* Pptase associated with the surfactin synthetase (Reuter *et al.*, 1999). In fungal systems there appears to be just one Pptase that modifies each NRPS (Keszenman-Pereyra *et al.*, 2003).

1.4.1.3 Condensation Domains

C Domains are approximately 450 amino acids in size and are found between each pair of A and T domains. There are seven sequence motifs found in C domains although the sequence is not as conserved as that of the A and T motifs (Marahiel *et al.*, 1997). C domains catalyse the condensation of the peptide bond between the carboxyl group of the upstream bound substrate and the free amino group of the downstream bound substrate, thereby translocating the growing chain onto the next module. C domains contain two sites, one for electrophiles (the donor) and one for nucleophiles (the acceptor) that control the directionality of the synthesis (Mootz *et al.*, 2002). The acceptor site binds the downstream substrate with high affinity until the incoming upstream substrate has bound to the donor site and the condensation process is completed (Linne and Marahiel, 2000). The acceptor site also appears to have an editing function; it discriminates, to a certain extent, against amino acids of opposite stereochemistry and with larger side chains. The donor site is, however, more tolerant (Belshaw *et al.*, 1999; Ehmann *et al.*, 2000).

1.4.1.4 Release Domains

Growing peptidyl chains are covalently attached to the NRPS enzymes via the T domain. Different types of C-terminal domains have evolved that catalyse the release of the non-ribosomal peptide once the full chain-length has been reached.

The most common form of release domain in bacteria is the thioesterase (Te) domain, named after their similarity to thioesterases found in fatty acid metabolism (Pazirandeh *et al*., 1989). The full-length acylthioester is passed from the final T domain to a serine residue in the Te domain. Keating *et al* (2001) postulate that the serine-bound peptide is then open to hydrolysis from water and released from the NRPS. When cyclic NRPs are released it is thought that specific binding sites in the Te domain for the amino- and carboxy-terminal substrates sequester the NRP long enough for intramolecular capture to occur before hydrolysis (Keating *et al*., 2001; Kohli *et al*., 2001).

Fungi, with a few exceptions such as ACV synthetases, which may be of bacterial origin (Section 1.3.3), seldom contain Te release domains and rather contain an additional C domain after the final T domain. No evidence exists that the peptidyl product is covalently transferred to the C release domain and it is thought that these domains may directly catalyse condensation between the amino- and carboxy-terminal groups of the non-ribosomal peptide (Keating *et al*., 2001). C domains may also directly cleave the thioester bond, as do acyl transferases (Dutnall *et al*., 1998).

The final known release domain is a reductase domain. This form of release domain uses NADPH as a co-substrate for hydride transfer, reducing the thioester bond to a hemithioaminal linkage which decomposes spontaneously to release the free non-ribosomal peptide, as described for the analogous system in lysine biosynthesis (Ehmann *et al*., 1999). NRPS systems containing this release domain include saframycin (Pospiech *et al*., 1996), mycobacterial peptidolipids (Billman-Jacobe *et al*., 1999), peptaibols (Wiest *et al*., 2002) and peramine in epichloë (Tanaka *et al*, 2005).

1.5 Ergot Alkaloids

Ergot alkaloid chemistry is reviewed in Floss, (1976). Ergot alkaloid genetics is reviewed in Panaccione and Schardl, (2002) and Tudzynski *et al.*, (2001b) and all aspects of ergot alkaloid synthesis, chemistry and significance were recently comprehensively reviewed in Schardl *et al.*, (2006).

1.5.1 Historical Overview

Ergot alkaloids are composed of clavines derived from dimethylallyl tryptophan, simple amides of lysergic acid, and peptide lysergic acid derivatives (ergopeptines). They are compounds found in the sclerotia, or ergots, of fungal plant pathogens from the genus *Claviceps,* which usually infect rye but can also infect wheat, barley, rice, corn, millet and oats (Tudzynski and Scheffer, 2004). The hard, dark-coloured, ergots take the place of grains in the host grass and can be eaten by grazing animals. The ergot alkaloids can also be eaten by humans when rye is harvested and ergots are milled along with the grains, the alkaloids thereby contaminating bread (Eadie, 2003). The ergot alkaloids are extremely toxic and contaminated ryebread was responsible for outbreaks of ergotism, called St Anthony's Fire, from the middle ages right up until the last recorded outbreak in Ethiopia in 1977 (Demeke *et al.*, 1979). The effects of ergotism could be fever, hallucinations, convulsions and gangrene (Eadie, 2003).

Ergot alkaloids were recognised for their pharmacologic potential in the Middle Ages when ergots were given to pregnant women to induce labour and to reduce haemorrhaging. The vasoconstrictive properties of ergot were recognised as a treatment for migraine in the 19th century (Woakes, 1868). Today ergopeptines are a common component of migraine pharmaceuticals and are also being used in Parkinson's Disease treatment (Tudzynski *et al.*, 2001b). The pharmacological effects of ergot alkaloids, mainly the ergopeptines, are generally due to the mimicking of neurotransmitters such as dopamine, seratonin and noradrenaline. The alkaloids thus function as partial agonists or antagonists at adrenergic, dopaminergic, and tryptaminergic receptors (Tudzynski *et al.*, 2001b).

1.5.2 Ergot Alkaloids in Endophyte-Infected Pasture

Tall fescue toxicosis, a syndrome similar to ergotism, as a result of stock feeding on tall fescue (*Festuca arundinacea*) was first recognised in New Zealand in 1948 (Cunningham, 1948). Livestock maladies associated with tall fescue toxicosis can include poor weight gain, hyperthermia, convulsions, reduced fertility and lactation, gangrene of the extremities and death (Bacon, 1995). The symptoms of the toxicosis can vary depending on environmental conditions. In warm conditions, affected animals lose control of body temperature and show poor weight gain. In cold conditions livestock are more likely to suffer gangrene of the limbs (Bacon, 1995). Agricultural losses as a result of ergot alkaloid poisoning are significant; the associated economic cost may currently approach US$1 billion per annum (Panaccione *et al.*, 2001). It was not until 1977 however that fungal endophytes were conclusively shown to be responsible for the toxicosis (Bacon *et al.*, 1977). Ergot alkaloid biosynthesis was later shown in *N. coenophialum* (Lyons *et al.*, 1986) and implicated in the toxicosis due to the similarities with ergotism. A focus of endophyte research has been isolating strains that are deficient in the production of ergot alkaloids, and the toxic lolitrems, but retain the endophyte properties desired for agriculture (Tapper and Latch, 1999). Some of these strains, such as *N. lolii* strains AR1 and AR37, and *N. coenophialum* strain AR542 (sold as MaxP/MaxQ), are now commercially available and have been shown to confer many benefits of endophyte infection with no, or reduced, toxicity (Parish *et al.*, 2003a; Parish *et al.*, 2003b). There is a trade-off with selection of non-ergot alkaloid-producing strains however as ergovaline is implicated in deterrence of black beetle (*Heteronychus arator*) (Ball *et al.*, 1997), an important pasture pest in warmer areas of New Zealand.

1.5.3 Ergot Alkaloid Biosynthesis

A simplified diagram of ergot alkaloid biosynthesis is shown in Fig. 1.4. The first committed step in ergot alkaloid biosynthesis is the formation of dimethylallyl tryptophan (DMAT) from primary metabolites tryptophan and dimethylallyl diphosphate. This reaction is catalysed by the prenyl transferase DMAT synthase (Gebler and Poulter, 1992). Serial redox reactions form clavine intermediate

compounds such as chanoclavine, agroclavine and elymoclavine, which are converted to D-lysergic acid (Tudzynski *et al*., 2001b). D-lysergic acid can be converted into several lysergyl amides such as D-lysergic acid *α*-hydroxyethylamide and ergometrine (Tudzynski *et al*., 2001b).

Ergopeptines are formed by the NRPS-catalysed addition of various peptides to activated lysergic acid. Ergotamine, the ergopeptine most abundantly synthesised by *C. purpurea,* is formed by the sequential addition of three amino acids, alanine, phenylalanine and proline, to lysergic acid (Walzel *et al*., 1997). Lysergic acid is activated and tethered to the LpsB (lysergyl peptide synthetase B, formerly Lps2) enzyme, a single-module NRPS (Riederer *et al*., 1996). C- domain mediated catalysis adds the activated lysergyl group to the activated alanine tethered to the first module of LpsA (formerly Lps1), a three-module NRPS (Riederer *et al*., 1996). Phenylalanine and proline are sequentially added and the tripeptide is heterocyclised and released from LpsA as lysergyl peptide lactam, a diketopiperazine (Walzel *et al*., 1997). One further heterocyclisation step is then undergone to yield the final product, ergotamine (Quigley and Floss, 1981). Various different ergopeptines are synthesised that differ in the first two amino acids of the tripeptide, with proline always in position three (Table 1.1). Ergovaline, the most abundant endophyte-produced ergopeptine, differs from ergotamine by the inclusion of valine rather than phenylalanine at the second amino acid.

Table 1.1. Variation of amino acids in positions one and two* of ergopeptines

	Position 1	
Position 2	ala	val
phe	ergotamine	ergocristine
val	ergovaline	ergocornine
leu	ergosine	ergokryptine

*Proline is invariant in position three.
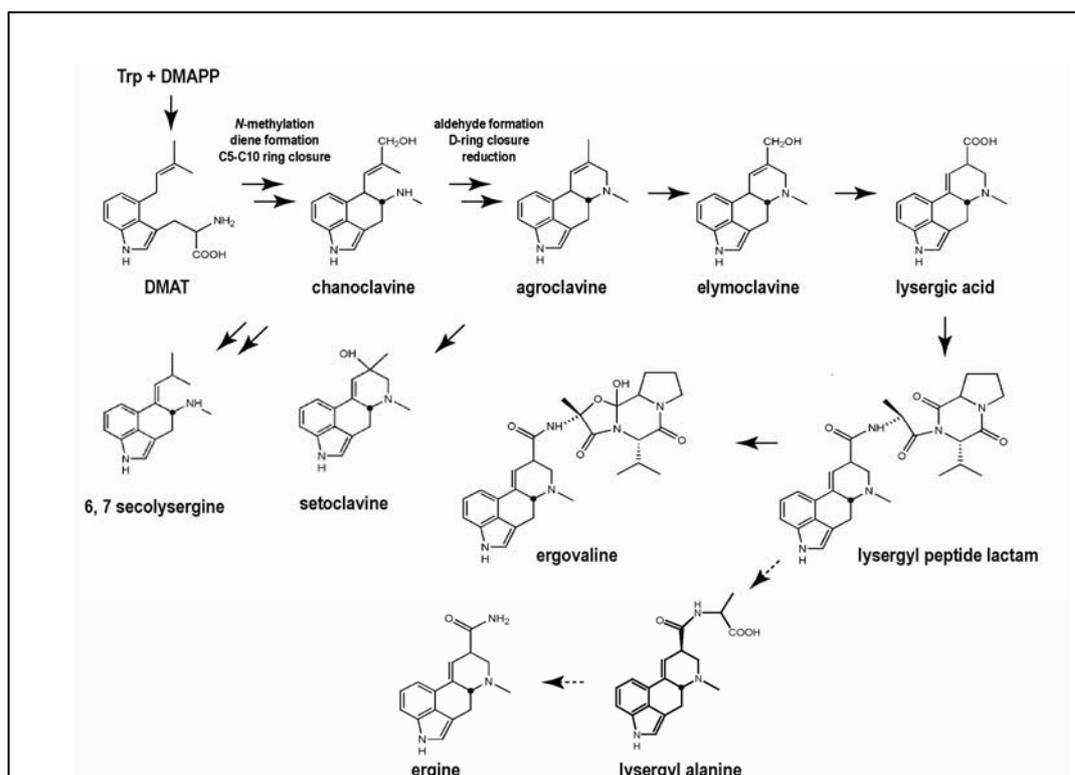
Figure 1.4. Simplified epichloë ergot alkaloid biosynthetic pathway. Solid arrows denote steps that are known or highly likely. Dashed arrows denote steps that are tentative based on Panaccione *et al* (2003). Two arrows denote more than one catalytic reaction (these steps are described above). Trp, tryptophan; DMAPP, dimethylallyl diphosphate; DMAT, dimethylallyl tryptophan.

1.5.4 Ergot Alkaloid Genetics

The gene coding for DMAT synthase, designated *dmaW*, was first cloned from *Claviceps fusiformis* using degenerate PCR primers based on partial protein sequence (Tsai *et al*., 1995). *cpd*1, a *dmaW* homologue, was subsequently cloned from *C. purpurea* and shown to be part of a cluster of genes including *lpsA-1* (formerly cpps1), the gene encoding LpsA, and two oxidoreductases, *easE* (cpox1) and *easD* (cpox2), hypothesised to be responsible for intermediate steps in the pathway (Tudzynski *et al*., 1999). Further sequencing and analysis of this region identified several more genes that may be involved in ergot alkaloid biosynthesis, including genes for a second LpsA NRPS (*lpsA-2*) two single-module NRPSs (*lpsB* and *lpsC*), a third oxidoreductase (*easA*), a cytochrome P450 monooxygenase (*cloA*) and two

hypothetical genes (*easF* and *easG*) (Tudzynski *et al.*, 2001b; Correia *et al.*, 2003; Haarmann *et al.*, 2005). Analysis of the duplicated *lpsA* genes showed divergence of the first two modules, suggesting a specific *lpsA* gene for each of the two ergopeptines produced by the *C. purpurea* strain analysed (Haarmann *et al.*, 2005). This was recently supported by functional analysis of *lpsA-1*, a mutant of which was able to synthesise ergocryptine but not ergotamine (Haarmann *et al.*, 2007).

Two other genes in the *C. purpurea* gene cluster have been functionally characterised. One of the single-module NRPS genes in the *C. purpurea* cluster, *lpsB*, was shown to be the gene encoding LpsB, which activates D-lysergic acid (Correia *et al.*, 2003). Expression of *lpsB* was analysed and shown to be downregulated by the presence of high phosphate (Correia *et al.*, 2003), which is known to repress ergot alkaloid synthesis in *C. purpurea* (Krupinski *et al.*, 1976). Mutation of the *cloA* gene, predicted to encode a P450 monooxygenase, showed it was responsible for the conversion of elymoclavine to lysergic acid. This was confirmed by feeding the mutant lysergic acid, which showed restoration of ergopeptine formation (Haarmann *et al.*, 2006).

Analysis of the *A. fumigatus* genome sequence revealed a gene cluster containing homologues of eight of the genes in the *C. purpurea* cluster (Coyle and Panaccione, 2005; Unsold and Li, 2005). These shared genes are predicted to encode enzymes that catalyse shared steps in the ergot alkaloid pathway, probably up to agroclavine, while genes unique to each species are likely to encode genes required for steps unique to each organism: lysergic acid and ergopeptine synthesis in *C. purpurea* (and epichloë) and further clavine derivatives in *A. fumigatus*. Functional analysis of the *dmaW* gene in the *A. fumigatus* cluster confirmed a role in ergot alkaloid synthesis (Coyle and Panaccione, 2005; Unsold and Li, 2005).

In endophytes the *dmaW* gene in *Neotyphodium* sp. Lp1 was recently cloned and shown by mutant analysis to be required for ergot alkaloid biosynthesis in that system (Wang *et al.*, 2004). The *N. lolii lpsA* gene was isolated using a *C. purpurea* probe and low-stringency hybridisation (Panaccione *et al.*, 2001) and subsequently

fully sequenced (Damrongkool *et al.*, 2005). Using the *N. lolii* gene sequence, targeted replacement of *lpsA* was performed in *Neotyphodium* sp. Lp1. The Δ*lpsA* mutant was unable to synthesise ergovaline, however no effect on plant infection or growth was apparent, indicating that ergovaline plays no role in the interaction between symbionts (Panaccione *et al.*, 2001). An examination of the biochemical outcome of the *lpsA* deletion showed that while ergovaline synthesis had been blocked, there was no build-up of intermediates above wild-type levels, indicating tight regulation of this pathway (Panaccione *et al.*, 2003). Interestingly, the synthesis of two other compounds, ergine and lysergyl alanine, were also blocked by the Δ*lpsA* mutation. It seems unlikely that these are direct products of the NRPS and it is proposed that they arise as break-down products of lysergyl peptide lactam (Panaccione *et al.*, 2003). Neither *lpsA* or *dmaW* were shown to be linked with other ergovaline biosynthetic genes in these studies.

1.6 Background and Aims of this Study

The gene to be characterised in this study was isolated in a screen for the peramine NRPS by Tanaka *et al* (2005). Degenerate primers were designed to the conserved A3 and A7 motifs of fungal NRPSs and PCR was performed using DNA extracted from *N. lolii*-infected *L. perenne* as template. Three different candidate NRPS genes were identified from this screen, ps7, ps9 and ps12. Expression studies showed that ps9 and ps12 were up-regulated *in planta*, as was expected for peramine synthetase as peramine is not produced in culture. Analysis of the taxonomic distribution of ps9 and ps12 showed that ps9 was present in all peramine-producing strains while ps12 was not. Thus ps9 was characterised further and shown to be the peramine NRPS gene, *perA*. The function of ps12 remained unknown and was to be the focus of this PhD research.

The first aim of this study was to determine the function of the candidate NRPS gene, ps12. Subsequent aims, based on sequence identity of ps12 with the *C. purpurea lpsB* gene, were to test the following hypotheses:

1. That ps12 was an orthologue of *C. purpurea lpsB*, involved in the synthesis of the toxic ergot alkaloid ergovaline

2. That the ps12 gene was physically linked with *dmaW, lpsA* and other yet to be isolated ergovaline biosynthetic genes as part of a gene cluster

3. That ergot alkaloid genes were co-ordinately expressed *in planta*, concurrent with alkaloid production

4. That the taxonomic distribution of different ergovaline biosynthetic genes corresponded with production of different ergot alkaloid compounds

Testing these core hypotheses lead to the finding that the ps12/*lpsB* gene was located in a complex cluster of genes, the *EAS* cluster, proposed to be required for ergot alkaloid synthesis. This cluster was extended by analysing the recently-released genome sequence of *E. festucae* E2368.

The next section of research, described in Chapter 4, aimed to characterise aspects of *EAS* gene regulation by testing various culture conditions for ps12/*lpsB* induction, analysing the *EAS* gene promoters for potential *cis* activating elements, and heterologous expression in *E. festucae* Fl1 of the candidate regulatory gene *laeA* from *A. nidulans*.

Research described in Chapter 5 describes the analysis of several transposons identified within the *EAS* cluster.

The final phase of research, described in Chapter 6, aimed to further characterise the *EAS* cluster in epichloë by comparative analysis of the cluster within four different ryegrass-infecting strains: *N. lolii* Lp19, *N. lolii* AR1, *E. festucae* Fl1 and *Neotyphodium* sp. Lp1. A key aim of this analysis was to determine whether the dramatic difference in ergot alkaloid production between these strains was due to gene expression, and, if so, whether this was due to differences in the organization or sequence of the clusters.

2. Materials and Methods

2.1 Biological Material

Fungal, plant and bacterial strains, plasmids and λ clones used in this study are described in Table 2.1.

Table 2.1. Biological material

Biological material	Relevant Characteristics	Reference
Fungal strains		
Epichloë festucae Fl1	Host: *Festuca longifolia*	(Young *et al.*, 2005)
E. typhina E8	Host: *Lolium perenne*	(Schardl *et al.*, 1994)
Neotyphodium lolii AR1	Host: *L. perenne*	(Johnson *et al.*, 2007b)
N. lolii AR40	Host: *L. perenne*	(Johnson *et al.*, 2007b)
N. lolii AR47	Host: *L. perenne*	(Johnson *et al.*, 2007b)
N. lolii AR59	Host: *L. perenne*	(Johnson *et al.*, 2007b)
N. lolii Lp11	Host: *L. perenne*	(Christensen *et al.*, 1993)
N. lolii Lp19	Host: *L. perenne*	(Christensen *et al.*, 1993)
Neotyphodium sp. Lp1	Host: *L. perenne*, LPTG-2 (AR6)	(Christensen *et al.*, 1993)
N. coenophialum AR512	Host: *L. arundinaceum*	(Christensen *et al.*, 1993)
N. coenophialum AR542	Host: *L. arundinaceum*	(Johnson *et al.*, 2007b)
N. coenophialum AR565	Host: *L. arundinaceum*	(Christensen *et al.*, 1993)
N. coenophialum Tf5	Host: *L. arundinaceum*	(Christensen *et al.*, 1993)
Neotyphodium sp. AR553	Host: *L. arundinaceum*, FATG-2 (Tf13)	(Christensen *et al.*, 1993)
Neotyphodium sp. AR501	Host: *L. arundinaceum*, FATG-3 (Tf16)	(Christensen *et al.*, 1993)
Neotyphodium sp. AR507	Host: *L. arundinaceum*, FATG-3	(Christensen *et al.*, 1993)
Neotyphodium sp. AR510	Host: *L. arundinaceum*, FATG-3	(Christensen *et al.*, 1993)
N. uncinatum Fp1	Host: *F. pratense*	(Christensen *et al.*, 1993)
DFM3	*E. festucae* Fl1/Δ*lpsB*::P*gpdA-hph*, Hyg^R	This study
DFM7	DFM3/pDF1/pII99, Hyg^R, Gen^R	This study
DFM8	DFM3/pDF1/pII99, Hyg^R, Gen^R	This study
DFM9	DFM3/pDF1/pII99, Hyg^R, Gen^R	This study
DFM10	*E. festucae* Fl1/ P*tef1-AnlaeA*-T*gluA*/pPN1688, Hyg^R	This study
DFM11	*E. festucae* Fl1/ P*tef1-AnlaeA*-T*gluA*/pPN1688, Hyg^R	This study
DFM12	*E. festucae* Fl1/ P*tef1-AnlaeA*-T*gluA*/pPN1688, Hyg^R	This study
DFM13	*E. festucae* Fl1/ P*tef1-AnlaeA*-T*gluA*/pPN1688, Hyg^R	This study
DFM14	*E. festucae* Fl1/pPN1688, Hyg^R	This study
DFM15	*E. festucae* Fl1/pPN1688, Hyg^R	This study
DFM16	*Neotyphodium* sp. Lp1/Δ*lpsA*::P*cpc1-hph*-T*trpC*, HygR	(Panaccione *et al.*, 2001)
DFM17	*Neotyphodium* sp. Lp1/Δ*dmaW*:: P*cpc1-hph*-T*trpC*, Hyg^R	(Wang *et al.*, 2004)
Aspergillus nidulans NRRL 194	wild type	(Pontecorvo *et al.*, 1953)
Plant symbiota		
EGH4982 – 4989	*L. perenne*/DFM3	This study
EGH4990 - 4997	*L. perenne*/*E. festucae* Fl1	This study
EGH4998 – 5002	*L.perenne*/*Neotyphodium* sp. Lp1	This study
EGH5003 – 5007	*L. perenne*/DFM16	This study
EGH5008 – 5012	*L. perenne*/DFM17	This study
EGH5013 – 5025	*L. perenne*, endophyte free	This study
EGH4667 – 4670	*L. perenne*/DFM10	This study
EGH4671 – 4674	*L. perenne*/DFM11	This study
EGH4683 – 4686	*L. perenne*/DFM12	This study
EGH4695 – 4698	*L. perenne*/DFM13	This study
EGH4679 – 4682	*L. perenne*/DFM14	This study
EGH4687 – 4690	*L. perenne*/DFM15	This study
WSC1	*L. perenne*/*N. lolii* AR1	W. Simpson, AgResearch
WSC2	*L. perenne*/*Neotyphodium* sp. Lp1	W. Simpson, AgResearch

Biological material	relevant characteristics	Reference
WSC3	*L. perenne/N. lolii* Lp19	W. Simpson, AgResearch
WSC4	*L. perenne/E. festucae* Fl1	W. Simpson, AgResearch

Bacterial strains		
XL1-Blue	*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* [F' *proAB lac*IqZΔM15 Tn*10* (Tet^R)]	(Bullock *et al.*, 1987)
SOLR	e14–(McrA–) Δ(*mcr*CB-*hsd*SMR-*mrr*)*171 sbc*C *rec*B *rec*J *uvr*C *umu*C::Tn*5* (Kan^R) *lac gyr*A96 *rel*A1 *thi*-1 *end*A1 λR [F′ *proAB lac*IqZΔM15] Su⁻	Stratagene
KW251	F⁻ *mcr*B1 *mcr*A *hsd*R2 *rec*D1014	Promega
Top 10	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G	Invitrogen

Plasmid		
pAN7-1	Amp^R, Hyg^R	(Punt *et al.*, 1987)
pBluescript	Amp^R	Stratagene
pCR2.1 TOPO	Amp^R, Kan^R	Invitrogen
pCR4Blunt-TOPO	Amp^R, Kan^R	Invitrogen
pDONR 221	Kan^R	Invitrogen
pDONR P2R-P3	Kan^R	Invitrogen
pDONR P4-P1R	Kan^R	Invitrogen
pDEST R4-R3	Amp^R,Cam^R	Invitrogen
pDF1	pBluescript excised *N. lolii* Lp19 λZAPII phagemid containing *lpsB*	This study
pDF2	pBluescript excised *N. lolii* Lp19 λZAPII phagemid containing 5'*lpsB – easE – easF*	This study
pDF3	pBluescript excised *N. lolii* Lp19 λZAPII phagemid containing *easE – easF –* 5' *easG*	This study
pDF6	pDEST R4-R3 multisite gateway clone containing 3 kb 5' *lpsB – hph* (from pAN7-1) – 3 kb 3' *lpsB*, Amp^R, Hyg^R	This study
pDF7	pDEST R4-R3 multisite gateway clone containing P*tef1-AnlaeA-TgluA*, Amp^R	This study
pDF8	pBluescript excised *N. lolii* Lp19 λZAPII phagemid containing 5' *lpsA*	This study
pENTRhph	pDONR 221 containing pAN7-1 *hph* cassette, Kan^R, Hyg^R	This study
pENTRlaeA	pDONR 221 containing *AnlaeA* ORF, Kan^R	This study
pENTRTEF	pDONR P4-P1R containing P*tef1*, Kan^R	R. Johnson and C. Voisey, unpublished
pENTRTERM	pDONR P2R-P3 containing T*gluA*, Kan^R	R. Johnson and C. Voisey, unpublished
pENTR3'lpsB	pDONR P2R-P3 containing 3-kb of 3' *lpsB* sequence, Kan^R	This study
pENTR5'lpsB	pDONR P4-P1R containing 3-kb of 5' *lpsB* sequence, Kan^R	This study
pII99	Amp^R, Gen^R	(Namiki *et al.*, 2001)
pPN1688	Amp^R, Hyg^R	(Young *et al.*, 2005)

Lambda clones		
λDF1	*N. lolii* Lp19 chimeric λGem12 clone containing *torA* [rearrangement] *easH – easA – easG – easF – easE*	This study
λDF2	*N. lolii* Lp19 λGem12 clone containing 3' *easG – easF – easE – lpsB*	This study

2.2 Fungal Growth and Media

Fungal strains were cultured at 22°C. Complex media used for growth was potato dextrose (PD) broth or agar, supplemented with hygromycin (150 – 200 µg/mL) or geneticin (220 µg/mL) where appropriate. Defined media used for expression analysis was either Mantle A (Mantle and Nisbet, 1976) with high (2.0 g/L) or low (0.2 g/L) KH_2PO_4, or Czapek-Dox (CD) salts plus nitrogen (100 mM NH_4Cl) and/or carbon source (100 mM glucose). Cultures incubated in plant extract were first grown in PD broth for three days, then filtered through sterile Whatman 3M paper and added to the extract for 30 min. The plant extract was prepared essentially using the method of Lev *et al*., (2005). Endophyte-free perennial ryegrass tillers were ground to a fine powder under liquid nitrogen and mixed 1:1 (w/v) with sterile distilled H_2O. The suspension was centrifuged at 13,000 rpm for 5 min and the supernatant removed for use.

2.3 Bacterial Growth and Media

E. coli strains were cultured at 37°C in Luria Bertani (LB) broth or agar. Ampicillin (100 µg/mL), kanamycin (50 µg/mL), isopropyl β-D-thiogalacto-pyranoside (IPTG, 20 µg/mL) and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-GAL, 40 µg/mL) were added as appropriate.

2.4 Enzymes and Chemicals

Enzymes and chemicals were purchased from Invitrogen, Roche and New England Biolabs. Antibiotics were purchased from Sigma, dissolved to an appropriate concentration in distilled water and filter sterilised through a 0.22 µm syringe filter. All chemicals were analytical grade and solutions were prepared in water unless otherwise indicated.

2.5 DNA Isolation

2.5.1 Plasmid DNA extraction

To extract plasmid DNA for routine screening, the boiling miniprep method, adapted from (Holmes and Quigley, 1981), was used. Cultures were grown at 37°C

with shaking at 200 rpm for approximately 18 hours in LB broth, then 1.5 mL harvested by centrifugation. Pellets were resuspended in 350 μL of HQ-STET buffer (8% [w/v] sucrose, 50 mM EDTA [pH 8], 5 mM Tris [pH8], 5% [w/v] TritonX-100). 25 μL of lysozyme (10 mg/mL in HQ-STET buffer) was added to each tube and the tubes were boiled for 40 s. Immediately after the boiling step, the tubes were centrifuged at 15000 × g for 10 min and the resulting gelatinous pellet was removed and discarded. Plasmid DNA was then precipitated by the addition of 450 μL of isopropanol followed by centrifugation at 15000 × g for 5 min. DNA pellets were washed in 70% ethanol, air dried and resuspended in 50 μL of filter-sterile Milli-Q water.

High quality plasmid DNA preparations to be used for sub-cloning, transformation or as templates for automated sequencing were prepared using either the QIAprep Spin Miniprep or Midiprep kits (Qiagen), according to the instruction manuals supplied by the manufacturers.

2.5.2 Genomic DNA Isolation

Fungal genomic DNA was extracted using methods adapted from Byrd *et al.*, (1990), Al-Samarrai and Schmid, (2000) and Wu *et al.*, (1995).

For routine screening of fungal transformants, a method adapted from (Wu *et al.*, 1995) was used to extract small amounts of genomic DNA. Fungal mycelium was cultured in PD broth for three days in 1.5 mL tubes, pelleted by centrifugation and freeze-dried, then ground to a fine powder using a micropestle. 150 μL of lysis buffer (100 mM Tris [pH 8], 100 mM EDTA [pH 8], 1% [w/v] SDS) was added, mixed by pipetting and incubated at 70°C for 30 min. Lysates were mixed with 150 μL of 5 M potassium acetate, incubated on ice for 10 min and centrifuged at 15000 × g for 10 min. DNA was precipitated from the supernatant by addition of 0.6 volumes isopropanol and centrifugation at 15000 × g for 15 min. DNA pellets were washed twice with 70% ethanol, air-dried and resuspended in 20 μL filter-sterile Milli-Q water.

Large amounts of genomic DNA for Southern analysis were prepared using the method of either Byrd (1990) or Al-Samarrai (2000). For the Byrd method, 30 mg of freeze-dried mycelium was ground to a fine powder under liquid N_2 and suspended thoroughly in 500 µL of extraction buffer (150 mM EDTA [pH 8], 50 mM Tris [pH 8], 1% sodium lauroyl sarcosine, 2 mg/mL proteinase K). The solution was incubated at 37°C for 20 min then at 65°C for 1 – 2 h followed by centrifugation at 15000 × g for 10 min. The supernatant was removed, mixed with an equal volume of phenol/chloroform (1:1 v/v) and centrifuged at 15000 × g for 10 min. The aqueous phase was removed and the procedure repeated until the aqueous phase was clear, then the aqueous phase was mixed with an equal volume of chloroform and centrifuged at 15000 × g for 10 min. To the aqueous phase obtained from the chloroform extraction, an equal volume of isopropanol was added, mixed by inversion and incubated at –20°C for 20 min followed by centrifugation at 15000 × g for 10 min. To remove remaining polysaccharides, the pellet was resuspended in 1 M NaCl and centrifuged at 15000 × g for 10 min, DNA was then precipitated from the supernatant by mixing with an equal volume of isopropanol, incubation at –20°C for 10 min and centrifugation at 15000 × g for 10 min. DNA pellets were washed three times with 70% ethanol, air-dried and resuspended in 50 µL TE buffer (10 mM Tris [pH 8], 0.1 mM EDTA [pH8]).

For the Al-Samarrai method, 30 mg freeze-dried mycelium was ground to a fine powder under liquid N_2 and resuspended in 500 µL lysis buffer (40 mM Tris [pH 8], 20 mM sodium acetate, 1 mM EDTA [pH 8], 1% [w/v] SDS) by approximately 30 cycles of vigorous pipetting, until the viscosity was significantly reduced. 2 µL RNase A (10 mg/mL) was added and the mixture incubated for 5 min at 37°C. Cell debris was removed by addition of 165 µL of 5 M NaCl, inversion and centrifugation at 15000 × g for 20 min. The supernatant was removed, mixed with an equal volume of phenol/chloroform (1:1 v/v) and centrifuged at 15000 × g for 20 min. The aqueous phase was removed and the procedure repeated with chloroform until the aqueous phase was clear. DNA was precipitated from the aqueous phase by addition of 0.1 volumes of 3 M sodium acetate and 2.5 volumes of 96% ethanol.

DNA pellets were washed three times with 70% ethanol, air-dried and resuspended in 50 μL TE buffer (10 mM Tris [pH 8], 0.1 mM EDTA [pH8]).

2.5.5 Chromosomal DNA Isolation in Agarose Plugs

To prepare chromosomal DNA for pulsed-field gel electrophoresis, fungal protoplasts (section 2.13.1) were resuspended in STC buffer (1 M sorbitol, 50 mM CaCl$_2$, 50 mM Tris [pH 8.0]) to a concentration of 1×10^9/mL, mixed with an equal volume of 1.4% low melting point agarose in GMB buffer (0.9 M sorbitol, 125 mM EDTA [pH 7.5]) and allowed to set in plug moulds. The plugs were then incubated at 50°C in SE buffer (2% [w/v] SDS, 250 mM EDTA [pH 8]) for 18 h, followed by 24 h at 50°C in 10 mL 10 x ET buffer (10 mM Tris [pH 8], 500 mM EDTA [pH 8]) with 20 mg of proteinase K (Invitrogen) and 100 mg of sodium lauroylsarcosine (Sigma). Plugs were finally washed four times with 1 x ET buffer and stored in 1 x ET buffer at 4°C.

2.5.4 λ DNA Isolation

Bacteriophage λ DNA was isolated from plate lysates using PEG precipitation and phenol/chloroform purification based on the method in Sambrook *et al*., (1989). Recombinant phage were plated for confluent lysis on LB agarose (1.5%) plates, overlaid with 5 mL of SM buffer (100 mM NaCl, 8 mM MgSO$_4$.7H$_2$O, 0.01% (w/v) gelatine, 50 mM Tris [pH 8]) and left overnight at 4°C. The lysate was collected and treated with DNase and RNase (1 μg/mL final) and incubated at 37°C for 30 min. 5 mL of PEG solution (20% PEG 6000 (w/v), 2 M NaCl) was added, the solution mixed and incubated on ice for 1 h. Phage were pelleted by centrifugation at 5800 g for 30 min at 4°C, resuspended in 500 μL of SM buffer with 5 μL of 10% (w/v) SDS and 10 μL of 250 mM EDTA (pH 8), and incubated at 68°C for 15 min. To remove phage proteins, an equal volume of phenol was added, the mixture was vortexed for 10 s, incubated at room temperature for 5 min and centrifuged at 15000 × *g* for 5 min. The aqueous phase was removed and the procedure repeated with phenol/chloroform (1:1 [v/v]), then chloroform. To the aqueous phase obtained from the chloroform extraction, an equal volume of isopropanol was added, mixed by inversion and incubated at –20°C for 20 min followed by centrifugation at 15000

× *g* for 10 min. DNA pellets were washed with 70% ethanol, air dried and resuspended in 50 μL TE buffer (10 mM Tris [pH 8], 0.1 mM EDTA [pH8]).

2.6 DNA Manipulation

2.6.1 DNA Quantification

DNA was quantified by comparison to λ DNA standards of known concentration or by spectrophotometric quantification using a NanoDrop (NanoDrop Technologies) spectrophotometer according to the manufacturer's instructions.

2.6.2 Restriction Digests

Restriction endonuclease digests were performed using appropriate commercial buffers with ~3 – 10 units of restriction enzyme/μg of DNA. Digests were performed at the recommended temperature in a water bath or incubator.

Chromosomal DNA embedded in agarose plugs for pulsed-field gel separation (section 2.6.4.2) was equilibrated three times in 10 mM Tris (pH 8), then in 1× reaction buffer and 1× BSA (New England Biolabs) for 30 min at 37°C. Plugs were then placed in 1× reaction buffer with 100 – 200 U of restriction enzyme, equilibrated on ice for 10 min and incubated overnight at the appropriate temperature.

2.6.3 DNA Purification and Precipitation

Purification of DNA following enzyme reactions and PCR product purification was performed using the QIAQuick and MinElute PCR purification kits (Qiagen) according to instruction manuals supplied by the manufacturer.

To extract DNA from agarose gels, gel slices were excised under long wavelength UV light and DNA fragments purified using the QIAQuick and MinElute gel extraction kits (Qiagen) according to manufacturers instructions.

DNA solutions were concentrated using either a Centricon column (Millipore) and the manufacturers instructions or precipitated and resuspended in lower volumes. For precipitations, either 0.1 volumes of 3 M sodium acetate and 2.5 volumes of 96% ethanol or 0.6 volumes of isopropanol were added, mixed by inversion and

incubated on ice for at least 10 min. DNA was pelleted by centrifugation at 15000 × g for 10 min, washed with 70% ethanol and resuspended in either H_2O or TE buffer (10 mM Tris [pH 8], 0.1 mM EDTA [pH 8]).

2.6.4 Agarose Gel Electrophoresis

2.6.4.1 Standard Electrophoresis

DNA was mixed with Blue Juice (Invitrogen) tracking dye and separated on 1 – 2% agarose gels as appropriate, made and run in 1× Tris-acetate (TAE) buffer (40 mM Tris [pH 8], 20 mM glacial acetic acid, 1 mM EDTA [pH 8]). Gels were stained with ethidium bromide (1 µg/mL) for 10 – 30 min and destained in 1× TAE buffer or H_2O for 30 – 120 min before visualisation on a UV transilluminator. Gel images were recorded using a BioRad gel documentation system. The 1 kb+ ladder (Invitrogen) was routinely run on gels as a size standard.

2.6.4.2 Pulsed-Field Gel Electrophoresis

Large DNA fragments were separated using contour-clamped homogeneous electric field (CHEF) electrophoresis in an agarose gel (Biorad, pulsed field grade) using a CHEF-DR® III electrophoresis unit at 14°C in 0.5× Tris-borate (TBE) buffer (44.5 mM Tris, 44.5 mM boric acid, 1.25 mM EDTA [pH 8.0]). To check the quality and concentration of DNA embedded in agarose plugs, chromosomes were separated on a 0.6% gel at 6 V/cm using a switch time of 50 s for 16 h. Large fragments were separated in a 1% gel at 6 V/cm using switch times of 50 – 90 s for 22 h.

2.6.5 Southern Analysis

2.6.5.1 Capillary Blotting

Following electrophoresis, gels were depurinated by soaking in 0.2 N HCl for 15 min, or until the bromophenol blue (a component of the tracking dye) turned yellow, then rinsed twice in distilled H_2O. To denature the DNA, gels were soaked twice in alkaline transfer buffer (1 M NaCl, 0.4 N NaOH) for 15 min per wash. Treated gels were then placed on top of 3 pieces of Whatman 3 MM paper (cut larger than the gel) soaked in alkaline transfer buffer, with overhanging wicks into

a reservoir of transfer buffer. A piece of Hybond N+ membrane (Amersham), cut to the exact size of the gel was pre-soaked in distilled water for 5 min, then transfer buffer for 5 min, then placed on top of the gel. The wells of the gel and any areas of the gel not covered by the membrane were removed and any air bubbles between the gel and the membrane were also removed. Three pieces of Whatmann 3MM cut to the exact size of the gel and soaked in transfer buffer were then placed on top of the membrane. A stack of paper towels (approx. 8 cm high) cut to the same size as the gel were then placed on top of the 3MM sheets and a weight added to the top of the stack. Parafilm strips were placed around the edges of the gel to prevent transfer buffer wicking into the paper towel stack and bypassing the gel and membrane.

After the DNA had been allowed to transfer for over 18 h, the membrane was soaked in neutralisation solution (0.5 M Tris [pH 7.5], 1 M NaCl) for 15 min and rinsed in 2× SSC (300 mM NaCl, 1.5 mM sodium citrate [pH 7]). Blotted gels were routinely re-stained in 1 μg/mL ethidium bromide and examined on a UV transilluminator to gauge the success of transfer.

2.6.5.2 Radioactive Hybridisation

Membrane prehybridisation was carried out at 65°C in a roller bottle containing 10 – 20 mL of modified Church and Gilbert prehybridisation solution (7% (w/v) SDS, 0.36 M Na_2HPO_4, 0.05 M NaH_2PO_4, 10 mM EDTA [pH 7.2] (Sambrook et $al.$, 1989)). Bottles were incubated in a rotating Hybaid hybridisation oven (Thermo Electron Corp.) for 40 – 60 min. Probe DNA was labelled with [α-^{32}P]dCTP or [γ-^{32}P)dATP (3000 Ci/mmol, Amersham) by primed synthesis with the Klenow fragment using a RadPrime Kit (Invitrogen) or end-labelling using T4 polynucleotide kinase (Roche). Following labelling, unincorporated isotope was removed using a Quant G50 Micro Column (Amersham). The probe was then denatured by boiling for 5 min and added to the bottle containing the prehybridised blot. Hybridisation bottles were hybridised at 65°C, or 60°C for lower stringency, for at least 18 h.

Following hybridisation, membranes were either washed three times at 50°C in 2× SSC, 0.1% (w/v) SDS for a total of 1 – 2 h or, for lower stringency washes, twice at room temperature for 15 min then once at 50°C for 15 min in 2× SSC, 0.1% (w/v) SDS. After washing, membranes were sealed in plastic and exposed to X-ray film (BioMax MS, Kodak) for 4 – 96 h at -80°C in the presence of X-Omatic (Kodak) intensifying screens. Films were developed using a 100 Plus X-ray processor (All-Pro Imaging).

Probes were stripped from membranes using boiling 0.5% (w/v) SDS. Boiling SDS solution was poured over the membrane and left shaking for 30 min, then the procedure was repeated. Stripped membranes were rinsed in 2× SSC, air-dried and then stored at room temperature.

2.6.6 Subcloning

2.6.6.1 Restriction Fragments
Vector DNA digested with appropriate enzyme or enzymes was dephosphorylated by adding 0.1 U of calf alkaline phosphatase (Invitrogen) and incubating at 37°C for 30 min followed by gel purification (section 2.6.3).

To ligate DNA fragments, 10 – 20 ng vector was mixed 1:2 or 1:3 vector:insert, or 1:2:2 for three-way ligations, with 4 μL 5× ligation buffer and 1 U of T4 DNA ligase (Invitrogen) in a total volume of 20 μL and incubated overnight at 16°C. Ligated DNA was then ethanol precipitated (section 2.6.3) and resuspended in 5 μL H_2O.

2.6.6.2 PCR Products
PCR products with T-overhangs were routinely cloned into pCR2.1-TOPO (Invitrogen) and blunt-ended PCR products were cloned into pCR-BluntII-TOPO (Invitrogen) using the manufacturer's instuctions.

2.6.6.3 Gateway® Cloning
Plasmids containing three separate inserts were constructed using the Gateway® Multisite (Invitrogen) system. Entry clones were constructed by amplifying three separate DNA fragments by PCR using appropriate *att*B-tailed primers as

prescribed by the manufacturer. These were then recombined into donor vectors in 3 separate reactions containing 20 – 50 fmol of *att*B PCR product, 150 ng of either pDONR P4-P1R, pDONR P2R-P3 and pDONR 221 as appropriate, 2 μL BP Clonase™ II (Invitrogen), which contains λ integrase (Int) and *E. coli* host integration factor (IHF) proteins and made up to 10 μL with TE buffer (pH 8). Reactions were mixed by vortexing and incubated at 25°C for 1 h followed by addition of 1 μL of proteinase K (2 μg/mL, Invitrogen) and incubation at 37°C for 10 min. 1 μL of the reaction mix was then used directly in *E. coli* transformation (section 2.6.6.4).

Entry clones obtained by BP recombination were used in a three-way recombination reaction containing 20 – 25 fmol of each entry clone, 60 ng of pDEST R4-R3, 4 μL LR Clonase™ Plus buffer, 4 μL LR Clonase™ Plus (Invitrogen), which contains λ Int and excisionase (Xis) and *E. coli* IHF proteins, and TE buffer up to 20 μL. Reactions were mixed by vortexing and incubated at 25°C for 16 h followed by addition of 2 μL of proteinase K (2 μg/mL, Invitrogen) and incubation at 37°C for 10 min. 1 μL of the reaction mix was then used directly in *E. coli* transformation (section 2.6.6.4).

2.6.6.4 Transformation

Either electrocompetent or chemically competent *E. coli* Top10 (Invitrogen) cells were transformed with plasmid DNA. Electrocompetent cells were thawed on ice, mixed with 1 – 2 μL of plasmid DNA and incubated on ice for 1 min. Cells were then transferred to a pre-chilled electroporation cuvette (0.1 cm, BioRad) and electrotransformed at 2500 V using a Gene Pulser (BioRad). Cells were then immediately resuspended in 250 μL of SOC media and incubated at 37°C, shaking at 200 rpm, for 40 – 60 min. Aliquots were then plated onto selective LB media.

Chemically competent cells were thawed on ice, mixed with 1 – 5 μL of plasmid DNA and incubated on ice for 30 min. Cells were heat-shocked at 42°C for 30 s and placed back on ice then resuspended in 250 μL of SOC media and incubated at

37°C, shaking at 200 rpm, for 40 – 60 min. Aliquots were then plated onto selective LB media.

2.7 Library Screening

N. lolii Lp19 λZAPII (prepared by Y. Itoh (Fleetwood *et al.*, 2007)), λGEM-12 (Dobson, 1997) and BAC (Johnson *et al.*, 2007b) libraries, as well as an *E. festucae* Fl1 fosmid genomic library (Johnson *et al.*, 2007a), were screened during the course of this thesis. The *N. lolii* Lp19 BAC library was ordered on nylon membranes and was screened using radioactive Southern hybridisation as described in section 2.6.5.2. The *E. festucae* Fl1 fosmid library was ordered in a pooled 96 well plate and was screened by PCR.

2.7.1 λ Library Plating and Blotting

E. coli host cultures were grown overnight in LB broth supplemented with 10 mM $MgSO_4$ and 0.2% (w/v) maltose. *E. coli* KW251 was used for the λGEM-12 library and *E. coli* XL1 Blue was used for the λZAPII library. Phage were diluted to a concentration of 3.0×10^6 PFU/mL in SM buffer and 3 μL aliquots were added to 200 μL aliquots of freshly grown host culture and incubated at 37°C for 15 – 20 min. The mixed *E. coli* and phage were then added to 3 mL aliquots of molten (50°C) LB agar, vortexed to mix, and poured onto LB agar plates and allowed to set. Plates were then incubated for 6 – 8 h until plaques were visible.

Recombinant phage DNA was blotted onto nylon membranes (Hybond N+, Amersham) by placing the membranes directly onto plates with forceps. Membranes were marked asymmetrically with a needle and the position of these was marked on the Petri plates for later alignment. Membranes were then placed DNA side up onto 250 μL of denaturing solution (0.4 N NaOH, 1.5 M NaCl), aliquoted onto cling-film spread on the laboratory bench, for 1 – 5 min. Following denaturation, membranes were transferred DNA side up onto 250 μL aliquots of neutralisation solution (0.5 M Tris [pH 7.2], 1 M NaCl) for 5 min, then rinsed in 2× SSC and air-dried on Whatmann 3MM paper.

2.7.2 DIG Hybridisation of Filters and Identification of Positive Clones

Hybridisation was performed using digoxygenin (DIG)-labelled probes and antibody detection. Filters were prehybridised in deep Petri dishes with 20 mL of Easy Hyb (Roche) solution per 10 filters for 30 min at 42°C with gentle agitation. Probes were labelled by PCR amplification with a DIG-labelling kit (Roche) using the manufacturers instructions, denatured by boiling for 3 min and added to the Petri dish. Hybridisations were then performed overnight at 42°C with gentle agitation.

Following hybridisation, membranes were washed 2 x 5 min in 2x SSC, 0.1% [w/v] SDS at room temperature and 2 x 15 min in 0.5X SSC, 0.1% [w/v] SDS at 65°C, then rinsed briefly in washing buffer (0.1 M Maleic Acid, 0.15 M NaCl, 0.3% [v/v] Tween 20). Membranes were then incubated for 30 min at 30°C in Blocking Reagent (Roche) then for a further 30 min in 20 mL of blocking solution with 1 μL of anti DIG-AP antibody (Roche). Membranes were then washed 3 x 10 min in 100 mL of washing buffer and equilibrated in 20 mL of detection buffer (0.1 M Tris, 0.1 M NaCl [pH 9.5]). Membranes were placed in plastic bags and 2 mL of 1x CDP *Star* (Roche, diluted in detection buffer) per 10 membranes was added and evenly spread over the membranes, these were incubated at room temperature for 5 min, then the excess liquid removed and the bag sealed. Finally bags were incubated at 37°C for 10 min and exposed to X-ray film (Biomax XAR, Kodak) for 15 – 25 min at room temp and developed using a 100 Plus X-ray processor (All-Pro Imaging).

Phage clones of interest were identified by aligning the hybridising spots with plaques using the marks on the membranes and the plates. These were excised using a cut-off P200 plastic pipette tip and placed into 500 μL of SM buffer with 50 μL of chloroform. To obtain pure clones, the screening process was performed again on a single plate for each clone with an appropriate amount of lysate to obtain ~ 20 plaques/plate.

2.7.3 Phagemid Excision

In order to excise clones from λ ZAPII phage vectors as pBluescript phagemids, the ExAssist interference-resistant helper phage (Stratagene) was used. XL1-Blue and SOLR *E. coli* strains were grown overnight in LB broth with 0.2% (w/v) maltose and 10 mM $MgSO_4$ at 30°C, pelleted by centrifugation, and resuspended in 10 mM $MgSO_4$ to an OD_{600} of 1.0. In a 15 mL polypropylene tube, 200 μL of XLI-Blue cells, 250 μL of phage lysate from a λ ZAPII clone of interest ($>1 \times 10^5$ pfu/mL) and 1 μL of ExAssist phage ($>1 \times 10^6$ pfu/mL) were mixed and incubated at 37°C for 15 min. 3 mL of LB broth was then added and incubated at 37°C for a further 3 – 18 h with shaking at 200 rpm. The tubes were then incubated at 70°C for 20 min and centrifuged at 1000 g for 15 min. Aliquots of the supernatant, 1, 10 or 100 μL, were mixed with 200 μL of SOLR cells and incubated at 37°C for 15 min. Serial dilutions of the cell mixture were then plated on LB plates with ampicillin selection at 37°C overnight; colonies that grew on ampicillin contained excised phagemids.

2.8 Polymerase Chain Reaction (PCR)

PCRs were performed using iCycler (BioRad) or 2720 (Applied Biosystems) thermocyclers. Oligonucleotide primers used in this study were synthesised by Invitrogen or Sigma Genosys and are described in Table 2.2.

2.8.1 Standard PCR Conditions

Unless otherwise stated, PCRs were performed in 20 or 50 μL using 5 – 20 ng fungal genomic DNA or 50 – 100 ng of epichloë-infected ryegrass genomic DNA, 1× *Taq* polymerase buffer (Invitrogen), 50 μM each dNTP, 200 nM each primer and 0.5 U *Taq* polymerase (Invitrogen). Thermocycling conditions used were initial denaturation at 94°C for 2 – 3 min followed by 30 cycles of denaturation at 94°C for 30 s, annealing at a temperature determined as 5°C lower than the T_m of the primer with the lowest T_m for 30 s and extension at 72°C for 45 s/kb of product, then 1 cycle of 72° for 5 – 10 min depending on size of product.

Table 2.2. PCR primers used in this study

Primer name	Sequence[a]	Gene/region	Use
DamP10	CCTTATAGAACACCTGCGCC	*lpsB*	probe
DamP11	GTTTGATCCACCGCATCAGG	*lpsB*	probe
DamP15	GGTGCTATCCGCCTCCCATG	*lpsB*	probe
DamP16	CATTGTATGCTGCCTTGACC	*lpsB*	probe
DamP21	TATCCACAGGAACTTCAACG	*easG*	probe
DamP22	TTGTTCAGTCTATCTTCTGC	*easG*	probe
DamP44	CGATTCCCGATGGTACAGGCTAC	*lpsB*	RACE
DamP45	CACAGGAATGAGCACGAAGAGGC	*lpsB*	RACE
DamP46	TCCGACGGTTGTCTCGACTCTAG	*easE*	RACE
DamP47	GTCGAGACAACCGTCGGATCGTC	*easE*	RACE
DamP62	CGTCAGGGAACGCACGAACAGTC	*easE*	RACE
DamP63	TTTGTGACTTGTCGAGATCGCTC	*lpsB*	RACE
DamP81	CGTGTTCACGAAGCTCAAGG	*easH*	PCR/RT-PCR
DamP82	CCAGCGGTGTCATAGTCTCG	*easH*	PCR/RT-PCR
DamP83	TGCCCACATGGTCTTTACCG	*easA*	PCR/RT-PCR
DamP84	CGTATCCAGCAGTGCCGTTC	*easA*	PCR/RT-PCR
DamP85	CTGCCATGGGACAGGTTCAC	*easG*	PCR/RT-PCR
DamP86	TCCGTTAATGTGCGGAATGC	*easG*	PCR/RT-PCR
DamP87	TGCGTTGGGACTTTGCAACT	*easF*	PCR/RT-PCR
DamP88	CGTAAGTGCCTCGCAGCTTG	*easF*	PCR/RT-PCR
DamP89	GACAACCGTCGGATCGTCAG	*easE*	PCR/RT-PCR
DamP90	CCGTTGGTATCGTTGGTGGA	*easE*	PCR/RT-PCR
DamP91	TTATCGTCCGAGGCAGCGTA	*lpsB*	PCR
DamP92	CGTTCACCGCAGAAGCAGTT	*lpsB*	PCR
DamP106	GCCCATCGGTCCTTCACGAT	*cloA*	PCR
DamP107	CCTCGACAGTGCATTGGGAC	*cloA*	PCR
DamP109	GAACACTGTGACCTTCATGG	*easD*	PCR
DamP110	GTTGACGATACTGCGGTTGC	*easD*	PCR
DamP111	CAGCATAGGGTCTTCTGAGG	*easC*	PCR
DamP112	GGATACCCTACTTGGACAGC	*easC*	PCR
DamP117	GCATGAGGGTTACTTATTGC	*dmaW*	PCR
DamP118	TCATATCACGACTAAGGAGATA	*dmaW* 5'	PCR
dmaWq-F	GTCCGTCGCAACTGGTAAAT	*dmaW*	real-time RT-PCR
dmaWq-R	TTCAAAGGCAGGACGGATAC	*dmaW*	real-time RT-PCR
easCq-F	GTCCATTGGTTCCTGTCATCC	*easC*	real-time RT-PCR
easCq-R	CTCGTATAGATCCCGTGTTGC	*easC*	real-time RT-PCR
easDq-F	TGTGTCTCGCCAGGAGTTAC	*easD*	real-time RT-PCR
easDq-R	CAGGTTCAACAATCGAGAAGC	*easD*	real-time RT-PCR
easEq-F	GATCGCCAAGCAGGATGTAT	*easE*	real-time RT-PCR
easEq-R	CTACCTATCGGCCCAAATGA	*easE*	real-time RT-PCR
easFq-F	TAGCTCCATTCGCTCCTCTT	*easF*	real-time RT-PCR
easFq-R	AAACTACTTCCGCCACTGGA	*easF*	real-time RT-PCR
easGq-F	GTGGTATGCCAGCTGAGGAC	*easG*	real-time RT-PCR
easGq-R	CCTGTTACATGCTTGACCACA	*easG*	real-time RT-PCR
Hyg-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGT TTATATCTTGCATCGTCCC	pAN71	replacement construct PCR
Hyg-R	GGGGACCACTTTGTACAAGAAAGCTGGGTAACG ACGGCCAGTGCCAAGC	pAN71	replacement construct PCR
KOLEnd-F	GGGGACTGCTTTTTTGTACAAACTTGTCTTCATA GCTGAGTTCACC	*lpsB*	replacement construct PCR
KOLEnd-R	GGGGACAACTTTGTATAGAAAAGTTGATTCCAG TACATTGTCCACG	*lpsB*	replacement construct PCR
KOprobe-F	AGTTATCAGCCGTCAACAGG	*lpsB*	probe
KOprobe-R	CATTGTATGCTGCCTTGACC	*lpsB*	probe
KOREnd-F	GGGGACAACTTTGTATAATAAAGTTGCGCTTGTT GCTTATCCTAGC	*lpsB*	replacement construct PCR
KOREnd-R	GGGGACAGCTTTCTTGTACAAAGTGGGAGATGA TAAAGCGGACTGG	*lpsB*	replacement construct PCR

Primer name	Sequence	Gene/region	Use
laeA-attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTTC GCCGCTGATGTTTGAG	*laeA*	expression construct PCR
laeA-attB2rev	GGGGACCACTTTGTACAAGAAAGCTGGGTGATG TGGTCGCCTTTTGCTC	*laeA*	expression construct PCR
laeA-F	TCAAGGCAGAACTCGGATGC	*laeA*	real-time RT-PCR
laeA-R	CTGTTCTTGCTCATCGCACG	*laeA*	real-time RT-PCR
lpsA-F	GCCCTGTGTCGTGAACCCAT	*lpsA*	probe
lpsA-R	CATCCACGGACATGCTTGGC	*lpsA*	probe
lpsBKOs-F	AACTGCTTCTGCGGTGAACG	*lpsB*	PCR
lpsBKOs-R	GTGTCCGTGTTTGAGTTTGC	*lpsB*	PCR
lpsBq-F	TGGCTACGATGATCAATACGA	*lpsB*	RT-PCR/real-time RT-PCR
lpsBq-R	ACACCGCTCGTTCAAATGTA	*lpsB*	RT-PCR/real-time RT-PCR
ltmGq-F	TCATTGCCCAGAGAGAGCTT	*ltmG*	real-time RT-PCR
ltmGq-R	CTCCATACCCTGACCACGAT	*ltmG*	real-time RT-PCR
p12-F	CCGTCTTCCCGTATACCGAA	*lpsB*	probe
p12-R	TACCCACTGCCTCGAACTTG	*lpsB*	probe
tub2q-F	ATGGTTCCTTTCCCTCGTCT	*tubB*	RT-PCR/real-time RT-PCR
tub2q-R	AGCTCAGGAACGCTAACAGC	*tubB*	RT-PCR/real-time RT-PCR

ᵃall primer sequences are written 5' to 3'

2.8.2 Colony PCR

Colony PCR was used to rapidly screen *E. coli* colonies for clones of interest. Colonies were picked into a standard 20 µL PCR reaction mix and amplified using standard thermocycling conditions, but with an initial denaturation at 95°C for 5 min.

2.8.3 Inverse PCR

Purified restriction digests of genomic DNA (200 ng) were self-ligated in 50-µL reactions with 1× T4 DNA ligase buffer and 5 U of T4 DNA ligase at 16°C overnight. Ligations were ethanol precipitated and resuspended in 20 µL H$_2$O and 4 µL used as template in a standard PCR reaction with primers designed facing outward from each other in the known genomic sequence. PCR products were then cloned and sequenced.

2.8.4 Long Template and High Fidelity PCR

Long template PCR was performed with the Triplemaster polymerase mix (Eppendorf) and tuning buffer using conditions recommended by the manufacturer. High fidelity PCR was performed using Pfx50™ DNA polymerase (Invitrogen) and conditions recommended by the manufacturer.

2.9 RNA Isolation and Analysis

2.9.1 RNA Isolation

RNA was isolated from mycelia and infected ryegrass pseudostems using TRIzol®
Reagent (Invitrogen). All apparatus used for RNA isolation was either autoclaved
for 45 min or cleaned thoroughly and wiped with RNAse AWAY (Invitrogen) as
appropriate. 50 – 100 mg of sample was ground to a fine powder under liquid
nitrogen and mixed with 1 mL of TRIzol®, followed by centrifugation at 12 000 g for
10 min at 4°C. The supernatant was incubated at room temperature for 5 min, 200
μL of chloroform was added and shaken vigorously for 15 s followed by a further 2
– 3 min incubation at room temperature and centrifugation at 12 000 g for 15 min at
4°C. To precipitate RNA while leaving polysaccharide in solution, the aqueous
phase was mixed with 250 μL of isopropanol followed by 250 μL of high salt
precipitation solution (0.8 M sodium citrate, 1.2 M NaCl), incubated at room
temperature for 10 min and centrifuged at 12 000 g for 10 min at 4°C. The RNA
pellet was washed with 75% [v/v] ethanol and centrifuged at 7500 g for 5 min at
4°C. The pellet was then air-dried for 5 – 10 min and resuspended in 50 μL of
RNase-free H_2O.

DNA was removed from RNA preparations by incubating 1 – 5-μg aliquots of total
RNA with 10 U of DNaseI (Roche) for 30 min at 37°C, followed by 10 min at 75°C.

2.9.2 cDNA Preparation

RNA was reverse transcribed at 60°C for 60 min in a reaction volume of 20 μl
containing 1× reaction buffer, 10 mM DTT, 1 mM each dNTP, 2.5 μM oligo(dT)$_{20}$,
2 U of RNAse inhibitor and 0.75 U of ThermoScript™ reverse transcriptase
(Invitrogen). Gene-specific amplification from cDNA was carried out using 1 μL of
cDNA (neat or diluted as described in text) as template, other conditions were as
described in section 2.8.1.

2.9.3 Real Time RT-PCR

Real-time PCR was performed using iQ™ SYBR® Green Supermix (BioRad) in 20 μl
reactions in a MyiQ Single-Color Real-Time PCR Detection System (BioRad).

Samples were prepared in duplicate for each condition and PCRs from each sample were performed in triplicate and the threshold (Ct) values averaged. To compare relative transcript abundance, the average cycling threshold for each experimental condition was normalised to *tubB* using the calculation $2^{-(\Delta Ctsample - \Delta CttubB)}$. Melt curve analysis and/or gel electrophoresis was used to determine the presence of non-specific amplification products or primer dimers.

2.9.4 Rapid Amplification of cDNA Ends (RACE)

In order to determine the 5' and 3' ends of cDNAs, RACE was performed with total RNA isolated from infected ryegrass pseudostems using the GeneRacer Kit (Invitrogen) and instructions supplied by the manufacturer. Gene-specific RACE PCR products were gel purified and cloned as described in section 2.6.6.2 and sequenced as described below.

2.10 DNA Sequencing

Plasmid, PCR products and λ DNA were sequenced using the dye terminator method with M13 forward and reverse, SP6, T7 and gene specific oligonucleotide primers. Dye-terminator reactions were carried out in 20 μL reactions with either 200 – 500 ng of double-stranded plasmid or λ DNA, or ~1 ng/100 bp of PCR-product, 3.2 pmol primer, 0.875× ABI dilution buffer (Applied Biosystems) and 1.5 μL Big-Dye (Applied Biosystems). Reactions were cycled in an iCycler (BioRad) thermocycler at 96°C for 3 min then 30 cycles of 96°C for 10 s, 50°C for 10 s and 60°C for 4 min. Separation and detection were carried out on ABI 377 and 3730 automated sequencers.

2.11 Bioinformatics

DNA sequences were edited and assembled into contigs using ContigExpress (Invitrogen). Sequence was analysed and annotated using Vector NTI suite 9 (Invitrogen). Open reading frames (ORFs) likely to encode genes were identified in Vector NTI and by genomic sequence comparison with public databases. Sequence comparisons were performed over the Internet, with GenBank databases at the National Centre for Biotechnology Information (NCBI) website

(http://www.ncbi.nlm.nih.gov/), and with fungal genome sequence databases at the Broad Institute website (http://www.broad.mit.edu/), using the BLASTN, BLASTX and BLASTP algorithms (Altschul *et al.*, 1990; Altschul *et al.*, 1997). Introns in ORFs were predicted as regions flanked by consensus splice sites (5' GT, 3' AG) that introduced frame-shifts, premature stop codons or gaps in the alignment of a translated genomic sequence with known protein sequences, the removal of which lead to a full-length polypeptide alignment.

Protein sequences were analysed using InterproScan at the European Bioinformatics Institute (EBI) website (Zdobnov and Apweiler, 2001) (http://www.ebi.ac.uk/InterProScan/) and the Conserved Domain Database (CDD) at NCBI (Marchler-Bauer *et al.*, 2005).

Extraction and alignment of the LpsB A-domain sequence and determination of 10-aa code were performed at the NRPSpredictor website (http://www-ab.informatik.uni-tuebingen.de/software/NRPSpredictor).

Alignment of nucleotide and amino acid sequences was performed in AlignX using the ClustalW algorithm (Higgins *et al.*, 1994).

E. festucae genome sequence, curated by C. Schardl at the University of Kentucky, and B. Roe at the University of Oklahoma, was analysed at the University of Oklahoma genome website (http://www.genome.ou.edu/fungi.html).

Upstream regions of genes were analysed for potential promoter elements using MEME (Multiple Em for Motif Elicitation) (Bailey and Elkan, 1994), the MATCH algorithm (Kel *et al.*, 2003) with the TRANSFAC database (Matys *et al.*, 2003) and various algorithms at the RSAT (Regulatory Sequence Analysis Tools) website (http://rsat.ulb.ac.be/rsat/).

Transposon and repeat sequences were identified and analysed using MEME, the etandem and einverted algorithms of EMBOSS (Rice *et al.*, 2000) and BLASTN with GenBank and fungal genome databases. Secondary structure was determined using MFOLD (Zuker, 2003) (http://frontend.bioinfo.rpi.edu/applications/mfold/).

2.12 Protoplasting and Transformation

2.12.1 Protoplast Preparation

Protoplasts of *E. festucae* Fl1 were prepared using the method of (Young *et al.*, (1998). Cultures were grown for three days in PD broth, filtered through sterile Whatman 3MM paper and washed with sterile H_2O followed by OM buffer (1.2 M $MgSO_4.7H_2O$, 10 mM Na_2HPO_4 [pH 5.8]). Washed mycelia were added to 15 mL of OM buffer with 10 mg/mL Glucanex (Chemcolour Industry) and incubated for ~3 h at 30°C to digest the cell walls. Protoplasts were filtered through nappy liners and overlaid with 2 mL ST buffer (0.6 M sorbitol, 100 mM Tris [pH 8.0]) followed by centrifugation at 1578 x g for 5 min, resulting in separation of two phases and concentration of protoplasts at the interface. Protoplasts were harvested using a p1000 pipette and cut-off tip, mixed with 5 mL of STC buffer (1 M sorbitol, 50 mM $CaCl_2$, 50 mM Tris [pH 8.0]) and pelleted by centrifugation at 2465 x g, followed by 2 further washes with 5 mL STC buffer. Protoplasts were finally resuspended to a concentration of 1.25×10^8/mL in STC buffer and used immediately for transformation or stored at -80°C (80 μl of protoplasts/20 μL of 40% PEG (40% PEG 4000, 50 mM $CaCl_2$, 1 M sorbitol, 50 mM Tris [pH 8.0])).

2.12.2 Fungal Transformation

Protoplasts were transformed with either 5 μg of a linear PCR product of pDF6 for targeted gene replacement of *lpsB*, or co-transformed with 5 μg of circular plasmid and 2.5 μg of either pII99 or pPN1688 for complementation or heterologous expression. DNA was added to 80 μL of protoplasts with 20 μL of 40% PEG

Transformants were selected on RG medium containing either hygromycin (150 μg/ml) for replacement constructs and *A. nidulans laeA* heterologous expression or geneticin (220 μg/ml) for complementation. To obtain clonal isolates the resulting transformants were purified by subculturing mycelia three times as described in Young *et al.* (2005).

2.13 Plant Growth Conditions and Inoculations

Perennial ryegrass seedlings were inoculated using the method of Latch and Christensen (1985) (Latch and Christensen, 1985). Endophyte-free perennial ryegrass cv. Nui seeds were surface sterilised by soaking in 50% H_2SO_4 for 30 min, rinsed 3 times in sterile H_2O, then soaked in 50% commercial bleach for 20 min, rinsed 3 times in sterile H_2O and dried on sterile Whatmann 3MM paper. Seeds were germinated on 4% water agar at 22°C in the dark to ensure etiolation. A very small piece of fungal mycelia was placed over a slit cut between the mesocotyl and coleoptile of each seedling using a sterile scalpel under a dissecting microscope. Inoculated seedlings were incubated at 22°C in the dark for 7 days followed by 7 days under lights. Seedlings were then transferred to root trainers with potting mix. Endophyte-infected plants were subsequently re-potted into individual bags and maintained in a glasshouse.

2.13.1 Detection of Endophyte Infection by Immunoblotting

Inoculated plants were tested for infection by tissue-print immunoblotting (Gwinn *et al.*, 1991) using polyclonal antibodies raised to *N. lolii* Lp5. Tillers were cut off at the base and blotted onto nitrocellulose membrane (BDH). Membranes were washed in blocking solution (20 mM Tris, 10 mM HCl, 50 mM NaCl, 0.5% non-fat milk powder [pH 7.5]) at room temperature for at least 2 hours then overnight at 4°C in anti-*N. lolii* antibody diluted 1/1000 in blocking solution. Membranes were rinsed in blocking solution and then incubated with anti-rabbit enzyme-conjugated secondary antibody (Sigma) diluted 1/2000 in blocking solution for two hours at room temperature. Finally, membranes were rinsed in blocking solution and developed in Fast Red chromogen (Sigma) for 15 min.

2.13.2 Microscopic Analysis with Aniline Blue

Endophyte growth *in planta* was examined by taking epidermal strips from the outermost leaf sheath of infected plants and staining with aniline blue (0.05% in lactic acid/glycerol/H_2O 1:2:1). Epidermal strips were placed onto a drop of stain on a microscope slide, covered with a coverslip and heated briefly until boiling. Stained hyphae were then examined with an Olympus BX50 light microscope.

2.14 Alkaloid Analysis

Pseudostems (2 – 3 per plant, approx. 35 mm in length) were collected from perennial ryegrass plants after at least 8 weeks growth in the greenhouse. Samples (approx. 50 mg) were extracted by maceration with a ceramic bead in a tissue disrupter (FastPrep FP120, Savant) for 60 s (setting 4) in 500 μl methanol – 1% aqueous acetic acid (1:1) (containing 0.112 μg/ml ergotamine tartrate as internal standard) and centrifuged at 15 000 × g for 2 min. Subsamples (15 μL) of the supernatant were eluted through a C18 Luna column (Phenomenex) (150 x 2 mm, 5 μm) at a flow rate of 200 μl/min using a Surveyor HPLC (Thermo Finnigan) with a solvent gradient, starting with 5% MeCN: 95% H_2O (containing 5 mM ammonium acetate) for 5 min and then increasing to 50% MeCN over 38 minutes followed by a column wash at 100 % MeCN. To limit carry-over between samples, 4 wash cycles were carried out between each injection. Cross-contamination between samples was very low but detectable (ca. 1/1000), and endophyte-free samples were interspersed between endophyte-infected samples to avoid ambiguity. Mass spectrometry was carried out by ESI in +ve mode with a linear ion trap mass spectrometer (LTQ, Thermo Finnigan). The spray voltage was 5.0 kV and the capillary temperature 275°C. The flow rates of sheath gas and auxiliary gas were set to 20 and 10 (arbitrary units), respectively. Metabolites were identified by comparison of elution times and product ion spectra from collision induced dissociation of selected precursor ions (35% collision energy) with authentic standards or, in the case of lysergyl-alanine and its isomer, and 6,7-secolysergine, with previously characterised metabolites of *N. sp.* Lp1 (Panaccione *et al.*, 2003). Measurements were carried out by selective reaction monitoring of the transitions *m/z* 269 → 223 (lysergic acid and isolysergic acid), *m/z* 340 → 223 (lysergyl-alanine and isolysergyl-alanine), *m/z* 268 → 223 (ergine and erginine), *m/z* 241 → 210 (6, 7-secolysergine and putative isomers), *m/z* 534 → 268 (ergovaline and ergovalinine) and *m/z* 582 → 268 (ergotamine and ergotaminine).

2.15 Insect Bioassay

The choice test bioassay was performed with 8-week-old ryegrass plants, either endophyte-free or infected with *E. festucae* Fl1 or DFM3. One plant of each treatment was planted in potting mix in a ~15 × 30 cm pot and blades cut back to ~12 cm from the base. Three adult black beetles (*Heteronychus arator*) were added to each pot and pots were sealed with clear plastic and mesh to prevent escape of the beetles and allow air flow. After 16 days the number of black beetle damaged and undamaged tillers were assessed. Logistic regression was used to model the number of beetle-damaged tillers out of the total number of tillers on each plant and the cage was accounted for in the model as a nuisance variable.

3. Results

Isolation and Characterisation of a Gene Cluster for Ergovaline Biosynthesis

3.1 Isolation and Analysis of the NRPS Gene PS12 and Upstream Sequence

PCR amplification of *N. lolii* NRPS adenylation domains in a previous study identified a clone, ps12, that was preferentially expressed *in planta* compared with culture (Tanaka *et al.*, 2005). To clone the full length gene the ps12 PCR product was used to probe genomic libraries. This region of the genome was found to be underrepresented in all libraries screened. The ps12 sequence was absent from three large insert libraries: *E. festucae* Fl1 cosmid (Tanaka *et al.*, 2005) and fosmid (Johnson *et al.*, 2007a) libraries, and an *N. lolii* Lp19 BAC library (Johnson *et al.* 2007b). The sequence was present, albeit underrepresented, in a *N. lolii* λZAPII small-insert library (Fleetwood *et al.* 2007) with just two clones (pDF1 and pDF2) isolated from ~80 000 recombinant phage plated.

pDF1 and pDF2 were sequenced using M13 forward and reverse, as well as custom primers and the sequences were assembled into a single contig that covered 9.8 kb (Fig. 3.1). Bioinformatics analysis revealed two open reading frames (ORFs) and a partial ORF encoding predicted proteins with 49%, 48% and 66% identity
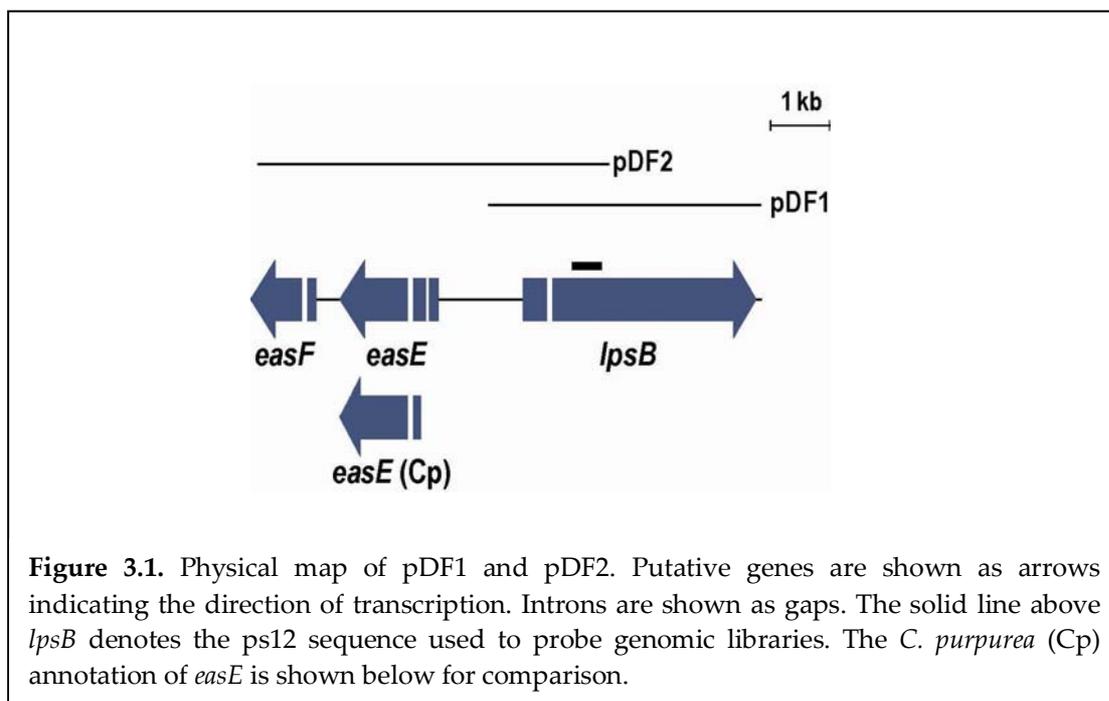
Figure 3.1. Physical map of pDF1 and pDF2. Putative genes are shown as arrows indicating the direction of transcription. Introns are shown as gaps. The solid line above *lpsB* denotes the ps12 sequence used to probe genomic libraries. The *C. purpurea* (Cp) annotation of *easE* is shown below for comparison.

respectively to genes for a single-module NRPS, *lpsB* (formerly *cpps2*), an oxidoreductase, *easE* (formerly *cpox1*), and a conserved hypothetical protein with no known function, *easF*, found in a gene cluster for ergopeptine ergot alkaloid biosynthesis in *C. purpurea* (Correia *et al*., 2003). The *easE* and *easF* genes are also found in the *Aspergillus fumigatus* clavine ergot alkaloid gene cluster (Coyle and Panaccione, 2006). These results suggested that the three putative genes, designated *lpsB*, *easE* and *easF* following the convention of Schardl *et al*. (2006), encode genes for the biosynthesis of ergovaline, the predominant ergopeptine ergot alkaloid product in *N. lolii*. The locus was thus labelled *EAS* (for ergot alkaloid synthesis). Analysis of *lpsB* and *easE* are described below, while analysis of *easF* is described in section 3.5.1, following isolation of the full-length gene.

3.1.1 The *lpsB* Gene

The predicted *lpsB* gene is 4130 bp in length and contains a 75-bp intron 594 bp from the translational start site, which was confirmed by cDNA sequencing and is conserved with the *C. purpurea* orthologue.

5' and 3' RACE analysis was performed using RNA extracted from perennial ryegrass infected with *N. lolii* Lp19. 5' RACE PCR products were amplified using DamP45 and GeneRacer 5' primers, these were then used as template in a nested PCR using DamP63 and the GeneRacer 5' Nested primers. The nested product was purified and cloned and 17 5' RACE clones were sequenced and shown to have nine different 5' UTR lengths. 5' UTRs represented multiple times were 44 (×2), 66 (×5) and 70 (×4) bp long. The longest UTR – 104 bp – was detected once.

3' RACE PCR products were amplified using DamP44 and GeneRacer 3' primers, these were then used as template in a semi-nested PCR using DamP44 and GeneRacer 3' Nested primers. The nested product was cloned and 14 3' RACE clones were sequenced and shown to have ten different 3' UTR lengths (Fig. 3.2). Interestingly, sequencing 3' RACE clones revealed a differentially spliced intron in the *lpsB* 3' UTR (Fig. 3.2). Of the three splice variants, the UTR with the intron

Figure 3.2. 3' RACE analysis of *lpsB*. An alignment of the sequences of 14 clones of nested PCR products from *N. lolii* Lp19 RACE cDNA (section 2.9.4) is shown, corresponding to nucleotide positions 33 – 129 downstream of the *lpsB* stop codon. The length of the 3' UTR is given based on the nucleotide position upstream of the *lpsB* stop codon, without any intron spliced.
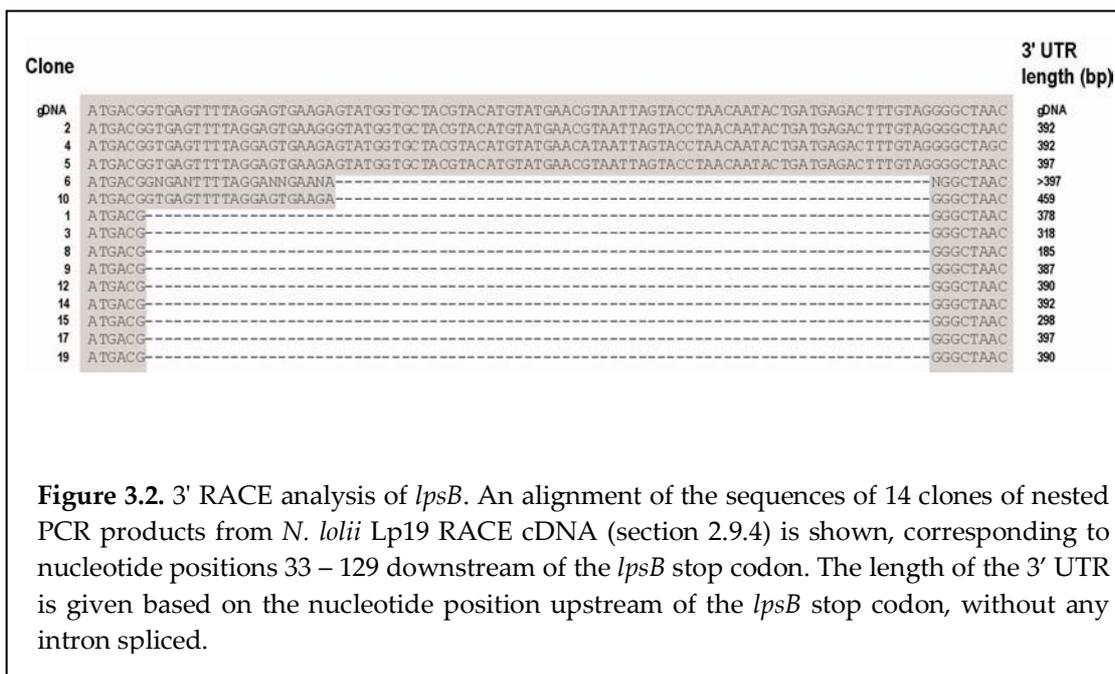
unspliced was detected three times and was polyadenylated at 392 (×2) and 397 bp downstream of the *lpsB* stop codon. Splice variant one (63-bp intron) was detected twice and one clone was polyadenylated at 459 bp downstream of the stop codon; the quality of sequence from the other clone was poor but the mRNA was polyadenylated at least 397 bp downstream of the stop codon. Splice variant three (83-bp intron) was detected nine times and had variable polyadenylation sites between 185 and 392 bp downstream of the stop codon (Fig. 3.2).

3.1.2 LpsB, a Single-Module NRPS

The predicted translated product, LpsB, is 1352 amino acids in length, a predicted unmodified molecular mass of 148 kD. Analysis of the predicted amino acid sequence by InterproScan revealed the single-module to contain an A-domain, T-domain and a C-domain (Fig. 3.3). A fourth uncharacterised domain of 301 amino acids is present at the amino terminal end of the protein. A query of this region against the Conserved Domain Database (CDD) returned a weakly significant match (E-value 0.00001) to the carboxy-end of C-domains but no conserved C-domain motifs were discernable in this region. Manual analysis showed that all ten sequence motifs conserved among A-domains are present in the predicted LpsB protein, as is the conserved T-domain motif (Table 3.1). Although C domain motifs

are typically not as conserved as those for the other domains, the C3, C5 and C6 motifs were detected (Table 3.1).

Figure 3.3. Domain structure of LpsB. A, A-domain; T, thiolation (PCP)-domain; C, condensation domain. Triangles indicate position of conserved motifs described in Table 3.1.
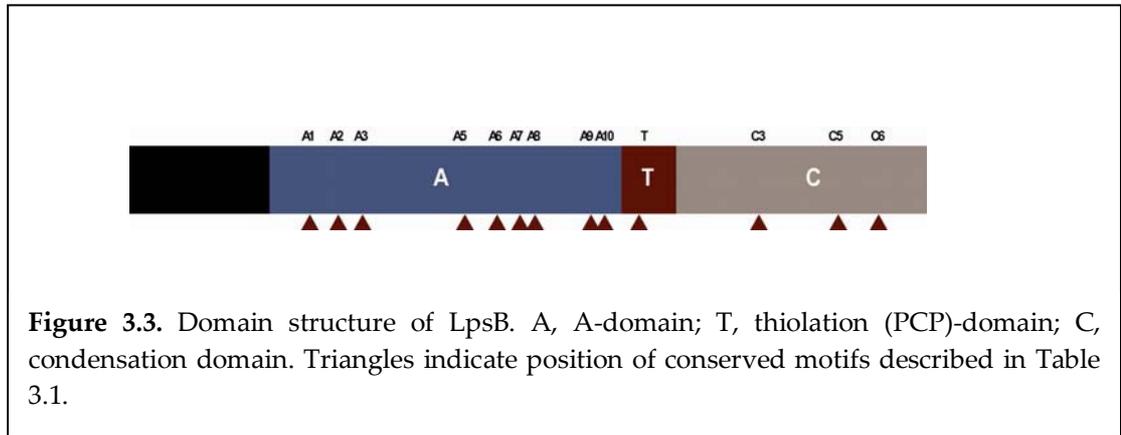
Table 3.1. Conserved motifs found in LpsB

Motif	LpsB	Consensus*
A1	LYSEEL	L(T/S)YxEL
A2	IKAGGAFVLLDP	LKAGxAYL(V/L)P(L/I)D
A3	LAEDLVTSIVEVGDDK	LAYxxYTSG(S/T)TGxPKG
A4[a]	FQFS, FDVT, FDTL	FDxS
A5	HLYGASE	NxYGPTE
A6	GELVIEGTIVGRGYL	GELxIxGxG(V/L)ARGYL
A7	YKSGDL	Y(R/K)TGDL
A8	GRKDTQVKLRGQRVELGEVE	GRxDxQVKIRGXRIELGEIE
A9	LPSYMVP	LPxYM(I/V)P
A10	NRKLLR	NGK(V/L)DR
T	DDHFFQRGGDSL	DxFFxxLGG(H/D)S(L/I)
C1	nd	SxAQxR(L/M)(W/Y)Xl
C2	nd	RHExLRTxF
C3	VHHAVYDGYT	MHHxISDG(W/V)S
C4	nd	YxD(F/Y)AVW
C5	TIPTVATIPCR	(I/V)GxFVNT(Q/L)(C/A)XR
C6	STDAWFE	(H/N)QD(Y/V)PFE
C7	nd	RDxSRNPL

*From Marahiel *et al*. (1997). [a]May be any of three different potential A4 motifs. nd, not detected

If the LpsB protein were involved in the synthesis of ergovaline, its substrate would be D-lysergic acid. Bioinformatics methods to predict A-domain substrates utilise the crystal structure of the GrsA PheA domain, for which the amino acids lining the binding site are known (Conti *et al.*, 1997). The sequence of the 10 amino acids lining the binding sites of A-domains constitutes a so-called A-domain specificity code (Stachelhaus *et al.*, 1999; Challis *et al.*, 2000). The LpsB A-domain sequence was aligned with that of PheA and the 10 amino acids likely to line the binding pocket were extracted at the NRPS Predictor website. Alignment of these amino acids with those in a database of NRPSs with known substrate was performed and showed that 9 out of 10 of the *N. lolii* LpsB amino acids matched those found in the *C. purpurea* LpsB binding pocket (Table 3.2). No other significant matches were identified using this software. As lysergic acid contains an indole ring from tryptophan the binding-pocket amino acids from tryptophan-activating A-domains from characterised bacterial NRPSs CdaI, CdaIII (Hojati *et al.*, 2002) and ComA (Chiu *et al.*, 2001) (Table 3.2) were manually aligned with those of LpsB. This showed a similar level of identity between the LpsB sequence and tryptophan-activating A-domain sequences as between the three tryptophan-domain sequences themselves. Sequence from an alanine-activating A-domain from HC toxin synthetase (Scott-Craig *et al.*, 1992) showed no such similarity (Table 3.2). This may suggest that the indole ring is the moiety of the lysergyl substrate that is recognised by the LpsB enzyme.

Table 3.2. Substrate-specificity determining amino acids

Protein[†]	position in A-domain*										Substrate
	235	236	239	278	299	301	322	330	331	517	
NlLpsB	D	V	F	S	V	G	L	I	M	K	
CpLpsB	D	V	F	S	V	G	L	V	M	K	LSA
CdaI (III)	D	A	W	S	V	G	L	T	T	K	Trp
CdaIII (II)	D	G	W	A	V	A	R	T	T	K	Trp
ComA (II)	D	V	A	V	V	G	E	V	V	K	Trp
HTS1 (II)	D	A	G	G	C	A	M	V	A	K	Ala

*based on the sequence of the phenylalanine-activatine domain of GrsA (Conti *et al.*, 1997).
†sources cited in text. LSA = lysergic acid, Trp = tryptophan, Ala = alanine

3.1.2.1 Analysis of Potential COM Domains in LpsB and LpsA

As the LpsB NRPS is predicted to interact with a second NRPS, LpsA, in order to catalyse formation of lysergyl peptide lactam, the *N. lolii* and *C. purpurea* LpsB carboxy terminal sequences were aligned with those of bacterial NRPSs involved in multi-enzyme complexes, which contain short COM domains required for protein-protein interaction (Fig. 3.4). No significant alignment was found with these domains although proline and leucine residues are found in conserved positions between LpsB proteins and bacterial COM domains.

Figure 3.4. Multiple sequence alignments with representative bacterial COM domains.
A. Alignment of *N. lolii* (Nl) and *C. purpurea* (Cp) LpsB carboxy terminal amino acid sequences with bacterial donor COM domains. The TPSD motif at the junction between bacterial epimerisation and COM domains is underlined. Invariant residues in bacterial domains are marked with an asterisk.
B. Alignment of *N. lolii* and *C. purpurea* LpsA amino-terminal sequences with bacterial acceptor COM domains. The PLS motif at the junction between bacterial condensation and COM domains is underlined. An invariant leucine residue in bacterial domains is marked with an asterisk.
Bb, *Brevibacillus brevis*; Bl, *Bacillus licheniformis*; Sp, *Streptomyces pristinaespiralis.* Accessions: GrsA, CAA33603; GrsB, CAA43838; BacB, AAC06347; BacC, AAC06348; SnbC, CAA72311; SnbDE, CAA72312

The amino termini of the LpsA protein sequence from *N. lolii* and the LpsA-1 and LpsA-2 protein sequences from *C. purpurea* were aligned with acceptor COM domains from bacterial systems (Fig. 3.4). This analysis showed no significant alignment between the three LpsB-interacting proteins and COM domains. Interestingly, the three LpsA homologues also had very little identity with each other.

3.1.3 The *easE* Gene and Predicted Oxidoreductase EasE

The predicted *easE* gene is 1951 bp in length and contains two introns. The gene was originally annotated based on the *C. purpurea* annotation as 1610 bp in length with one conserved 59-bp intron. However, 5'RACE analysis (see below) indicated that the coding sequence is likely to start 341 bp further upstream and include a second 74-bp intron (Fig. 3.1). The sequence upstream of the *C. purpurea easE* gene is not deposited in GenBank so it was not possible to determine whether the shorter *easE* gene in *C. purpurea* was likely to be a mis-annotation.

5' and 3' RACE analysis was performed using RNA extracted from perennial ryegrass infected with *N. lolii* Lp19. 5' RACE PCR products were amplified using DamP47 and GeneRacer 5' primers. These were then used as template in a nested PCR using DamP62 and the GeneRacer 5' Nested primers. The nested product was purified and cloned and six 5' RACE clones were sequenced and shown to have three different 5' UTR lengths of 19 (×4), 25 and 55 bp.

3' RACE PCR products were amplified using DamP46 and GeneRacer 3' primers. These were then used as templates in a semi-nested PCR using DamP46 and GeneRacer 3' Nested primers. The nested product was cloned and seven 3' RACE clones were sequenced showing six different 3' UTR lengths of 22, 32, 41, 48, 58 (×2) and 60 bp.

The predicted protein EasE is 605 aa long, a predicted unmodified molecular mass of 66 kD. Analysis of the predicted aa sequence using the CDD and InterProScan, revealed the presence of an FAD-binding domain between positions 49 and 179 (pfam01565.13), that shows 26% identity with the functionally characterised

domain from GlcD, the 6-hydroxy-D-nicotine oxidase, from *Arthrobacter oxidans* (Brandsch *et al.*, 1987). A berberine-like domain between positions 526 and 571 (pfam08031.1) was also detected that showed 22% identity to that of GlcD. Berberine domains are found in a variety of FAD-dependent oxidoreductases and were characterised in the berberine bridge enzyme, required for benzophenanthridine alkaloid metabolism in plants (Dittrich and Kutchan, 1991). The EasE protein is 54% and 49% identical respectively to proposed orthologues from *C. purpurea* and *A. fumigatus*. EasE may be involved in the conversion of chanoclavine to agroclavine, which is thought to be catalysed in part by an FAD-binding oxidoreductase (Floss, 1976).

3.2 Creating an *E. festucae lpsB* Mutant

To confirm that *lpsB* was required for ergovaline biosynthesis, deletion of a 276 bp region of the adenylation domain, removing the A2 and A3 sequence motifs, was performed. A replacement construct was prepared using Gateway recombination, and then transformed into *E. festucae* Fl1. *E. festucae*, the sexual progenitor of *N. lolii*, was used rather than *N. lolii* because of its genetic tractability.

3.2.1 Preparation of an *lpsB* Deletion Construct

The targeted deletion construct, pDF6, was constructed using a MultiSite Gateway system (Fig. 3.5). Flanking sequences of 3.1-kb, 5' and 3' of *lpsB*, were amplified by PCR with attB1- and attB4-tailed primers (KOLEnd-F/KOLEnd-R) and attB3- and attB2-tailed primers (KOREnd-F/KOREnd-R), respectively from *E. festucae* Fl1 genomic DNA. A third 4.1-kb fragment containing P*gpdA-hph* (encoding hygromycin resistance) was amplified from pAN7-1 (Punt *et al.*, 1987) with attB1- and attB2-tailed primers (Hyg-F/Hyg-R). These three fragments were then recombined, first into three separate entry vectors and then into a single destination vector (section 2.6.6.3) to yield pDF6. The orientation of each fragment was confirmed by restriction digests and sequencing across the junctions between the recombined fragments.

A linear product of pDF6, containing the *lpsB* flanking regions and *hph* cassette, was amplified by PCR using TripleMaster polymerase (section 2.8.4). This was then used to transform *E. festucae*.

3.2.2 Transformation and Screening of *E. festucae*

Five µg of the linear product of pDF6 was transformed into *E. festucae* Fl1 protoplasts, as described in section 2.13, and transformants selected on media containing hygromycin. After growth for one month, clonal isolates of hygromycin-resistant transformants were obtained by serial subculturing three times from the edge of the colony. This process leads to nuclear purification as epichloë compartments are mononucleate (Young *et al.*, 2005).
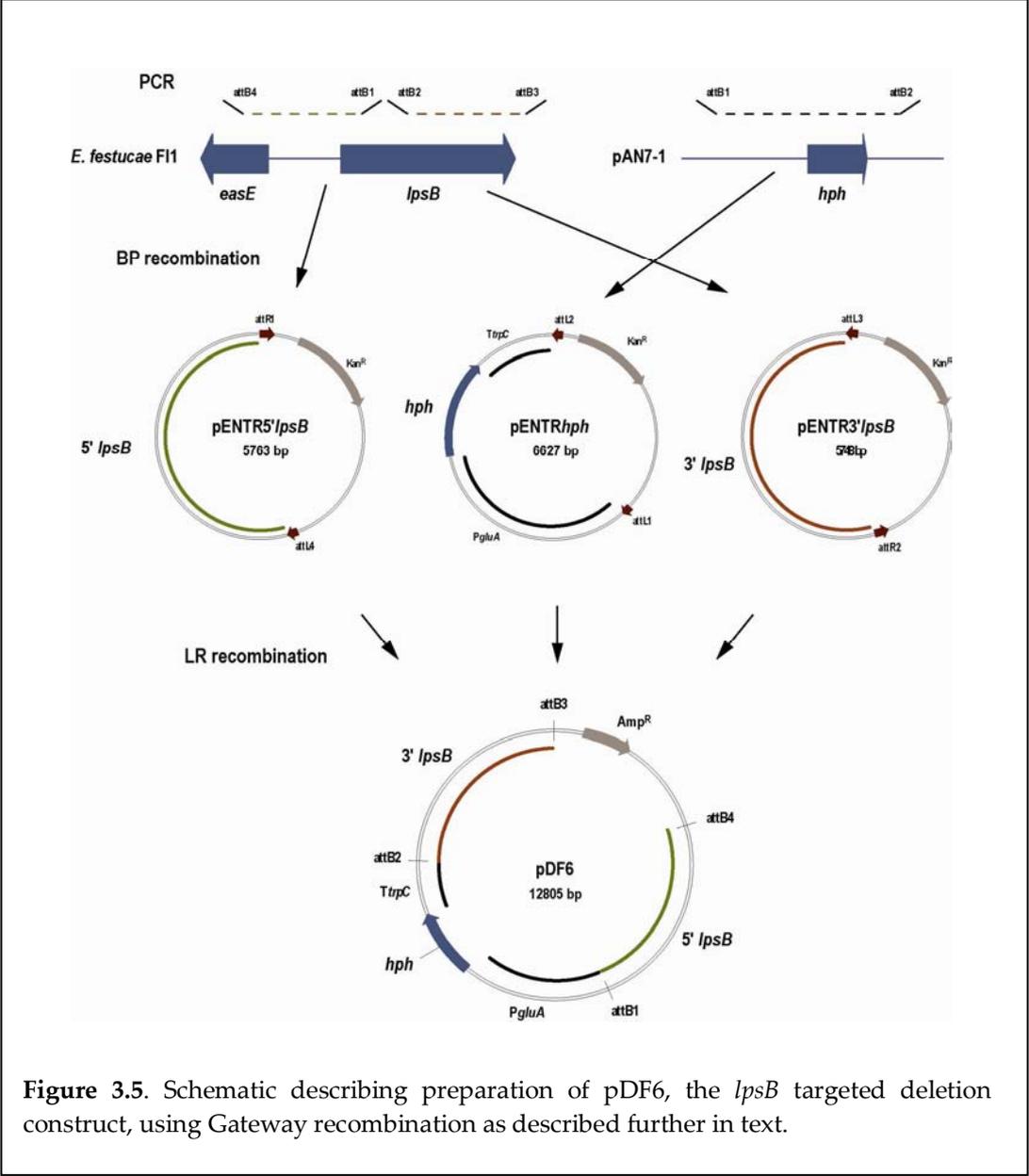
Figure 3.5. Schematic describing preparation of pDF6, the *lpsB* targeted deletion construct, using Gateway recombination as described further in text.

Transformants were initially screened for homologous recombination of the deletion construct by PCR. Genomic DNA was extracted from transformant mycelia by alkaline lysis (section 2.5.2) and used as template in a PCR with primers lpsBKOs-F and lpsBKOs-R, which flank the hygromycin resistance cassette. Of 48 transformants screened, seven (15%) amplified a mutant band, but not a smaller wild-type band, indicating a homologous replacement event. A wild-type band was amplified from the remaining isolates, indicating ectopic insertion of the replacement construct in the genome.

A selection of putative mutants were further screened by PCR to confirm that the sequence 5' and 3' of the integration was intact (Fig. 3.6), as occasionally large deletions can be caused by single-crossover events at the integration site (C. Young, personal communication). No external deletions were detected in any of the screened isolates.

Figure 3.6. Targeted gene replacement of *lpsB*
A. Diagrammatic representation of homologous recombination at *lpsB*. Wild type (WT) gene locus and targeted replacement construct (KO) are shown. Numbered pairs of arrows represent primers for PCR shown in **B**. The solid line above *lpsB* represents the PCR probe amplified with primers KOprobe-F and KOprobe-R used in **C**.
B. PCR screening of selected transformants with primer pairs shown in **A** (4 = DFM3). a = lpsBscreen-F/lpsBscreen-R, b = lpsBKOLEnds-F/lpsBKOLEnds-R, c = DamP10/DamP11
C. Southern blot analysis. Wild type (WT) and *lpsB* mutant DFM3 (KO) genomic DNA was digested with *Eco*RI and transferred to a nylon membrane, which was hybridised with a ^{32}P-labelled probe to *lpsB* shown in **A**.
E = *Eco*RI site

The Δ*lpsB* mutant strain DFM3 was finally screened by Southern analysis to confirm an integration event (Fig. 3.6C). A ~16-kb band hybridised in the wild-type and a 4-kb band hybridised in the mutant, due to introduction of an *Eco*RI site in the deletion construct. DFM3 was thus confirmed as a deletion mutant and was analysed phenotypically.

3.3 Phenotypic Characterisation of DFM3

3.3.1 Growth in Culture and *In Planta*

No difference in morphology or growth rate on solid or liquid media was observed between the mutant, parent isolate and a strain with an ectopically integrated deletion construct. To determine whether the *lpsB* mutation had an effect on symbiosis with perennial ryegrass, etiolated ryegrass seedlings were inoculated with wild-type and mutant strains (section 2.14). These plants were grown for approximately five weeks and screened for infection using tissue-print immunoblotting (section 2.14). Ten infected plants were examined and mutants showed a wild-type plant-interaction phenotype, with no observable difference in plant or hyphal growth (examples shown in Fig. 3.7).
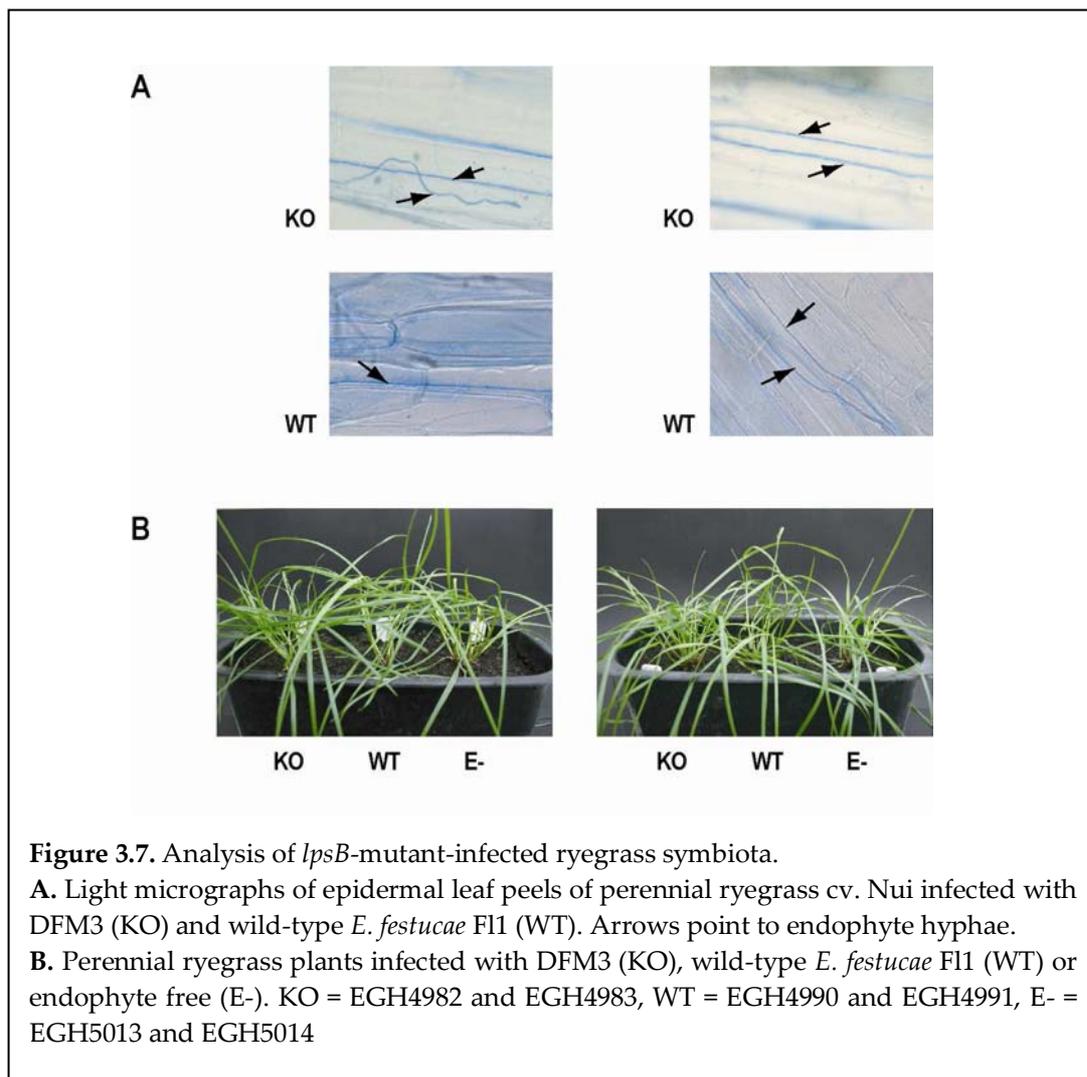
Figure 3.7. Analysis of *lpsB*-mutant-infected ryegrass symbiota.
A. Light micrographs of epidermal leaf peels of perennial ryegrass cv. Nui infected with DFM3 (KO) and wild-type *E. festucae* Fl1 (WT). Arrows point to endophyte hyphae.
B. Perennial ryegrass plants infected with DFM3 (KO), wild-type *E. festucae* Fl1 (WT) or endophyte free (E-). KO = EGH4982 and EGH4983, WT = EGH4990 and EGH4991, E- = EGH5013 and EGH5014

3.3.2 Alkaloid Analysis of the *lpsB* Mutant

LC-MS/MS with selected reaction monitoring was performed to determine the effect of the *lpsB* mutation on ergot alkaloid production (Fig. 3.8). Ergovaline and its stereoisomer ergovalinine were not detected in samples from symbiota containing the *lpsB* mutant compared with wild type, confirming the expected requirement of *lpsB* for ergovaline biosynthesis. Lysergic acid and its C8-stereoisomer were observed to accumulate to elevated levels in *lpsB* mutant symbiota compared to wild-type associations.

Lysergyl alanine, an amide of lysergic acid previously identified by Panaccione *et al.* (Panaccione *et al.*, 2003) as a missing compound in an *lpsA* mutant of *Neotyphodium* sp. Lp1, was also identified together with its C8-stereoisomer in
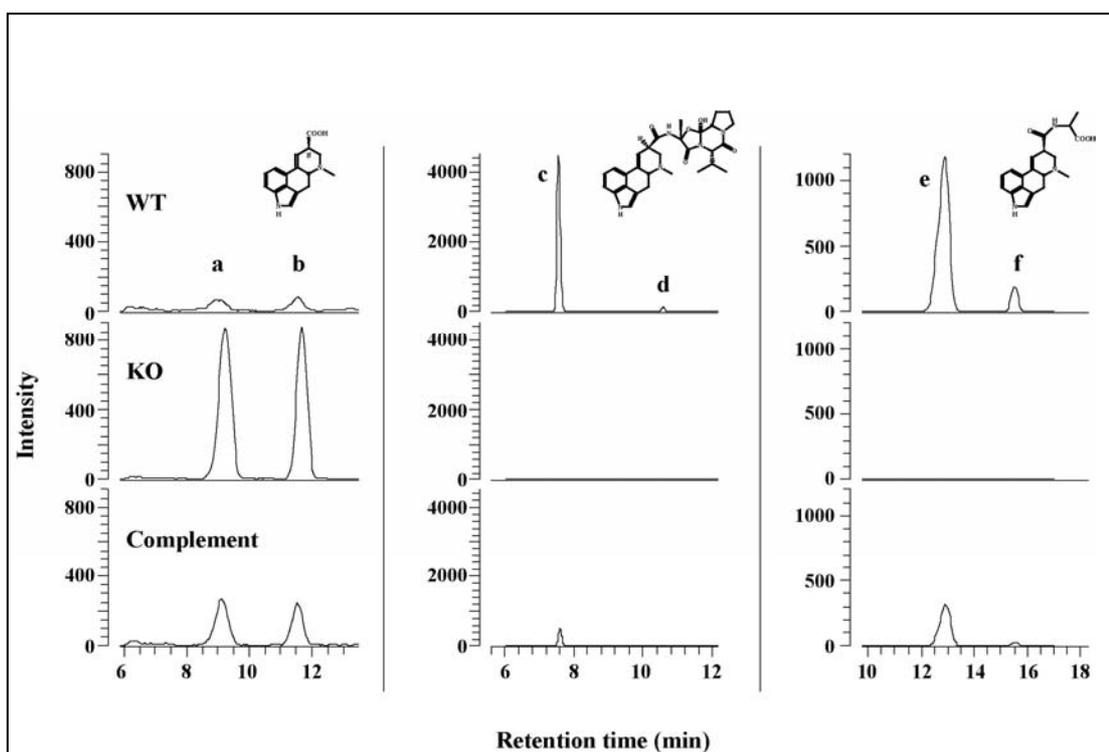
Figure 3.8. Extracted ion chromatograms from LC-MS/MS of extracts of perennial ryegrass plants infected with wild type *E. festucae* Fl1 (WT), *lpsB* mutant DFM3 (KO), or DFM3 transformed with pDF1 (Complement), showing accumulation of lysergic acid (a, shown above) and isolysergic acid (b) (m/z 269 → 223), ergovaline (c, shown above) and ergovalinine (d) (m/z 534 → 268), and lysergyl-alanine (e, shown above) and isolysergyl-alanine (f) (m/z 340 → 223). Relative abundance values are on the same (arbitrary) scale for each extract.

symbiota containing *E. festucae* strain Fl1 but was absent in *lpsB* mutant symbiota. Samples from *Neotyphodium* sp. Lp1 wild type and *lpsA* mutant symbiota were included in this analysis as reference associations. Using the sensitive mass spectrometry method used here, small amounts of lysergyl alanine were detected in the *lpsA* mutant samples (~14 – 190-fold less than wild type, Table 1), suggesting that at least some synthesis of this compound is possible without the LpsA enzyme.

Synthesis of ergine, another lysergic acid amide, was also previously shown to be blocked by the *lpsA* mutation in *Neotyphodium* sp. Lp1 (Panaccione *et al.*, 2003) and was also absent from *lpsA* mutant samples in this study. While no samples of *E. festucae lpsB* mutant-infected symbiota contained ergine, production of ergine was not always detected in wild-type symbiota.

The novel clavine compound 6, 7-secolysergine, observed to accumulate in the *Neotyphodium* sp. Lp1 *lpsA* mutant (Panaccione *et al.*, 2003), was not observed at detectable levels in any *E. festucae* association. However, an unknown peak detected by LC-MS/MS monitoring of the selective reaction characteristic of 6, 7-secolysergine (*m/z* 241 $\rightarrow$ 210) did appear to increase in concentration in the *E. festucae lpsB* mutant symbiota (Table 3.3). This compound was absent in *Neotyphodium* sp. Lp1 associations. Analysis of other clavine intermediate compounds did not show any significant reduction or accumulation resulting from *lpsB* mutation (Table 3.3).

3.3.3 Complementation of DFM3

In order to complement the *lpsB* mutation, pDF1 (Fig. 3.1), containing the wild-type *lpsB* gene and 379 bp of upstream sequence, was co-transformed into *lpsB* mutant protoplasts with pII99, which confers geneticin resistance. An arbitrary selection of these transformants (DFM7, DFM8 and DFM9) was used to infect perennial ryegrass seedlings and the ergot alkaloid phenotype of these associations was determined by LC-MS/MS (Fig. 3.8, Table 3.3). These strains were able to synthesise ergovaline and lysergyl alanine, and on occasion, ergine.

Table 3.3. LC-MS/MS analysis of mutant/grass symbiota

strain*	no. of samples	chanoclavine	agroclavine	elymoclavine	lysergic acid	ergovaline	lysergyl alanine	ergine	unknown clavine
Fl1	5	36.4 – 118..1 [a]	3.1 – 11.2	2.8 – 8.3	0.4 – 0.9	15 – 40.9	5.8 – 11.4	ND – 1.7	0.8 – 1.3
DFM3	6[†]	24.2 – 113.8	2.8 – 7.1	ND – 12.9	1.6 – 5.6	ND	ND	ND	1.7 – 4.0
DFM7	3	64.4 – 160.4	5.0 – 8.1	9.1 – 19.3	2.1 – 11.8	ND – 2.1	ND – 1.4	ND	2.0 – 5.0
DFM8	3	67.3 – 159.8	5.2 – 23.9	3.7 – 15.1	1.8 – 11.8	0.4 – 5.3	1.1 – 2.6	ND – 1.1	1.1 – 2.7
DFM9	3	50.3 – 65.4	6.7 – 11.8	3.2 – 19.9	2.7 – 5.8	ND – 2.6	ND – 1.2	ND	1.6 – 2.1
Lp1	2	913.0 – 1872.6	4.6 – 14.2	25.2 – 62.6	0.7 – 6.7	61.1 – 584.7	37.0 – 312.3	5.2 – 13.5	ND
DFM16	3	1113.3 – 2583.3	7.3 – 21.6	52.9 – 93.7	28.8 – 139.4	ND	0.6 – 1.6	ND	ND
E-	7	ND	ND	ND	ND	ND	ND	ND	ND

*Fl1 = *E. festucae* Fl1;DFM3 = *E. festucae* Fl1, Δ*lpsB*; DFM7, DFM8, DFM9 = individual transformants of DFM3 transformed with pDF1; Lp1 = *Neotyphodium* sp. Lp1; DFM16 = *Neotyphodium* sp. Lp1, Δ*lpsA*; E - = endophyte-free ryegrass.

[a]Estimates of accumulation of metabolites are in arbitrary units relative to ergotamine internal standard, corrected for sample dry weight, but uncorrected for differences in mass spectrometric response and isomerisation of the standard. [†]Two samples from each of three plants

3.4 Insect Bioassays with the *lpsB* Mutant

Ergovaline is a known feeding deterrent of black beetle (*Heteronychus arator*) (Ball *et al.*, 1997). To determine whether intermediate ergot alkaloids, or other *E. festucae* Fl1 compounds, have activity against black beetle, a feeding choice test was performed. Seven replicate cages were prepared with three perennial ryegrass cv. Nui plants infected with *E. festucae* Fl1 wild type, the *lpsB* mutant DFM3, or without endophyte infection. After 16 days plants were analysed for tiller damage. There was no observable feeding in two of the cages and so the data from five cages were analysed. A generalised linear model was used to model the number of damaged tillers out of the total number of tillers on each plant. There was much higher preference (significant at the 5% level) for endophyte-free plants over those infected with either wild type or mutant *E. festucae* (Fig. 3.9), indicating a strong endophyte effect on feeding choice. There was a small preference for *lpsB* mutant *E. festucae-* over wild type-infected plants (Fig. 3.9), although this was not significant at the 5% level. These data indicate that compounds other than ergovaline are able to deter black beetle feeding. It cannot be determined from these data whether these compounds are ergot alkaloid intermediates or other metabolites synthesised by the endophyte.

Figure 3.9. Feeding preference of black beetle. Data shown is the mean percentage of tiller damage from beetle feeding from five replicate cages which included one perennial ryegrass plant for each treatment. Error bars indicate standard error. E- = endophyte free, WT = *E. festucae* Fl1-infected, KO = DFM3-infected
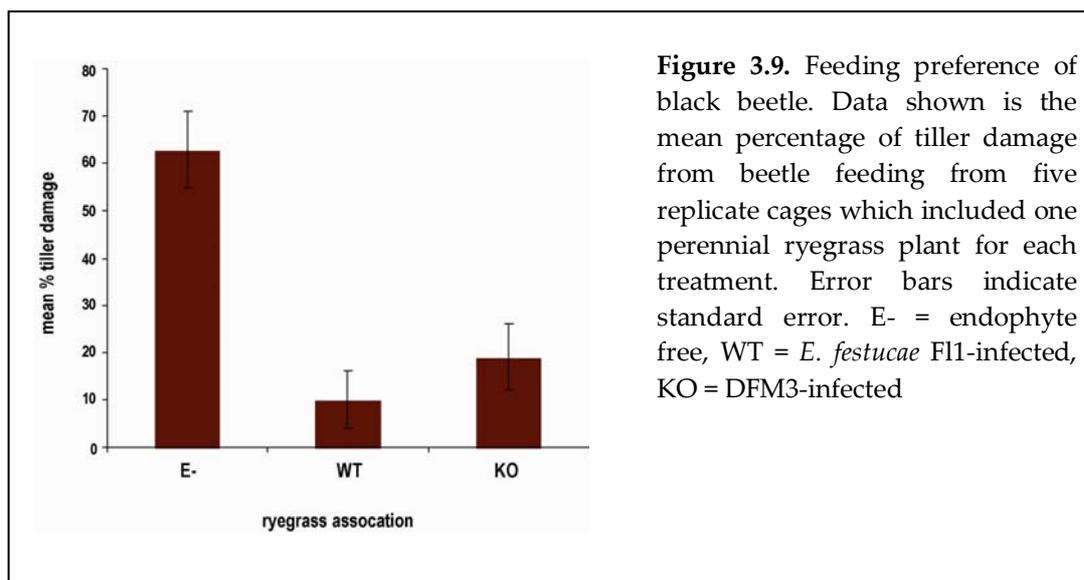
3.5 Extension of the *EAS* Cluster

To extend the sequence of the cluster, additional clones were isolated by screening λZAPII and λGEM-12 libraries, and by inverse PCR. No clones could be isolated from the λZAPII library using a probe to the right end of the cluster amplified with primer set DamP10/DamP11. However, inverse PCR of a 1.8 kb *Hin*dIII fragment, that was determined by Southern analysis to extend beyond the right end of pDF1, extended the sequence by 1.4 kb (Fig. 3.10). This sequence was extremely AT-rich (80%) from 601 bp 3' of *lpsB* and was thus not extended further.

A third λZAPII clone, pDF3, was isolated using a probe to the left end of the cluster amplified with primers DamP21/DamP22 that extended the sequence only 791 bp (Fig. 3.10). Left-end sequence was present, albeit underrepresented, in the λGEM-12 library with four clones, λDF1 – 4, isolated from 110 000 recombinant phage plated. The four clones spanned 19.4 kb and extended the previously isolated sequence by 4.1 kb at the left end and, unexpectedly, 4.1 kb at the right end (Fig. 3.10).
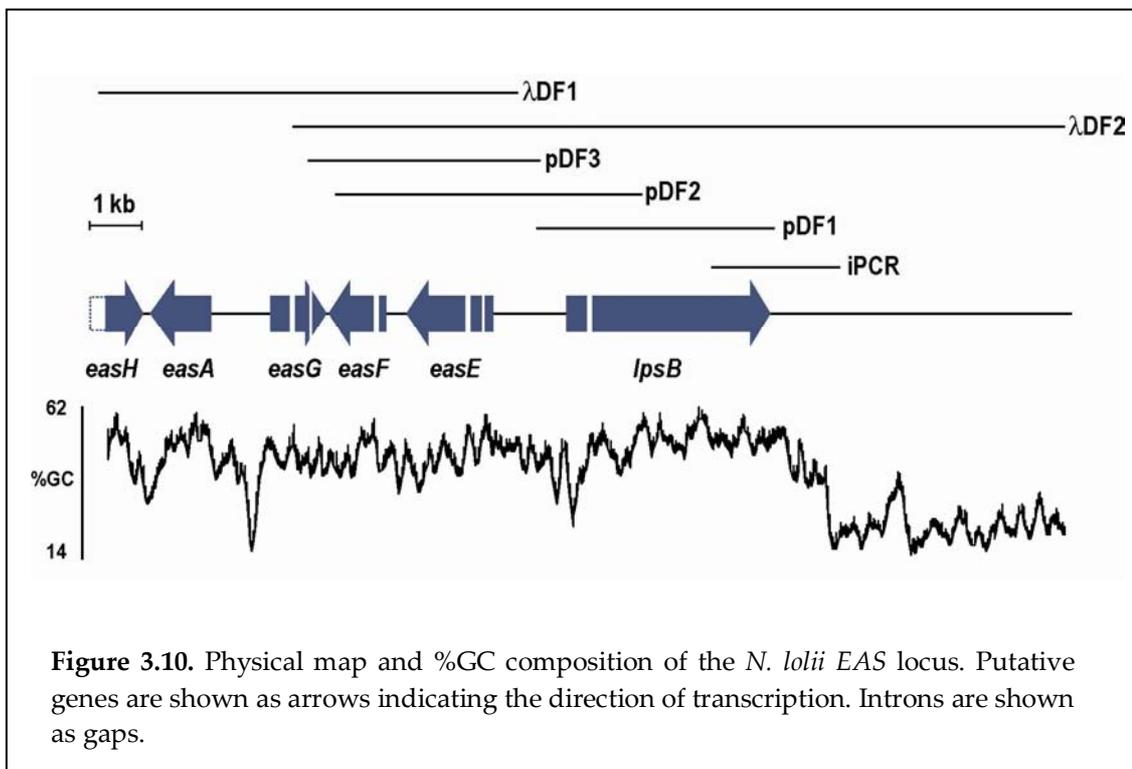
Figure 3.10. Physical map and %GC composition of the *N. lolii EAS* locus. Putative genes are shown as arrows indicating the direction of transcription. Introns are shown as gaps.

Bioinformatics analysis of the 4.1-kb left-end sequence identified three further putative genes encoding a second oxidoreductase (*easA*), a conserved ORF with unknown function (*easG*) and a dioxygenase (*easH*), as well as the full length *easF* (Table 3.4 and Fig. 3.10). The *easH* gene was truncated at the 5' end within the lambda clone, *λ*DF1 (Fig 3.11A). Upstream of this truncated sequence was 313 bp of sequence without similarity to sequences in public databases, then 2.8 kb of a putative ORF with 91% identity to the 5' end of the *torA* gene from *Fusarium graminearum*. It thus appeared that *λ*DF1 was a chimeric clone, which was confirmed by PCR analysis (Fig. 3.11B). The genuine *easH* sequence was determined based on predicted protein sequence alignment with the *C. purpurea* EasH sequence.

The 5.5-kb right end sequence was extremely AT-rich (76%), compared with the composition of the *EAS* genes (50%). BLASTX analysis of this region indicated that this sequence is derived from two nested retrotransposons that are highly degenerate, with no evidence of intact ORFs. This degeneracy is probably a result of repeat-induced point mutation (RIP), which results in C:G to T:A transitions (Cambareri *et al.*, 1989) and was previously observed in two epichloë retroelements, Tahi and Rua, associated with the *LTM* gene cluster (Young *et al.*, 2005). *EAS* cluster transposons and RIP are further analysed in Chapter 5.

Figure 3.11. *λ*DF1 is a chimeric clone. **A.** Physical map of the "left" end of *λ*DF1. The vertical dashed line indicates the point where the re-arrangement occurred and where the genuine *EAS* cluster sequence ends. Primers DamP72 (1) and DamP70 (2), used for PCR in **B** are shown above. **B.** PCR analysis showing *λ*DF1 does not represent genomic sequence
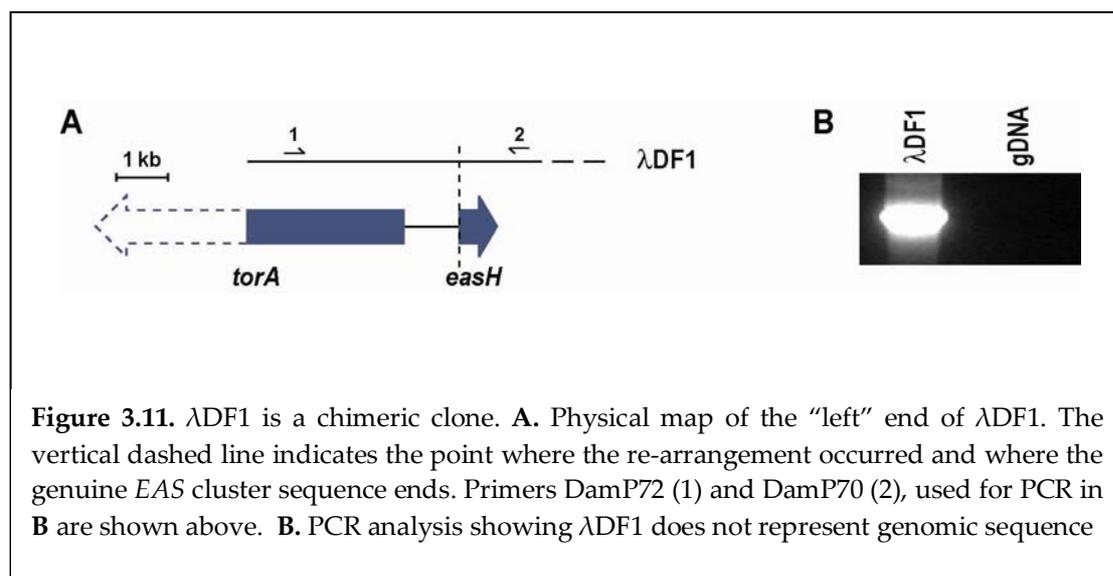
Each of the additional putative genes in the *EAS* cluster had homologues in the *C. purpurea* and *A. fumigatus* ergot alkaloid gene clusters (Table 3.4). The presence of these genes in each organism, with the exception of *easH* which appears to be a pseudogene in *A. fumigatus*, suggests they play a role in the shared early and intermediate steps of ergot alkaloid synthesis. Sequence analysis of each of the three gene clusters (Fig. 3.12) showed substantially different gene order between the three clusters. Interestingly though, there was no more conservation of gene order between the two clavicipitaceous species, *N. lolii* and *C. purpurea*, than between either *N. lolii* or *C. purpurea* and *A. fumigatus* clusters.

Table 3.4. Similarity of *N. lolii eas* genes with *C. purpurea* and *A. fumigatus* orthologues

Gene[a]	*C. purpurea*		*A. fumigatus*	
	Identity (%)	E-value	Identity (%)	E-value
lpsB	49	0.0	-	-
easE	54	e-123	49	e-113
easF	61	e-118	60	e-115
easG	52	4e-78	50	1e-64
easA	75	e-152	57	3e-131

[a]the partial *easH* gene is not included as the *C. purpurea easH* gene is not deposited in GenBank and the *A. fumigatus easH* is a pseudogene
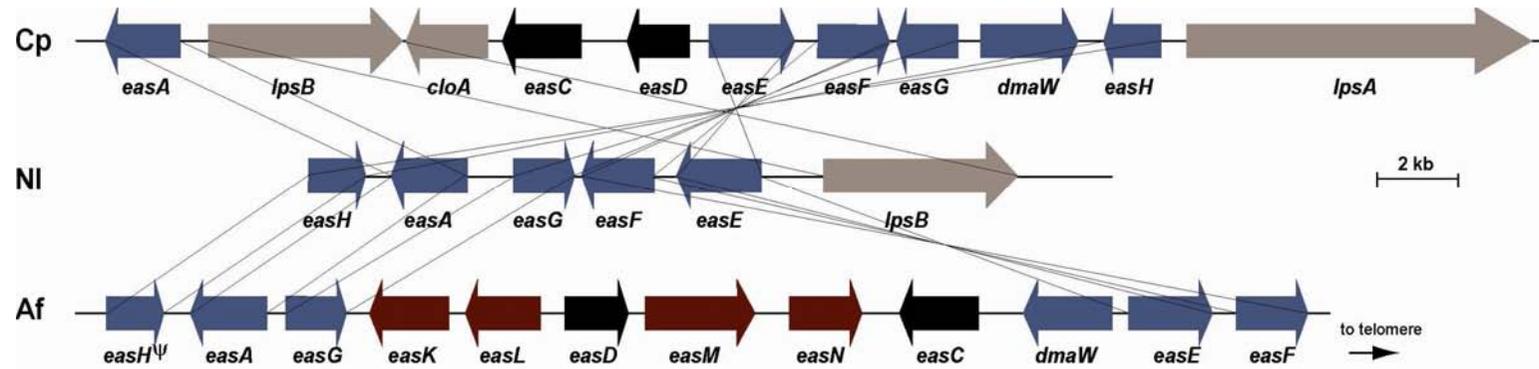
Figure. 3.12. Comparative *eas* gene order among *N. lolii*, *C. purpurea* and *A. fumigatus*. Cp, *C. purpurea*; Nl, *N. lolii*; Af, *A. fumigatus.* Blue arrows, genes shared between each organism; black arrows, genes proposed to be required for ergot alkaloid synthesis but not yet identified in *N. lolii* (later identified in *E. festucae*, section 3.8); light brown arrows, genes found in *N. lolii* and *C. purpurea* but not *A. fumigatus*; dark brown arrows, genes found only in *A. fumigatus*. *N. lolii dmaW*, *cloA* and *lpsA* are not shown as their location relative to the *EAS* cluster is not yet known.

3.5.1 Analysis of *easA*, *easF* and *easG*

Each of the predicted genes in the extended *EAS* cluster were analysed using various bioinformatics software. Data are summarised in Tables 3.4 and 3.5. The full-length *easH* gene, which is truncated in the cloned *N. lolii* sequence, was later found in the sequenced genome of *E. festucae* E2368 and so is described in section 3.8.

3.5.1.1 The *easA* Gene and Predicted Protein EasA

The predicted *easA* gene is 1143 bp long and does not contain introns; the predicted protein is 380 aa long, a predicted unmodified molecular mass of 42 kD. Analysis with the CDD and InterproScan revealed an old yellow enzyme (OYE)-like flavin mononucleotide (FMN)-binding domain (cd02933.2) between residues 6 and 348. The closest functionally characterised protein matches in GenBank were 12-oxo-phytodienoic acid reductases, with the OPR5 gene from *Zea mays* being the closest match (40% identical). These, and other OYE-like enzymes, bind FMN as a prosthetic group and oxidise NADPH to reduce phenolic substrates (Williams and Bruce, 2002). The EasA gene has 75% and 57% identity respectively with *C. purpurea* and *A. fumigatus* orthologues and may be involved in the epimerisation of chanoclavine aldehyde, proposed to be a step in the conversion of chanoclavine to agroclavine (Schardl *et al*., 2006).

3.5.1.2 The *easF* Gene and Predicted Protein EasF

The predicted *easF* gene is 1103 bp long and contains one 69-bp intron 265 bp from the predicted translational start site (Fig. 3.10), which was confirmed by cDNA sequencing.

The predicted EasF protein is 344 aa long, a predicted unmodified molecular mass of 38 kD. Analysis of the predicted sequence using the CDD showed this protein belongs to a family of conserved proteins with unknown function (COG4301). This was reflected in the BLASTP results, with many similar, but no functionally characterised proteins present in the database. The top match to a protein with a predicted function was to a predicted S-adenosyl-L-methionine (SAM)-dependent methyltransferase from *Mycobacterium ulcerans* (accession YP_907775), which was

26% identical to EasF. Analysis with InterproScan revealed the presence of a SAM-dependent methyltransferase domain (SSF53335) from the Structural Classification of Proteins (SCOP) database and also an α/β-hydrolase domain (SSF53474). EasF is 61% and 60% identical respectively to orthologues in *C. purpurea* and *A. fumigatus*. If EasF does have methyltransferase activity, it could be predicted to be involved in the early step of *N*-methylation of dimethylallyl tryptophan, a step in the synthesis of chanoclavine (Schardl *et al.*, 2006).

3.5.1.3 The *easG* Gene and Predicted Protein EasG

The predicted *easG* gene is 1059 bp long and contains two introns of 78 and 53 bp, 480 and 757 bp from the predicted translational start codon respectively (Fig. 3.10), which were confirmed by cDNA sequencing.

The predicted EasG protein is 309 aa long, a predicted 34 kD. Analysis of the predicted protein using CDD predicted a domain between aa 22 and 296 found in predicted nucleoside-diphosphate-sugar epimerases (COG0702). InterproScan analysis predicted a NAD(P)H-binding Rossmann-fold domain (SSF51735). Similar to EasF, there were no functionally-characterised protein hits from a BLASTP analysis but many conserved hypothetical proteins. The top hit to a gene with a predicted function is to a predicted sugar epimerase from *Terrabacter* sp. DBF63, which was 28% identical to EasG. EasG is 52% and 50% identical respectively to *C. purpurea* and *A. fumigatus* orthologues. An obvious function for EasF cannot be determined from sequence analysis, however, the presence of an NAD(P)H-binding domain suggests the protein may catalyse a reduction/oxidation reaction, several of which are required in the early part of the ergot alkaloid pathway.

Table 3.5. Bioinformatic analysis of genes within the *N. lolii EAS* cluster

Gene	Function or putative function	Size (bp)	(aa)	Conserved domains (position)	Non-EAS match[†] (accession)	Organism	BLASTP E-value
lpsB	Single-module NRPS	4130	1352	AMP-binding (303 – 709) PP-binding (835 – 900) Condensation (940 – 1205)	PerA (BAE06845)	*E. festucae*	0.0
easE	Reductase/dehydrogenase	1951	605	FAD-binding (49 – 179) Berberine-like (526 – 571)	AN9231.2 (XP_682500.1)	*A. nidulans*	2e-129
easF	Methyltransferase	1103	344	COG4301 (1 – 344) SAM-dependent methyltransferase* hydrolase*	MUL_4285 (YP_907775)	*M. ulcerans*	1e-11
easG	Reductase/dehydrogenase	1059	309	COG0702 (22 – 296) NAD(P)H-binding Rossmann fold*	ORF59 (BAE45078)	*T.* sp. DBF63	2e-13
easA	Reductase/dehydrogenase	1143	380	OYE-like-FMN (6 – 348)	OPR5 (AY921642.1)	*Z. mays*	1e-70
easH [a]	Oxygenase/hydroxylase	558[a]	188[a]	PhyH (1 – 110)	Fum3p (AAG27131)	*G. moniliformis*	5e-13

[a]*easH* is truncated at the 5' end in λDF1 and therefore only a partial sequence was studied. *exact position is not output for domains identified by SCOP database. [†]nearest functionally-characterised non-*EAS* match (or with predicted function where no functionally characterised proteins are similar) from BLAST

3.6 Linkage of the *EAS* Gene Cluster with *dmaW* and *lpsA* and Identification of *cloA*

Two genes for ergovaline biosynthesis have been previously characterised: *dmaW*, a dimethylallyl tryptophan synthase-encoding gene required for the first committed step in ergot alkaloid biosynthesis (Tsai *et al*., 1995) and *lpsA*, a gene encoding a trimodular NRPS that, along with a second NRPS, predicted to be LpsB, is required for formation of lysergyl peptide lactam (Panaccione *et al*., 2001). To test the hypothesis that these genes were physically linked with the gene cluster isolated in this study, long PCR was attempted with primers facing outward from *lpsB* and *easH* ends of the *EAS* cluster combined with primers facing outward from each end of *dmaW* and the 5' end of *lpsA* in appropriate combinations. A specific fragment was not amplified and so the λZAPII library was screened for clones containing these genes. One clone, pDF4, was isolated for *lpsA* (Fig. 3.13) containing 1204 bp of unique sequence 5' of the *lpsA* sequence deposited in GenBank (accession AF368420). This sequence contained 655 bp of a Tahi LTR retrotransposon relic at the 5' end. The AT-richness (76%) of this element precluded further chromosome walking.

Figure 3.13. Cloning the 5' end of *lpsA*. The *lpsA* gene and Tahi retrotransposon LTR are shown as arrows indicating the direction of transcription. *lpsA* introns are shown as gaps. *lpsA* CDS accession AF368420 (Damrongkool *et al* 2005)
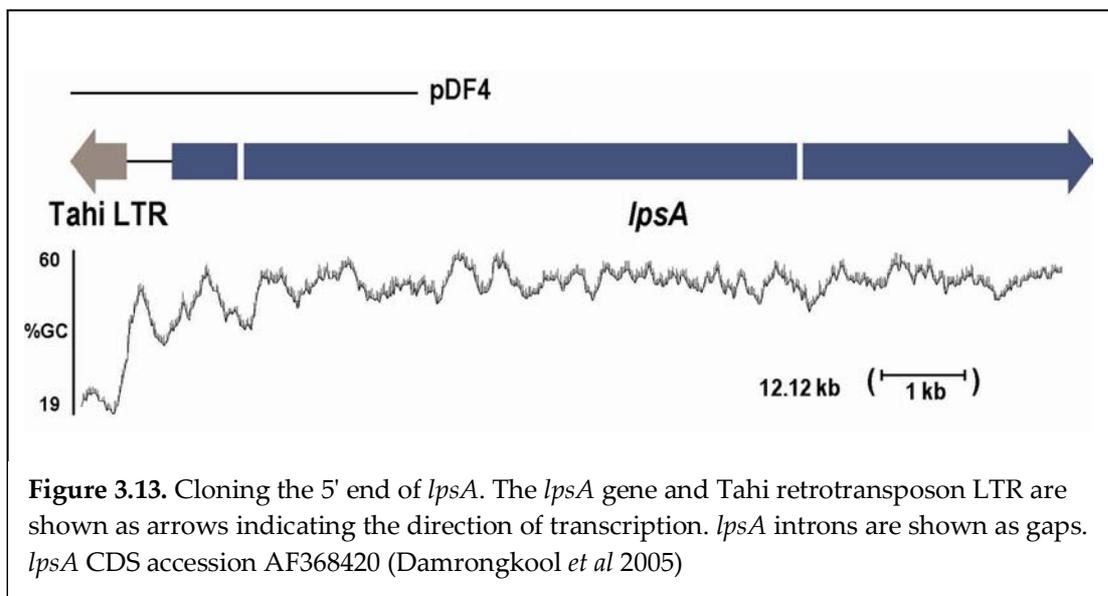
No clone could be isolated for *dmaW*, however, BLASTX analysis of the *N. coenophialum dmaW*-2 flanking sequence in GenBank (accession AY259839.1) revealed the presence of the 3' end of a P450 monooxygenase gene orthologous with the *C. purpurea cloA*-encoded monooxygenase required for conversion of elymoclavine to D-lysergic acid (Haarmann *et al.*, 2006) (Fig. 3.14A). PCR analysis showed *cloA* to be present in *N. lolii* Lp19 and also *E. festucae* Fl1 but immediately adjacent to *dmaW* only in *E. festucae* (Fig. 3.14B).

As cloning was not successful, Southern analysis was performed on *Not*I digests of *N. lolii* Lp19 and *E. festucae* Fl1 genomic DNA (Fig. 3.15). These digests were separated by pulsed-field gel electrophoresis and blotted onto a nylon membrane, then hybridised with ^{32}P probes. Probes for *lpsB*, *lpsA* and *dmaW* each hybridised to a 114-kb fragment in *E. festucae* and a 340-kb fragment in *N. lolii*, indicating that *dmaW* and *lpsA* were linked with the *EAS* cluster in each species.

Figure 3.14. Identification of *cloA*. **A.** Physical map of *dmaW*-2 locus from *N. coenophialum* (accession AY259839). Primers used for PCR in **B** are shown above. 1, dmaWqF; 2, dmaWqR, 3, DamP106; 4, DamP107. **B.** PCR showing presence of *cloA* and linkage with *dmaW* in *E. festucae*. *dmaW*, primers 1 & 2; *cloA*, primers 3 & 4; *dmaW-cloA*, primers 1 & 4. Nl, *N. lolii* Lp19; Ef, *E. festucae* Fl1
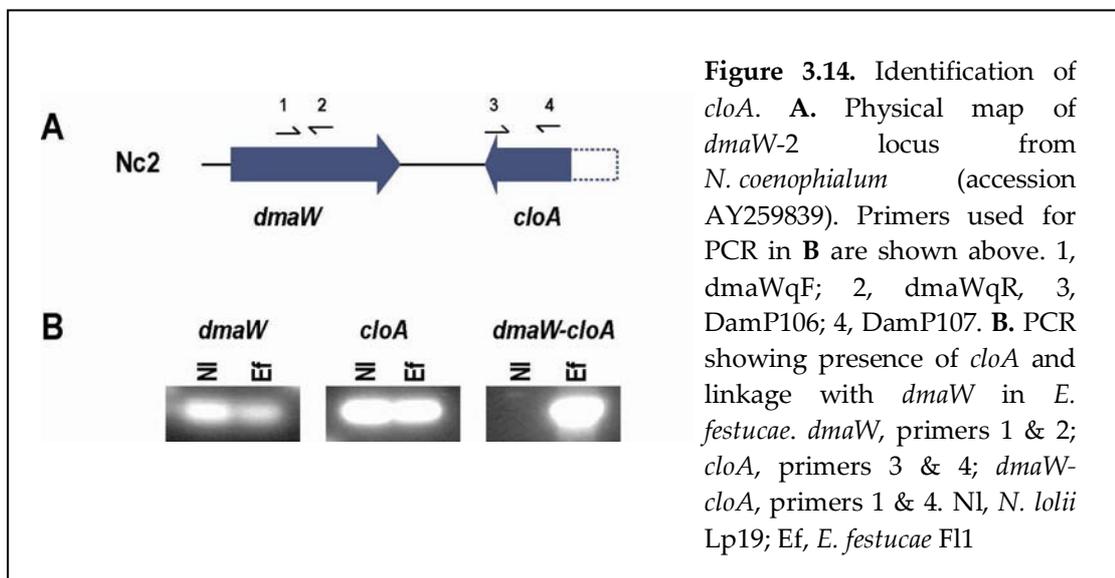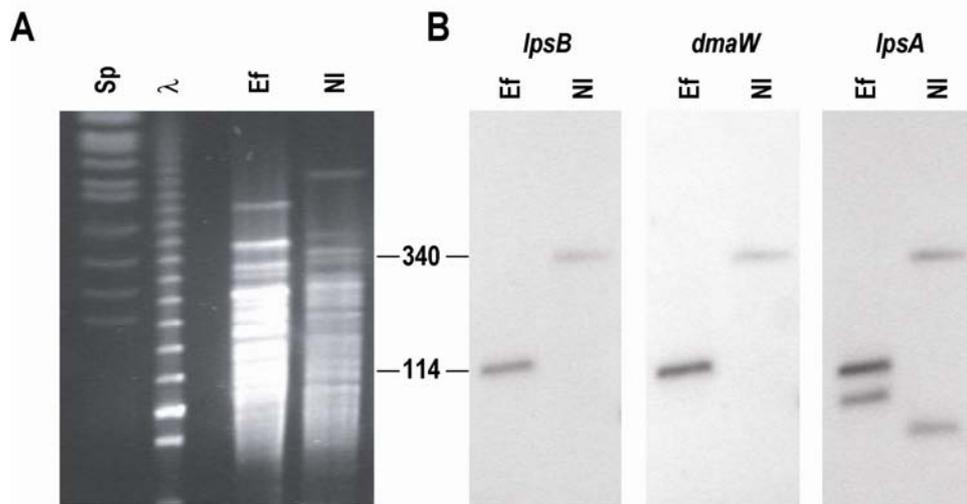
Figure 3.15. Southern blot analysis indicating linkage of *lpsB, lpsA* and *dmaW*. **A.** Ethidium bromide stained gel of *N. lolii* Lp19 (Nl) and *E. festucae* Fl1 (Ef) genomic DNA digested with *Not*I and separated by CHEF pulsed-field gel electrophoresis. Sp, *Schizosaccharomyces pombe* chromosomes; λ, concatamers of bacteriophage λ DNA. **B.** Southern blot of **A** hybridised with ³²P-labelled probes to *lpsB* (pDF2), *lpsA* (amplified with primers lpsA-F and lpsA-R), and *dmaW* (amplified with primers dmaWqF and dmaWqR). The *lpsA* probe contains a *Not*I site and hence hybridises to two bands.

3.7 The *EAS* Cluster is Located Near a Telomere

Recent fungal genome projects have shown that secondary metabolite gene clusters are frequently clustered near the telomeres (Galagan *et al.*, 2005; Nierman *et al.*, 2005; Rehmeyer *et al.*, 2006). To test whether the *EAS* cluster may be also be found in this dynamic region of the genome, a Southern blot of genomic *Not*I digests of *N. lolii* Lp19 and *E. festucae* Fl1 was probed with *lpsB* and the telomeric repeat TTAGGG (Fig. 3.16). The telomeric repeat hybridised with several fragments, one of which corresponded with the same sized band hybridised by *lpsB* in both *E. festucae* and *N. lolii*, indicating that the *EAS* cluster is found near the telomere in each of these species.
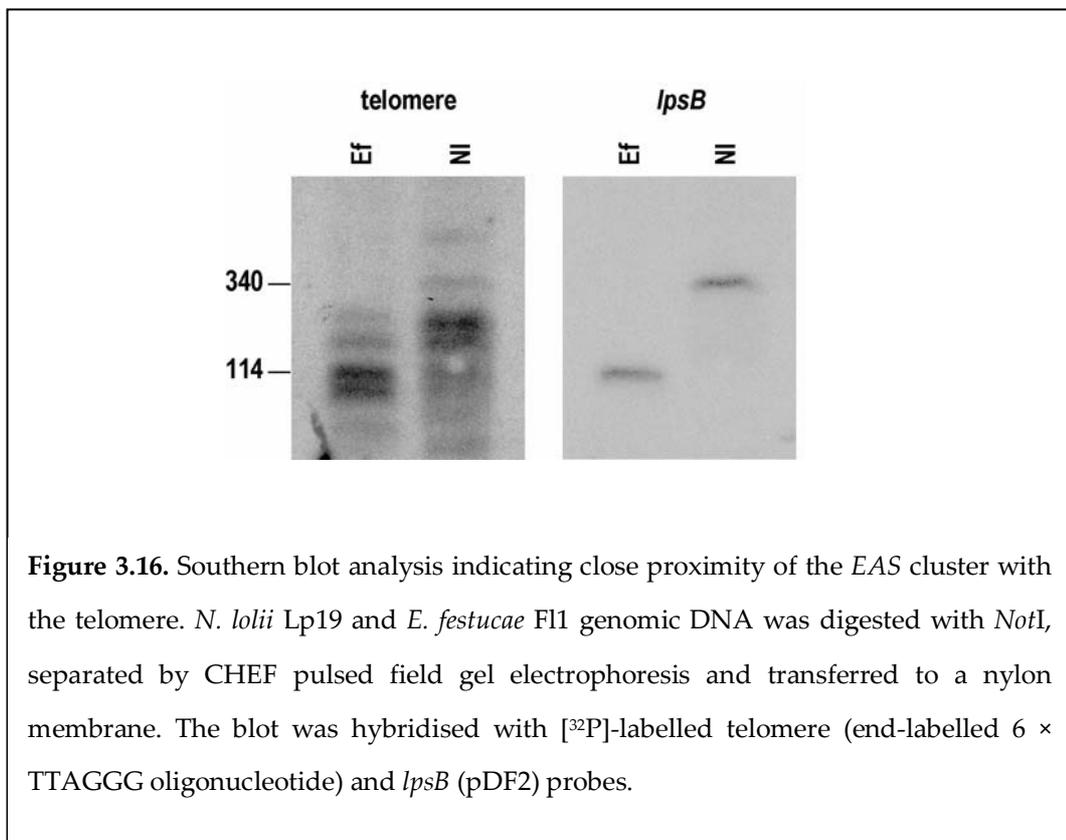
Figure 3.16. Southern blot analysis indicating close proximity of the *EAS* cluster with the telomere. *N. lolii* Lp19 and *E. festucae* Fl1 genomic DNA was digested with *Not*I, separated by CHEF pulsed field gel electrophoresis and transferred to a nylon membrane. The blot was hybridised with [^{32}P]-labelled telomere (end-labelled 6 × TTAGGG oligonucleotide) and *lpsB* (pDF2) probes.

3.8 Genome Sequence Analysis – Completion of the Cluster

The *E. festucae* E2368 genome was recently sequenced in the laboratory of Professor Chris Schardl at the University of Kentucky. To determine whether the *eas* genes identified in this study were present, BLASTN analysis of the genome sequence was performed. This revealed that each of the *eas* genes identified in this study, along with *lpsA* and *dmaW,* were present (Table 3.6) with almost identical nucleotide sequences in *N. lolii* Lp19 and *E. festucae* E2368*.* The *EAS* cluster sequence in the genome assembly was not as connected as the cloned *N. lolii* cluster, being found on four different contigs (Table 3.6).

Two genes, *easC* and *easD*, were present in the *C. purpurea* and *A. fumigatus EAS* clusters but not identified in this study. TBLASTN analysis was thus undertaken, using the *C. purpurea* predicted EasC and EasD protein sequences, to determine whether these genes were present in the *E. festucae* genome sequence. Both *easC* and *easD* were identified (Fig. 3.17), completing the suite of *eas* genes in epichloë that are present in the *C. purpurea EAS* cluster.

Table 3.6. *N. lolii EAS* genes in the *E. festucae* E2368 genome

Gene	Identity (identical bp/total bp)	*E. festucae* contig (length)	Position
lpsB	4125/4130	1241 (10033)	4708 – 8837
easA	1143/1143	3372 (1391)	163 – 1305
easE	1942/1951	1241 (10033)	1488 – 3438
easF	1008/1018*	1241 (10033)	1 – 1018
	54/54*	2216 (1771)	1718 – 1771
easG	1058/1059	2216 (1771)	616 – 1674
easH	556/558	2048 (2125)	1297 – 1854
dmaW[†]	1483/1487[a]	1139 (14909)	3020 – 4506
lpsA[‡]	10890/10922	1183 (12626)	842 – 11763

*easF is incomplete on two contigs, total gene length 1103 bp. [†]The *dmaW* sequence from *Neotyphodium* sp. Lp1 (AY259837) was used to query the *E. festucae* genome as the *N. lolii* gene was not sequenced. [a]Total *Neotyphodium* sp. Lp1 gene length 1485. [‡]Accession AF368420

Both *easC* and *easD* were found on the 14909-bp contig 1139, along with *dmaW* and *cloA* (Fig. 3.17). This contig was very AT-rich at each end, from 1.7 kb upstream of *dmaW* at the left end and from 1.24 kb upstream of *easD* at the right end. The AT-rich sequence at the right end of the cluster had 87 % identity with a Tahi LTR retrotransposon relic from the *N. lolii LTM* cluster. The left end sequence did not match transposon sequences from the database but did have several repeats in the *E. festucae* genome sequence and thus was likely to represent a novel transposable element. (This and other transposons from the *EAS* cluster are further analysed in Chapter 5.)

Genome sequence analysis, combined with *N. lolii* data, thus indicates that the *EAS* cluster is found in three mini-clusters separated by AT-rich sequence in *E. festucae*: cluster one contains *easA, easE, easF, easG, easH* and *lpsB*, cluster two contains *dmaW, cloA, easC* and *easD*, and cluster three contains just *lpsA*. The *N. lolii* cluster may be even more disjointed, as *dmaW* does not appear to be linked to *cloA* (Fig. 3.14).

Figure 3.17. Physical map and GC% of *E. festucae* E2368 Contig 1139. Predicted genes are shown as arrows indicating the direction of transcription. Introns are shown as gaps. The end of a Tahi retrotransposon relic is shown as a dark brown bar.
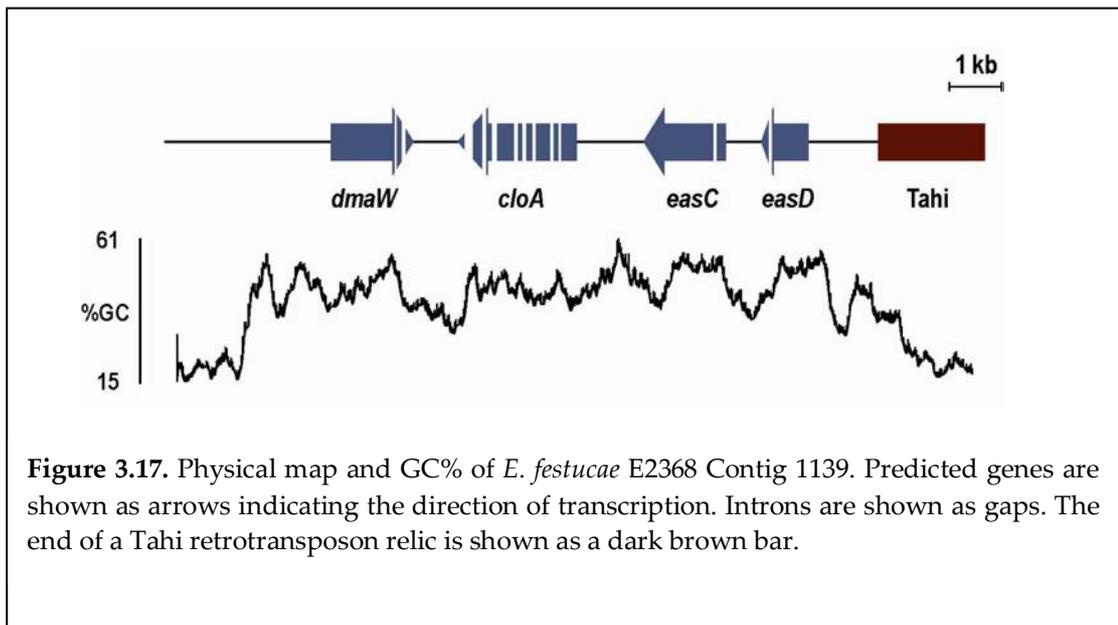
3.8.1 Bioinformatics Analysis of *E. festucae easC*, *easD,* and Full-Length *easH* and *cloA*

3.8.1.1 The *easH* Gene and Predicted Dioxygenase EasH

The *easH* gene was truncated at one end of the *EAS* cluster sequence isolated from *N. lolii*. This sequence was used to identify the full-length gene on the 2125-bp contig 2048 (Fig. 3.18). *easH* is the only predicted gene found on this contig, with the other end of the contig being highly AT-rich (Fig. 3.18). The predicted full-length *easH* gene is 921 bp in length and contains no introns. The *C. purpurea easH* sequence is not deposited in GenBank and so percentage identity could not be determined and the *easH* gene found in *A. fumigatus* is a pseudogene and is unlikely to encode a protein product.

The predicted EasH protein is 307 aa long, a predicted unmodified molecular mass of 33.6 kD. Analysis with the CDD revealed a conserved phytanoyl-CoA dioxygenase (PhyH) domain (pfam05721.5). The top functionally-characterised protein match was to Fum3p from *Gibberella moniliformis* (30% identity), a dioxygenase involved in the fumonisin pathway (Ding *et al.*, 2004). EasH is thus likely to catalyse a hydroxylation step subsequent to the steps shared with *A. fumigatus*, possibly the conversion of agroclavine to elymoclavine or lysergyl peptide lactam to ergovaline (Schardl *et al.*, 2006).
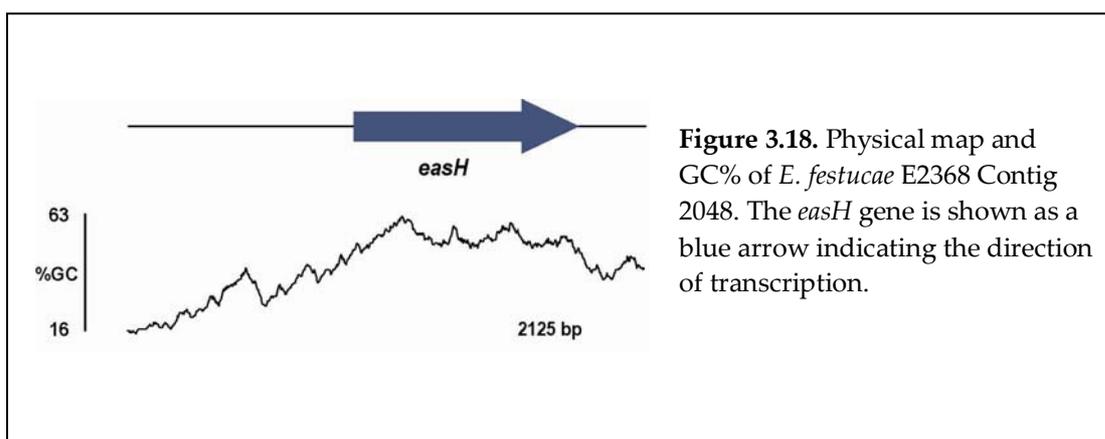
Figure 3.18. Physical map and GC% of *E. festucae* E2368 Contig 2048. The *easH* gene is shown as a blue arrow indicating the direction of transcription.

3.8.1.2 The *cloA* Gene and P450 Monooxygenase CloA

The 3' end of the *cloA* gene was identified downstream of *dmaW* in *N. coenophialum* (section 3.6) and the full-length gene was identified in *EAS* cluster 2, on contig 1139. The *E. festucae cloA* gene is 2140 bp in length and is predicted to contain eight introns (Fig. 3.17), substantially more than other *eas* genes.

The predicted CloA protein is 527 aa long, a predicted unmodified molecular mass of 60 kD. CloA is likely to be an orthologue of the *C. purpurea* protein, to which it is 63% identical. CloA is a P450 monooxygenase required for the conversion of elymoclavine to lysergic acid (Haarmann *et al.*, 2006).

3.8.1.3 The *easC* Gene and Predicted Catalase-Like Protein EasC

The *easC* gene is 1491 bp and manual annotation predicts it to contain a rare GC-AG intron of 63 bp (Fig 3.17). The predicted protein is 476 aa long and a predicted unmodified molecular mass of 54 kD. Analysis with the CDD revealed a catalase domain (cd00328.2) covering the length of the protein. BLASTP results reflected this, showing EasC to be similar to many predicted and characterised catalases. The top functionally characterised match (38% identical), for which a structure has also been solved, was to Catalase A (Cta1p) from *Saccharomyces cerevisiae* (Mate *et al.*, 1999). EasC is 73% and 62% identical respectively with the C. *purpurea* and *A. fumigatus* orthologues. A role for a catalase in ergot alkaloid synthesis is not immediately obvious.

3.8.1.4 The *easD* Gene and Predicted Oxidoreductase EasD

The *easD* gene is 850 bp in length and is predicted to contain one intron of 64 bp (Fig 3.17). The predicted EasD protein is 262 aa long and a predicted unmodified molecular mass of 28 kD. Analysis with the CDD predicted a short-chain dehydrogenase domain (COG4221, COG1028). These enzymes are NAD(P)H dependant and reduce a wide range of substrates (Jornvall *et al.*, 1995). BLASTP hits were largely to hypothetical oxidoreductases. The top hit to a functionally characterised gene was to ABA4 (29% identical) from the *Botryotinia fuckeliana* abscisic acid gene cluster, which is proposed to be required for the conversion of 1', 4'-*trans*-diol-abscisic acid to abscisic acid (Siewers *et al.*, 2006). EasD is 72% and

67% identical respectively to the *C. purpurea* and *A. fumigatus* orthologues respectively. There are several oxidation/reduction reactions in the ergot alkaloid pathway that EasD could potentially catalyse.

4. Results

Expression and Promoter Analysis

4.1 *EAS* Cluster Expression Analysis

Ergot alkaloids are synthesised preferentially during biotrophic growth and are rarely detected in subcultured mycelia (Bacon, 1988). In order to gain insight into the regulation of *eas* genes, transcription was analysed from mycelia grown *in planta* and under a number of different culture conditions by RT-PCR and/or real time RT-PCR.

4.1.1 *In Planta* Expression

To test whether *eas* genes are preferentially expressed *in planta*, semi-quantitative RT-PCR analysis was performed on *EAS* cluster one genes with RNA extracted from *L. perenne* Nui infected with *E. festucae* Fl1 and from *E. festucae* Fl1 grown axenically in PD broth (Fig. 4.1 and 4.2). The amount of fungus in the total plant biomass has been estimated at around 0.3 – 1% using DNA comparisons (Panaccione *et al.*, 2001; Young *et al.*, 2005), therefore the amount of fungal RNA in an infected plant sample will be extremely dilute compared to RNA extracted from mycelia grown in culture. To compensate for this difference, fungal RNA levels were normalised in each sample by dilution, as has been described by Tanaka *et al.* (2005) and Young *et al.* (2005). cDNA derived from mycelia grown in PD broth was serially diluted to 1/100, at which dilution a band of similar intensity to the *in planta* cDNA was amplified by PCR using primers to *E. festucae* Fl1 *tubB*, the constitutively expressed gene for β-tubulin (Fig. 4.1A). RT-PCR amplification from this diluted sample was then compared with amplification from infected plant RNA (Fig. 4.1B). This analysis showed that each gene in *EAS* cluster one was strongly expressed *in planta* but only weakly, in the case of *easG* and *easH*, or not at all expressed in axenic culture when grown in complex media.

Expression analysis was repeated for *lpsB*, *easE*, *easF* and *easG* using real-time RT-PCR (Fig. 4.2). Expression of each gene was determined relative to *tubB*, using the $2^{-(Ct_{tubB} - Ct_{sample})}$ calculation (Livak and Schmittgen, 2001) (Ct = the averaged threshold value of three duplicate samples). This analysis showed a similar result

94

to the semi-quantitative approach, with each of the genes expressed *in planta* but weakly or not at all expressed in axenic culture.
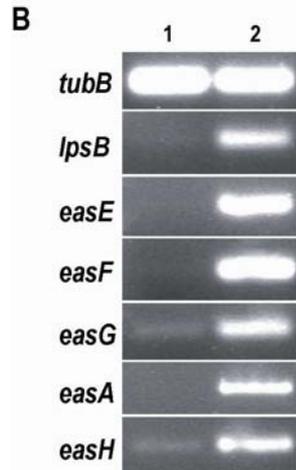
Figure 4.1. RT-PCR analysis.
A. Normalisation of cDNA from *in planta* and cultured mycelium. *tubB* was amplified from cDNA from *E. festucae* Fl1-infected perennial ryegrass and a dilution series of *E. festucae* Fl1 cDNA prepared from cultured mycelium. 1, *in planta* no-RT; 2, culture no-RT; 3, neat in planta; 4, neat culture; 5, 1/5; 6, 1/10; 7, 1/20; 8, 1/40; 9, 1/80; 10, 1/100
B. Amplification of *tubB* and *eas* genes from cDNA from *E. festucae* Fl1-infected perennial ryegrass (1) and 1/100 dilution of cDNA from cultured mycelia (2). Primers described in Table 2.2.
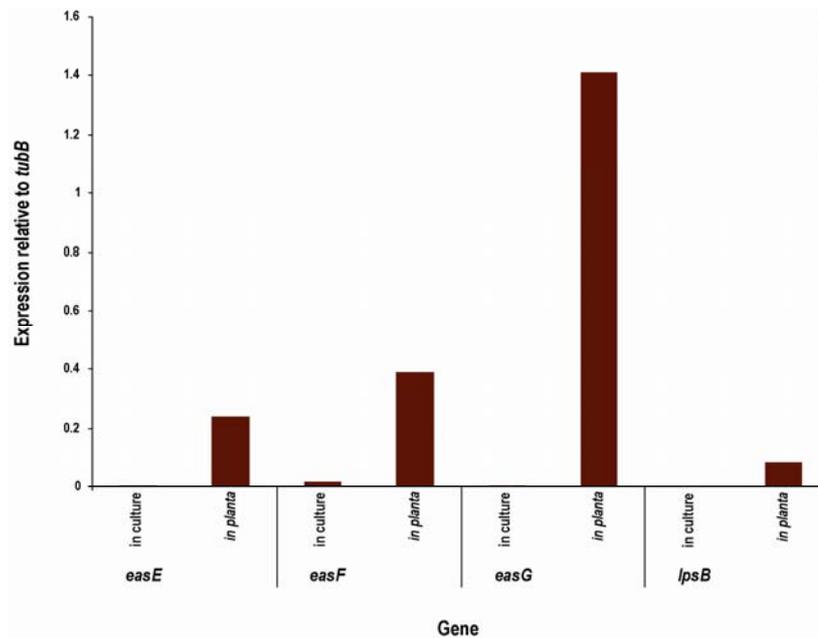
Figure 4.2. Real time RT-PCR analysis of selected genes from the *eas* cluster. Expression levels are shown relative to *tubB* as described in the text. Primers described in Table 2.2.

4.1.2 Expression Analysis of *lpsB* in Axenic Culture

To test the hypothesis that ergovaline genes may be subject to catabolite repression in culture, RNA was analysed from *E. festucae* Fl1 grown in carbon, nitrogen and phosphate limited minimal media, as well as from PD media and infected perennial ryegrass.

For carbon and nitrogen-limited conditions, Czapek Dox medium was used with glucose and ammonium chloride as carbon and nitrogen sources respectively. Cultures were grown under limiting conditions for carbon, nitrogen or both, as well as in media complete for both carbon and nitrogen. RNA isolated from these cultures was tested for *lpsB* expression by real-time RT-PCR and no expression could be detected under any of the growth conditions in two separate cultures for each condition.

In ergot alkaloid producing strains of *C. purpurea*, *eas* gene expression is repressed by high phosphate levels. *E. festucae* Fl1 was grown in *C. purpurea* phosphate-repressing (2.0 g/L KH_2PO_4) and non-repressing (0.5 and 0.2 g/L KH_2PO_4) conditions. Real-time RT-PCR analysis of RNA extracted from these cultures showed no expression of *lpsB*.

Two conditions that epichloë endophytes could be expected to face in the plant apoplast are strongly oxidising conditions (Pignocchi and Foyer, 2003) and exposure to reduced light wavelengths due to a filtering effect of green coloured plant tissues. Many examples exist of fungal gene expression in response to both of these conditions (Lev *et al.*, 2005; Schmoll *et al.*, 2005; Casas-Flores *et al.*, 2006; Narasaiah *et al.*, 2006). However, exposure of *E. festucae* Fl1 cultures to oxidative stress (20mM H_2O_2 for 15 or 30 min) or growth in the dark again failed to induce *lpsB* expression.

Failure to induce expression of *lpsB* in culture suggested that a specific plant factor may be required for induction. However, transfer of *E. festucae* Fl1 grown in PD media to a perennial ryegrass extract for 30 min failed to induce *lpsB* expression.

4.2 *EAS* Cluster Promoter Analysis

EAS promoter regions were analysed in order to identify potential *cis* regulatory motifs. Sequence was analysed upstream of the translational start site of each gene in *N. lolii* Lp19 *EAS* cluster one, *N. lolii* Lp19 *lpsA* and *dmaW* and *E. festucae* E2368 *cloA*, *easC*, *easD* and *easH*. For those genes that are divergently transcribed and share upstream regions, the full intergenic region was analysed. For other genes, 800 bp upstream of the translational start site was examined. Transposable element sequences (see Chapter 5) were removed from each upstream region sequence analysed.

Promoter regions were analysed with four different algorithms: Oligo Analysis, Dyad Analysis, Consensus and MEME. Oligo Analysis searches sequences for shared "words," (oligonucleotide sequences), Dyad Analysis searches for spaced pairs of words, while Consensus and MEME are matrix based algorithms. Oligo Analysis yielded one palindromic motif (motif 1) with a significance index (minus $\log_{10}$ transformation of the E-value) over 1. This motif, TAGCTA is a palindrome consisting of an inverted TAG. No 5 – 14 bp spaced dyads were detected shared among the sequences. Consensus identified four motifs (motifs 2 – 5,) with E values less than 1e-15. MEME failed to identify any significant motif shared among the sequences, however, a short microsatellite (8 × GCACG) was identified 442 bp upstream of *cloA*.

The TRANSFAC database of transcription factor binding sites was queried with each of the motifs. TRANSFAC does not accept degenerate sequences and so for the four motifs identified by Consensus, a sequence based on the most prevalent base at each position of the matrix was used to query the database. Neither motif 1 or 2 had matches in the database, while each of motifs 3 – 5 had matches to the reverse complement of ADR1 (98 – 100%) and STRE (80 – 100%) binding sites, the core of both of which is a run of four cytosine nucleotides. Adr1 is a transcription factor for peroxisomal proteins and is also required for growth on glycerol in yeast. Msn2/4p, which binds STRE, the yeast stress response element, are involved

in transcriptional response to stress. Motifs similar to Adr1 and STRE binding sites have also been identified shared among loline gene promoters (Kutil *et al.*, 2007).

Each of the *EAS* promoter regions, including transposon sequences, were searched for the five identified motifs. The upstream region of five genes not expected to be co-regulated with *EAS* genes were examined as controls. These were *ggsA*, which encodes the primary geranylgeranyl diphosphate synthase; *tubB*, which encodes β tubulin; *hmgA*, which encodes HMG CoA reductase; *pyr4*, which encodes orotidine 5-phosphate decarboxylase; and *gcnA*, which encodes a constitutively expressed β 1-6-glucanase. Other alkaloid pathway genes, *LTM, LOL* and *perA,* for indole diterpene, loline and peramine synthesis respectively, and nc25, to date the most highly expressed plant regulated gene (Johnson *et al.*, 2003), were also examined. This analysis showed that motif 1 was not found in all *EAS* promoters and was found in two out of the five control gene promoters, including five copies upstream of *pyr4* (Fig. 4.3). This motif is therefore not likely to be specifically involved in co-regulation of *eas* genes.

Motif 2 is enriched in *EAS* and some other plant-regulated genes compared to control genes, with the exception of *hmgA* (Fig. 4.3). Although motif 2 is not found upstream of *lpsA* or between *easA* and *easG*, a search allowing one mismatch found motifs 554 and 492 bp upstream of *lpsA* and 255, 51 and 48 bp upstream of *easG*.

Although no single C-rich motif (3 – 5) is found in every *EAS* promoter, and a motif 5 sequence is found upstream of *ggsA*, these motifs are highly enriched in *EAS* gene promoter regions compared with control and other plant-induced genes (Fig. 4.4).
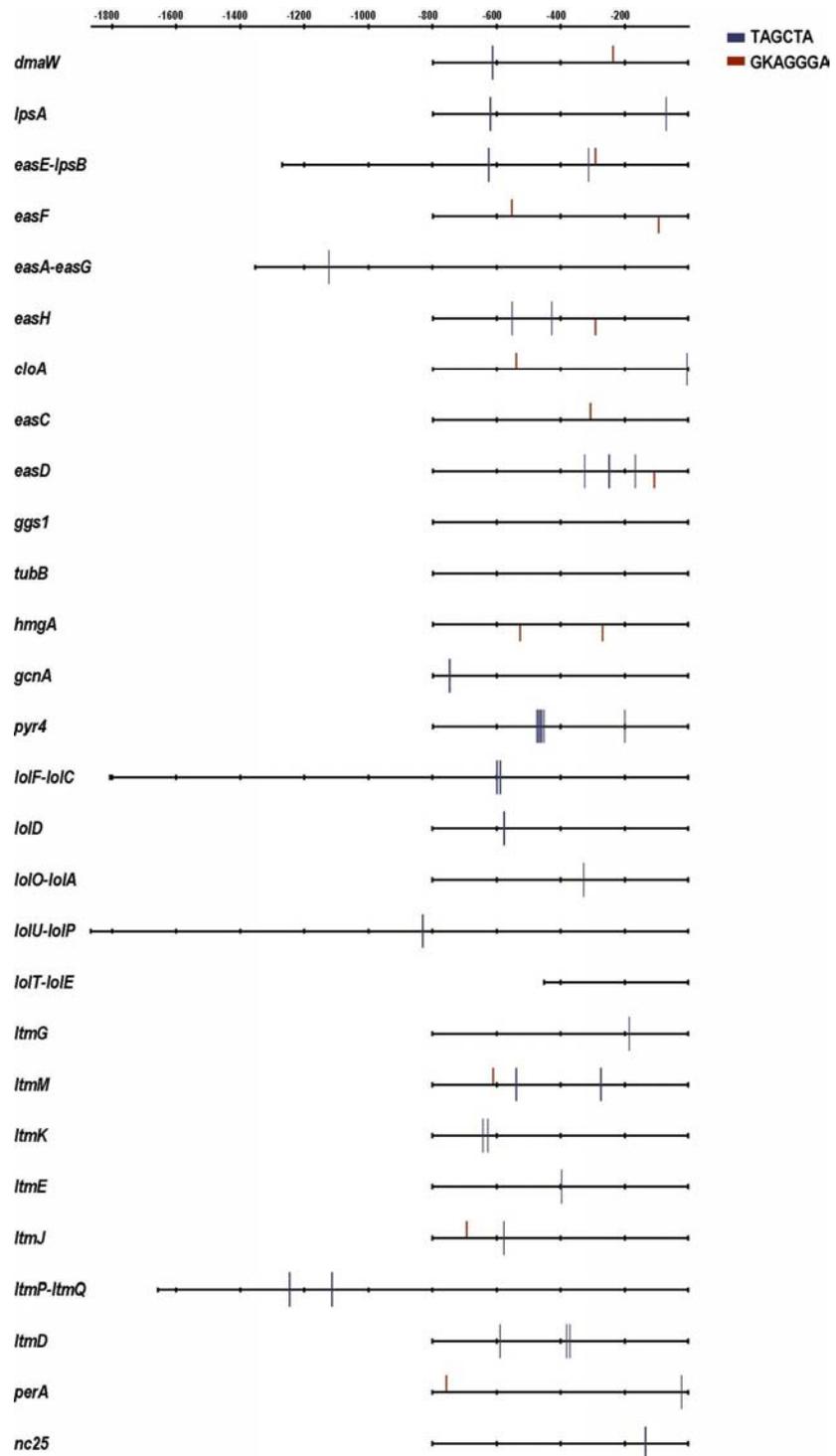
Figure 4.3. Distribution of motifs 1 (blue) and 2 (red) in plant-induced and control promoters. Motifs are shown as coloured bars relative to translational start sites. For regions intergenic between two genes, numbering is relative to the right-hand gene. Motif 2 is shown based on the most prevalent base at each position of the matrix.
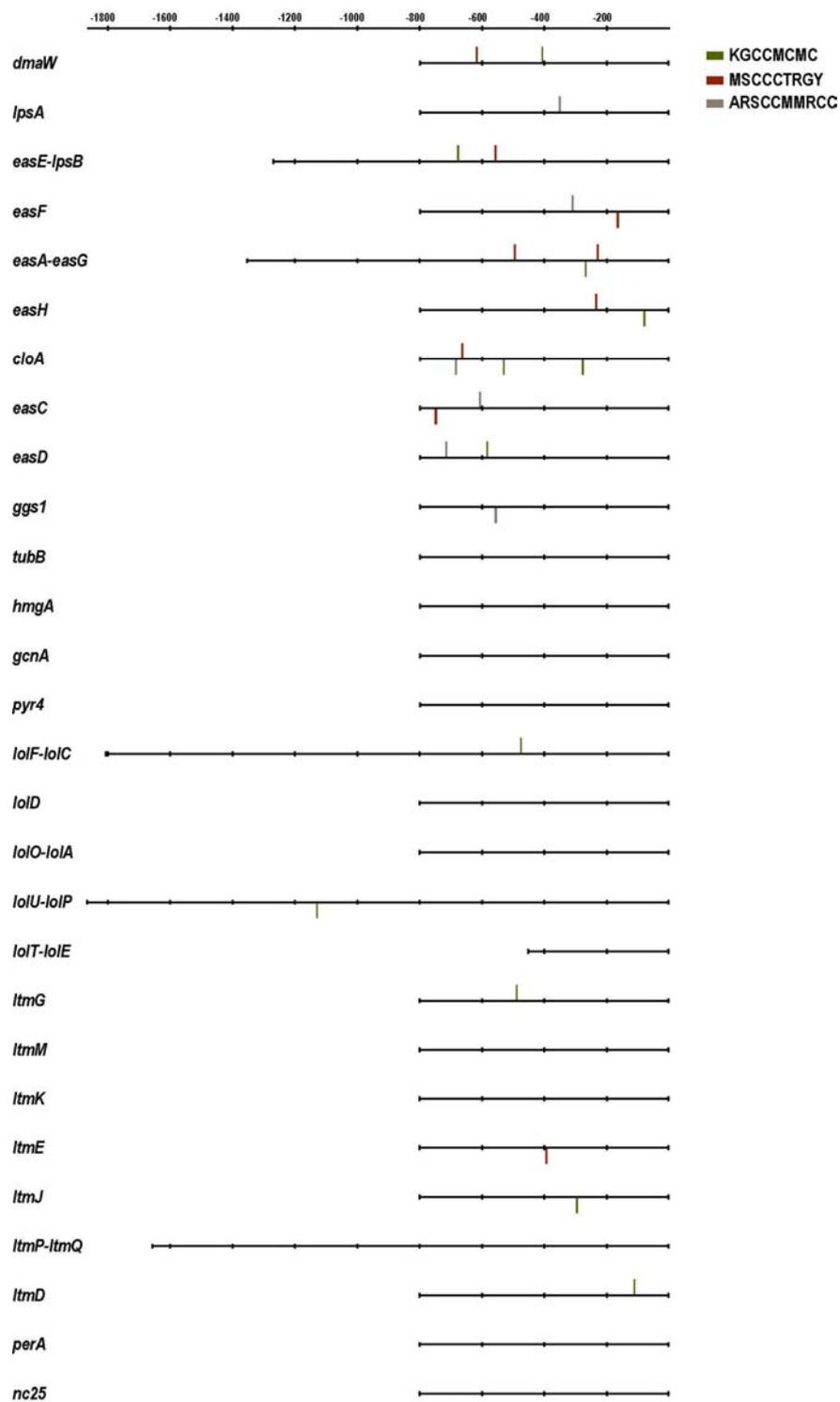
Figure 4.4. Distribution of motifs 3 (green), 4 (red) and 5 (brown) in plant-induced and control promoters. Motifs are shown as coloured bars relative to translational start sites. For regions intergenic between two genes, numbering is relative to the right-hand gene. Motifs are shown based on the most prevalent base at each position of the appropriate matrix.

4.3 *laeA*, a Candidate Regulatory Gene

In aspergilli LaeA, a histone methyl transferase, is a global regulator of secondary metabolism (Bok and Keller, 2004; Bok *et al.*, 2006a), and was recently shown to regulate *eas* genes in *A. fumigatus* (Perrin *et al.*, 2007). This gene was therefore a target for a candidate gene approach to identifying a regulator of *eas* genes in epichloë.

4.3.1 Cloning Strategies

When *E. festucae*, *M. grisea*, *N. crassa*, and *F. graminearum* genomes were queried with the LaeA amino acid sequence, several matches were identified in each genome. When the top match to LaeA from each genome tested was then queried against the *A. nidulans* or *A. fumigatus* genomes however, several other aspergillus proteins were found to be more similar than LaeA itself. Alignment of ascomycete methyltransferases showed significant similarity between these proteins in all genera examined, however aspergillus LaeA proteins segregated into a distinct clade (Fig. 4.5). A clear orthologue outside the aspergilli, based on sequence similarity, was thus not able to be identified. In order to facilitate synteny-based cloning, the genes surrounding *laeA* were identified and examined in *E. festucae* and other ascomycetes (Fig. 4.6). Microsynteny is conserved for most of the genes that surround *laeA* in *A. nidulans*, however, there are no protein methyltransferases found linked to these genes in other ascomycete genomes.

Figure 4.5. Phylogenetic analysis of selected ascomycete protein methyl-transferases. LaeA proteins are highlighted. The unrooted tree (prepared by A. Khan) is based on a maximum likelihood analysis of an alignment using the top two TBLASTN matches to LaeA from the *E. festucae* (Ef) genome (named by contig), the top BLASTP matches from *N. crassa* (Nc), *M. grisea* (Mg) and *G. zeae* (Gz (= *F. graminearum*)) and the top BLASTP matches to these proteins in the *A. nidulans* (An or En) and *A. fumigatus* (Af) genomes. Non-*E. festucae* proteins are named by accession number. LaeA proteins are highlighted.

Figure 4.6. Conserved microsynteny at the *A. nidulans laeA* locus. Gene order is largely conserved with the exception of the *laeA* gene itself, which is not present at this locus in other ascomycete genomes. An, *A. nidulans*; Fg, *F. graminearum*; Mg, *M. grisea*; Nc, *N. crassa*; Ef, *E. festucae*. *E. festucae* labels are FGENESH calls trained on the *F. graminearum* dataset. Other gene labels are from respective genome sequences.
[a]EFT-17 is found at the edges of two contigs, A.3.6282 and A.1.8862 are thus not confirmed as linked.
*This sequence is present but not currently annotated in the *E. festucae* E2368 GBrowse.

4.3.2 Heterologous Expression of *laeA* in *E. festucae*

As no *laeA* orthologue was identifiable from sequence comparison, the *A. nidulans laeA* gene was expressed heterologously in *E. festucae* Fl1, to determine whether this gene could affect expression of epichloë secondary metabolite genes.

4.3.2.1 Preparation of an Expression Construct and Transformation of *E. festucae*

An expression construct was prepared using a MultiSite Gateway system (Fig. 4.7A). A 1.3-kb fragment containing the *laeA* ORF was amplified from *A. nidulans* genomic DNA with *att*B1- and *att*B2-tailed primers (laeA-attB1/laeA-attB2rev) and recombined into the entry vector pDONR 221 to yield pENTRlaeA. This was then recombined with pENTRTEF (R. Johnson and C. Voisey, unpublished), which contains the translation elongation factor promoter, P*tef1*, from *Aureobasidium pullulans*, and pENTRTERM (R. Johnson and C. Voisey, unpublished), which contains the *A. nidulans* glucoamylase terminator, to yield pDF7. The orientation
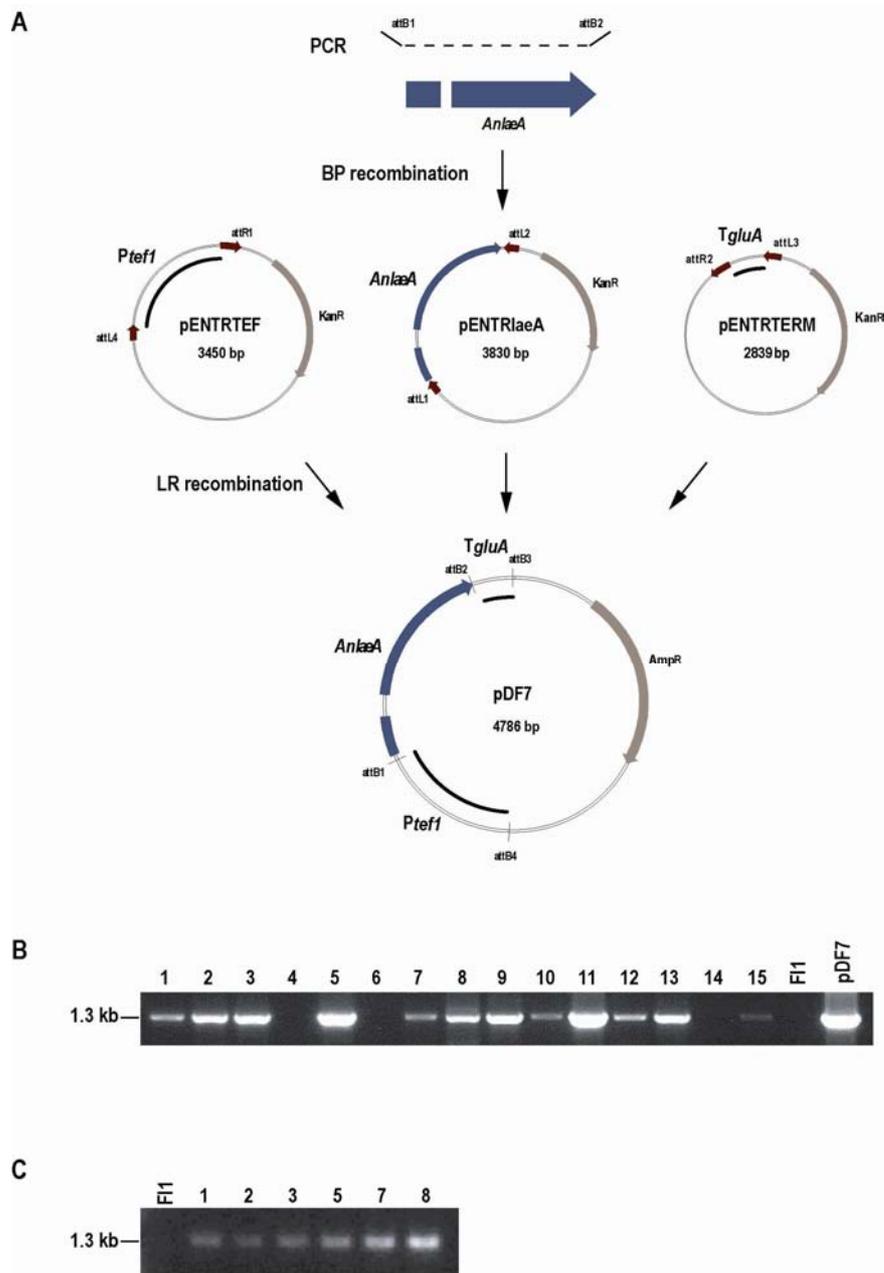
Figure 4.7. Preparation of an *laeA* expression construct and screening of transformants.
A. Schematic describing preparation of pDF7, the *laeA* construct for heterologous expression using Gateway recombination.
B. PCR screening of Hyg[R] transformants for integration of *laeA*.
C. RT-PCR analysis of selected transformants for *laeA* expression in cultured mycelium. Isolates chosen for further analysis *in planta* were 1 (DFM10), 2 (DFM11), 5 (DFM12) and 8 (DFM13). Isolates not containing *laeA* selected as controls were 4 (DFM14) and 6 (DFM15)

of each fragment was confirmed by restriction digests and sequencing across the junctions between the recombined fragments.

Five µg of pDF7 and 2.5 µg of pPN1688, which contains a hygromycin resistance cassette, were co-transformed into *E. festucae* Fl1 protoplasts as described in section 2.13, and selected on media containing hygromycin. After growth for one month, clonal isolates of hygromycin-resistant transformants were obtained by serial subculturing three times from the edge of the colony. Fifteen hygromycin resistant colonies were obtained.

Transformants were screened for integration of the expression construct by PCR (Fig. 4.7B). Genomic DNA was extracted from transformant mycelia by alkaline lysis (section 2.5.2) and used as template in a PCR with primers laeA-attB1 and laeA-attB2rev, which were used to amplify the *laeA* gene from *A. nidulans*. Of 15 transformants screened, an *laeA* product was amplified from 11, indicating integration of the expression construct. No difference in morphology or growth rate on solid or liquid media was observed between transformants and the parent isolate.

4.3.2.2 Phenotypic Characterisation of *laeA* Transformant Strains

In order to determine whether *laeA* was expressed in transformed *E. festucae* Fl1, RNA was extracted from transformant and wild-type mycelia grown in PD broth and analysed by RT-PCR. This analysis showed that each transformant was expressing the full-length *laeA* gene (Fig. 4.7C). RT-PCR using primers to *lpsB* and *ltmG*, the gene for the first commited step in indole diterpene biosynthesis, failed to amplify a product and *perA* was not amplified above low wild-type levels. Expression of *laeA* was therefore not sufficient to induce these genes in culture.

To determine whether *laeA* expression had an effect on expression of *eas* or *ltm* genes *in planta*, etiolated ryegrass seedlings were inoculated with wild-type and mutant strains. These plants were grown for approximately five weeks, after which *laeA* transformants showed a wild-type plant-interaction phenotype.

Figure 4.8. Real time RT-PCR analysis of *laeA*-expressing strains *in planta*
A. Expression of *laeA, lpsB, dmaW* and *ltmG* relative to *tubB*, deduced using the method of (Livak and Schmittgen, 2001).
B. Expression of each gene relative to wild-type (Fl1). *As wild type *E. festucae* does not express *laeA*, this expression is arbitrarily shown relative to DFM113.
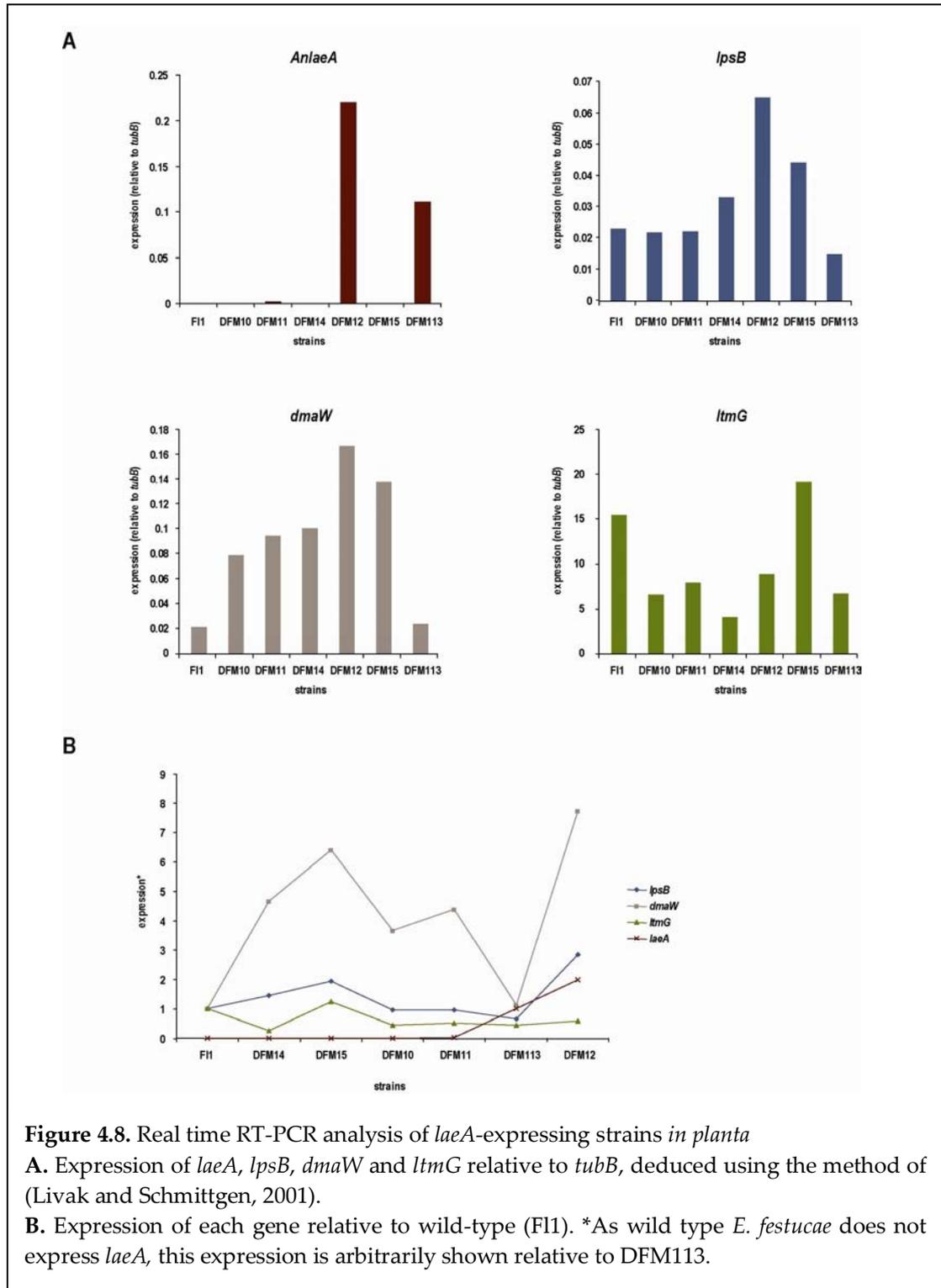
RNA extracted from three pooled plants infected with strains expressing *laeA* was analysed in triplicate by real-time RT-PCR for *laeA*, *dmaW*, *lpsB* and *ltmG* expression relative to *tubB* (Fig. 4.8). Two strains that were transformed with pPN1688 but not the *laeA* expression vector, DFM14 and DFM15, were included as

controls. This analysis showed that the level of expression of *laeA in planta* was much more variable than was apparent in culture. There was no correlation between the level of *laeA* expression and alkaloid gene expression. The strain expressing *laeA* at the highest level also expressed *dmaW* and *lpsB* the highest out of all the strains, however, DFM15, which does not contain the *laeA* gene, had the next highest expression of these genes and the second highest *laeA*-expressing strain contained the lowest levels of expression for these genes of the strains tested. From this experiment, there was thus no evidence that *laeA* could affect expression of *eas* or *ltm* genes in *E. festucae*.

5. Results

EAS Cluster Transposable elements

5.1 Transposable Elements Within the *EAS* Cluster

The first transposons described in epichloë were the Tahi and Rua (Maori for one and two) retrotransposon relics. These degenerate elements comprise large AT-rich tracts of sequence interrupting the *N. lolii LTM* gene cluster for indole diterpene synthesis (Young *et al.*, 2005). As described in Chapter 3, analysis of the sequence of the epichloë *EAS* cluster revealed the presence of highly AT-rich sequence at the end of each of the sequenced blocks. BLAST analysis of the AT-rich sequence downstream of *lpsB* in cluster one and upstream of *lpsA* in cluster three showed a weak match to a retrotransposon *pol* gene and a Tahi element respectively (sections 3.5 and 3.6), and a long terminal repeat (LTR) from a Rua retrotransposon had previously been shown to be found downstream of *dmaW* in *Neotyphodium* sp. Lp1 (Young *et al.*, 2006). It thus appeared that the *EAS* cluster may be interspersed with transposon platforms similar to the *LTM* cluster.

A detailed sequence analysis was undertaken of the AT-rich regions flanking the *EAS* sequences. These regions contained relics of multiple transposable elements (Fig. 5.1). Analysis of intergenic regions also revealed transposon sequences present between *lpsB* and *easE*, and between *easA* and *easG* (Fig. 5.1). In total ten different transposon sequences were identified, including autonomous and non-autonomous DNA transposons, and autonomous retrotransposons (including Tahi and Rua elements). A detailed analysis of each element is described in sections 5.2 – 5.5. In the course of characterising the integration sites of some of these elements, several other repeat sequences were identified. These were not characterised in any detail and were labelled with EFT (*Epichloë festucae* transposon) numbers.
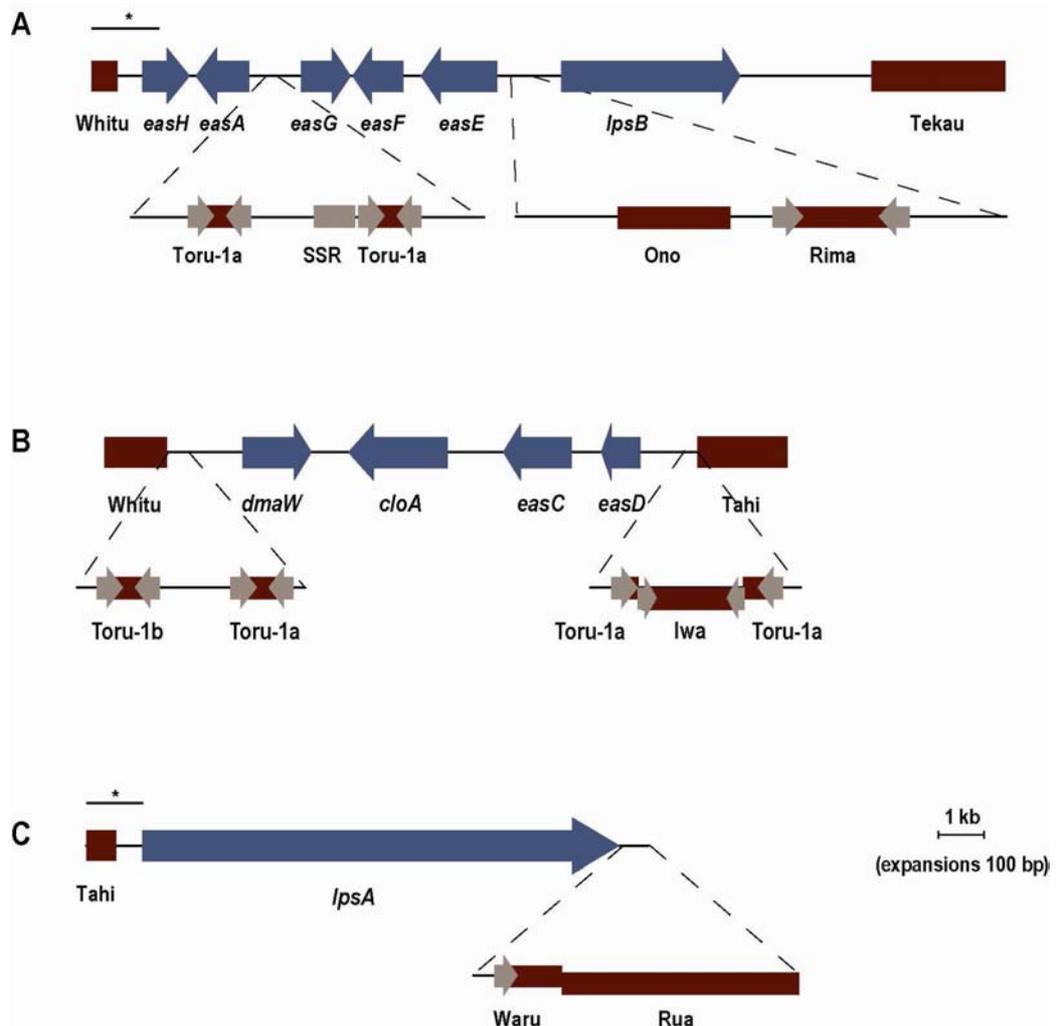
Figure 5.1. Location of transposons within the *EAS* cluster. DNA transposons are represented by dark brown rectangles with light brown inverted arrows, retrotransposons are represented by dark brown bars without arrows. Sequences are from either *N. lolii* Lp19 or *E. festucae* E2368, whichever had the more complete sequence for each cluster, supplemented with sequence from the other species where noted

A. *N. lolii* cluster one. *overlined sequence from *E. festucae* E2368
B. *E. festucae* E2368 cluster two.
C. *E. festucae* E2368 cluster three. *overlined sequence from *N. lolii* Lp19.

5.2 MITEs - Toru, Rima and Iwa

Miniature inverted-repeat transposable elements (MITEs) are short high-copy repeat sequences found in the genomes of many eukaryotes. They are characterised by their small size (typically 150 – 500 bp), high copy number, terminal inverted repeats (TIRs) and target site duplications (TSDs) (Feschotte *et al.*, 2002). MITEs are non-autonomous DNA transposons, that is, the transposase required for transposition is encoded on a parent autonomous DNA transposon (section 5.3). Parent autonomous elements can be identified based on shared TIR and TSD sequences which are specific to different transposases (Feschotte *et al.*, 2002). A parent element was recently identified for the fungal MITE *mimp* from *F. oxysporum*, which was shown to be mobilised by a *Tc-1*-like element, *impala* (Dufresne *et al.*, 2007). Only two fungal MITEs have been described, *mimp* and *guest*, an uncharacterised MITE-like element from *N. crassa* (Yeadon and Catcheside, 1995), suggesting they may be rare in fungi. Analysis of the *EAS* cluster however revealed three putative MITEs in association with these genes.

5.2.1 Toru

Manual analysis of the intergenic sequence between *easA* and *easG* revealed the presence of two short repeated sequences that had the hallmarks of a MITE (Fig 5.1). They are 82% identical, 139 and 141 bp in size respectively, have 61-bp imperfect TIRs and a putative AT TSD. The 24 bp between the TIRs is AT rich and has no similarity with any sequence in the public databases. The element was named Toru (Maori for three) following the convention adopted for naming the two retrotransposons in the *LTM* cluster called Tahi and Rua (one and two).

BLASTN analysis showed that Toru was also present upstream of epichloë *dmaW*, again in two copies (Fig. 5.2), and the *N. uncinatum LOL*-1 cluster where a single full-length copy is found 320 bp upstream of *lolU* and a solo TIR is present 1.1 kb upstream of *lolD*. A 26-bp fragment of a TIR is also associated with the B9 microsatellite in *Neotyphodium* sp. Lp1 (Moon *et al.*, 1999).
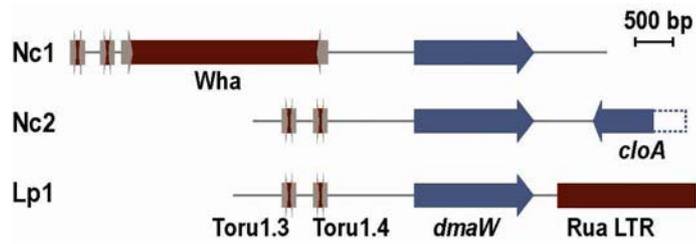
Figure 5.2. Sequence diversity at *dmaW* loci
Nc1, *N. coenophialum dmaW* locus 1 (accession AY259838); Nc2, *N. coenophialum dmaW*
locus 2 (accession AY259839); Lp1, *Neotyphodium* sp. Lp1 *dmaW* locus (accession
AY259837).
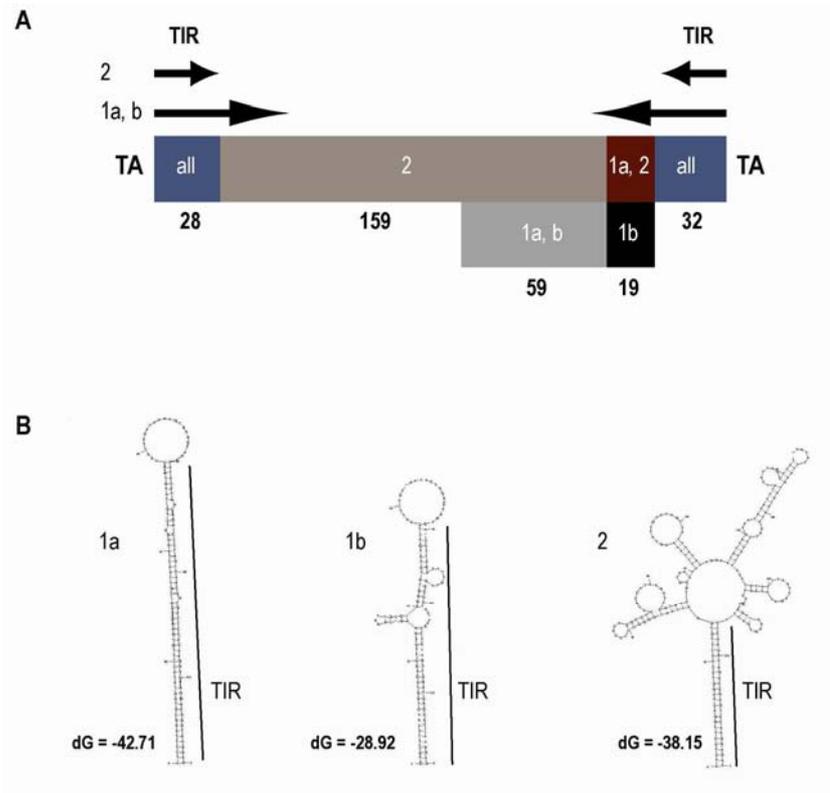
Figure 5.3. Toru structure.
A. Schematic representation of Toru family elements. Unique or shared regions are
shown as coloured rectangles. Toru elements containing these regions are labelled
inside boxes, lengths in bp of regions in respective consensus sequences are shown
below. TIRs for each element are shown above and TA TSDs are shown to each side.
B. MFold predicted secondary structure of the three Toru family elements

A 241-bp element related to Toru was found 489 bp upstream of *ltmG* in the *LTM* cluster of *N. lolii* and *E. festucae*. This sequence shared 28 bp of the left TIR and 51 bp of the right TIR and contained 159 bp of unique sequence between these repeats. This variant was named Toru-2 and the original Toru element was labelled Toru-1. The structure of these elements is shown graphically in Fig 5.3.

The *E. festucae* E2368 genome was searched for Toru sequences using the consensus sequence from an alignment of the identified Toru-1 elements. There were 214 matches with the default expect value of 10 and 154 matches with the most stringent value of 0.0001. When the 59 bp of unique sequence from Toru-1 elements was used to query the genome database, 78 matches were returned with an E value of 1e-05 or less. When the 159 bp of unique sequence from Toru-2 was used to query the genome database, 36 matches were returned with the same E value cut-off, a total of 114 matches to the combined elements. It may be that the extra sequences returned to the BLAST search with the full Toru-1 sequence are degenerate in these regions or possibly there are other variants of Toru not yet identified. Access to the actual genome sequence database that would allow the identification of any further variants through the use of computer algorithms was not available at the time of writing this thesis.

An alignment of 28 Toru-1 sequences was performed (Appendix E). This confirmed the presence of two variants of Toru-1 (1a and 1b) that differ in a 19 bp region within the Toru-1 right hand TIR (Fig. 5.3). Of the 28 aligned elements, 24 had a putative 2-bp AT TSD, three elements had a putative 3-bp TSD of TAT, TAA or ATA and one element did not appear to contain a sequence duplication at the target site.

MITEs have been found to integrate preferentially into genic regions in other organisms (Bureau and Wessler, 1994a, b; Bureau *et al*., 1996). To investigate whether this may also be true in epichloë, the integration site was determined for Toru-1 elements in 12 contigs, which contained the elements with the top BLAST matches to the *EAS* Toru-1 element consensus (Table 5.1). These contigs were

downloaded, the integration site was determined and the FGENESH predicted ORFs (trained on the *F. graminearum* dataset) or different repeat elements flanking the integrations were examined. Two contigs contained two Toru elements and so 14 elements were examined in total. Five elements were found in the intergenic region between two divergently transcribed predicted genes. Four further elements were located in the 5' region of a predicted gene with either the 3' end of a gene (×2) or another repeat element (×2) on the other side of the integration site. As mentioned above, Toru is also found intergenic between *easA* and *easG* and *lolU* and *lolP*, and upstream of *dmaW*. Toru elements were thus found in the 5' regions of 19 predicted genes and for 12 of these integration was within 1 kb of the predicted translational start. One further element was found between the 3' end of a gene and the contig end. The remaining four elements were found between two other repeat elements or a repeat element and the contig end. Although a small sample, it thus appears that Toru elements, like other MITEs, do preferentially integrate into genic regions and 5' regions in particular.

Table 5.1. Insertion sites of Toru elements in 12 contigs

Contig	Var	Position	Feature left Feature[a]	BLAST match (%ID, accession)	Distance (bp)	Feature right Feature[a]	BLAST match	Distance (bp)
535	1a	18689 – 18821	A.5.8692 (5')	*Cg* hypothetical protein (54%, XP_001220650) *Nf* Ser/Thr protein phosphatase (44%, XP_001264274)	680	A.6.8692 (5')	*Fg* hypothetical protein (31%, XP_391517) *Ac* glucosyl transferase (30%, XP_001275448)	1237
535	1b	26592 – 26725	A.7.8692 (5')	*Fg* hypothetical protein (51%, XP_390219) *An* developmental protein *fluG* (37%, XP_662423)	999	A.8.8692 (5')	*Sn* hypothetical protein (EAT88683)	630
338	1a	174 – 311	contig end	n/a	174	A.1.7822 (3')	*Fg* hypothetical protein (54%, XP_381615)	4184
180	1b	40895 – 41023	A.15.6720 (5')	*Ao* hypothetical protein (66%, BAE58473)	3366	Iwa	n/a	1258
180	1a	42775 – 42908	EFT-11		2	Iwa	n/a	2
1013	1a	6569 – 6710	A.2.6174 (5')	*Fg* hypothetical protein (64%, XP_390361) *Nf* DnaJ chaperone (40%, EAW20949)	2856	A.3.6174 (5')	*Fg* hypothetical protein (62%, XP_390360) *Os* MAP3K-like protein kinase (26%, ABF96605)	546
663	1a	1150 – 1299	EFT-12		226	Iwa	n/a	8
528	1a	4794 – 4927	*racA* (3')	*Ef* RacA (100%, BAF36499)	987	A.3.2399 (5')	*Cg* hypothetical protein (54%, EAQ90127)	212
89	1a	6462 – 6593	A.2.9023 (3')	*Hj* glucose transporter (78%, AAR23147)	704	A.5.2721 (5')	*Fg* hypothetical protein (51%, XP_382225)	2816
247	1a	8082 – 8216	Iwa	n/a	3	contig end	n/a	74
36	1a	3961 – 4096	EST[†] (5')	no match	526	A.1.7969 (5')	*Fg* hypothetical protein (77%, XP_388748) *Af* Rho GTP exchange factor (53%, XP_753782)	694
1560	*	668 – 762	Whitu	n/a	0	Iwa	n/a	469
818	1a	17486 – 17618	A.3.8957 (5')	*Fg* hypothetical protein (38%, XP_389928)	749	EFT-13	n/a	1966
746	1b	4236 – 4368	A.1.8895 (5')	*Nf* hypothetical protein (35%, XP_001264724)	1923	A.2.8895 (5')	*Fg* hypothetical protein (49%, XP_389035)	622

[a]features are either FGENESH CDS predictions from the *E. festucae* E2368 Gbrowse, previously-identified epichloë genes, or repeat elements. Where an EST was annotated but there was no FGENESH prediction, the EST was labelled as the feature. When the feature associated with a Toru element is a predicted gene, it is annotated as in either the 5' or 3' region, and the top BLASTP match, along with the top match with some indication of function where available, is given.

[†]EST CLS_CLiFpEfSpn_4a3_1_j15cLibkit5LD_E08

*truncated sequence, variant unable to be determined

Cg, Chaetomium globosum; Nf, Neosartorya fischeri; Fg, Fusarium graminearum; Ac, Aspergillus clavatus; An, Aspergillus nidulans; Sn, Staganospora nodorum; Ao, Aspergillus oryzae; Os, Oryza sativa; Hj, Hypocrea jecorina; Af, Aspergillus fumigatus

5.2.2 Rima

A second putative MITE, Rima (Maori for five) was identified by MEME analysis of the sequence between the *lpsB* and *easE* genes (Fig. 5.1). This element is 295 bp in size and has two 63-bp imperfect TIRs, a putative TA TSD and AT-rich, non-coding sequence between the TIRs (Fig. 5.4). No other copies of Rima were found in GenBank by BLASTN analysis.

When the *E. festucae* E2368 genome sequence was queried with the Rima sequence 10 sequences were matched with an E value less than 1e-05, indicating that the Rima element is found in considerably lower copy-number than Toru. All but one of the sequences, with at least one end of the Rima sequence in the database, contained the same putative 2-bp TA TSD found for the element located in the *EAS* cluster. This TSD is a signature of Tc1/*mariner* family transposases, a member of which is thus likely to have been responsible for Rima mobilization. Rima elements appear to be more degenerate than Toru elements with each of the copies containing many SNPs and short insertions or deletions (Appendix E). This may suggest that the Rima element was active in epichloë some time before Toru.

Analysis of the integration site for Rima elements (Table 5.2) did not reveal as marked a preference for 5' genic regions as was seen for Toru. Nevertheless, five out of the ten elements (including the element intergenic between *lpsB* and *easE*) were found in the 5' regions of at least one gene. Interestingly, three of the four elements found within one kb 5' of a predicted gene were found associated with

Figure 5.4. Rima structure.
A. Schematic representation of Rima. Lengths in bp of TIRs (light brown arrows) and internal sequence (dark brown bar) of the *EAS* Rima element are shown below. TA TSDs are shown to each side.
B. MFold predicted secondary structure of Rima
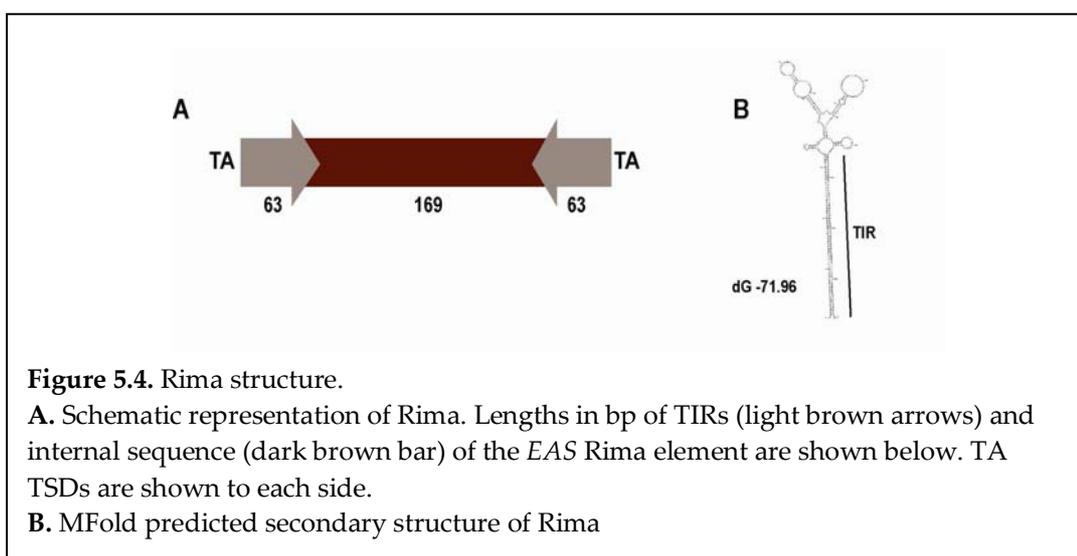
Table 5.2. Insertion sites of Rima elements

Contig	Position	Left Feature[a]	BLAST match	Distance	Right Feature[a]	BLAST match	Distance
52	76121 – 76338	A.21.8677 (5')	*Fg* hyp (41%, XP_381916) *Ac* cytochrome P450 (36%, XP_001275759)	570	A.22.8677 (5')	*Fg* hyp (69%, XP_383027) *At* fumarylacetoacetase (63%, XP_001215102)	680
420	1 – 263	contig end	n/a	0	A.1.8291 (5')	*An* hyp (66%, XP_001398268) *Ab* aminopeptidase (36%, CAC85297)	179
1372	5474 – 5838	contig end	n/a	321	A.2.290 (3')	*Mg* calmodulin-dependent protein phosphatase (72%, XP_367545)	456
946	6719 – 6985	A.3.9077 (5')	no hit	519	A.4.9077 (5')	no hit	915
647	367 – 553	EFT-18	n/a	0	A.1.2508 (5')	*Ac* NRPS (28%, XP_001270326) E 0.0	965
344*	26 – 90	contig end	n/a	26	EFT-11	n/a	0
140	24918 – 25152	A.9.6510 (3')	*Fg* hyp (51%, XP_386415) *An* hypercellular protein A (38%, AAP04416)	468	EFT-17	n/a	111
566*	12776 – 12874	EST[+] (3')	no hit	339	Iwa TIR	n/a	280
849*	23611 – 23684	A.5.8989 (3')	*Fg* hyp (79%, XP_389451) *Ac* cell differentiation protein Rcd1 (86%, XP_001272092)	4366	A.6.8989 (5')	*Fg* hyp (52%, XP_389453)	3233

[a]features are either FGENESH CDS predictions from the *E. festucae* E2368 Gbrowse, previously-identified epichloë genes, or repeat elements. Where an EST was annotated but there was no FGENESH prediction, the EST was labelled as the feature. When the feature associated with a Toru element is a predicted gene, it is annotated as in either the 5' or 3' region, and the top BLASTP match, along with the top match with some indication of function where available, is given.

[+]EST CLS_cLiFpEfI12na_5a3_1_h15cLibkit5LD_D08

*truncated sequence

Fg, Fusarium graminearum; Ac, Aspergillus clavatus; An, Aspergillus nidulans; Ao, Aspergillus terreuse; Mg, Magnaporthe grisea; Ab, Agaricus bisporus

putative secondary metabolite genes. Rima elements were found in the *EAS* cluster, as described above, and upstream of a putative NRPS gene (contig 647) and a putative cytochrome P450 gene (contig 52), which appears to be located in a biosynthetic gene cluster as there is another NRPS gene immediately downstream.

5.2.3 Iwa

A third putative MITE, termed Iwa (Maori for nine), was discovered by manual analysis of an insertion within a Toru-1 element at the right end of the *E. festucae* E2368 *EAS* cluster 2 (contig 1139, Fig. 5.1). Iwa is 238 bp long, with imperfect 29-bp TIRs and a putative 2-bp TA TSD (Fig. 5.5).

There were 253 matches to Iwa from a BLASTN search (Expect value = 10) in the *E. festucae* E2368 genome, slightly more copies than for Toru. Alignment of the top ten matches from BLASTN showed these to be significantly less degenerate than was seen for either Toru or Rima elements, with fewer SNPs, deletions or insertions. This may suggest more recent activity of Iwa than the other two MITEs. Alignment also showed each element to contain a putative 2-bp TA TSD, the same TSD as for Rima elements and again indicating mobilisation by a Tc1/*mariner* family transposase.

Analysis of the integration site of ten Iwa sequences showed fewer Iwa elements associated with either 5' or 3' regions of predicted coding regions than was observed for either Toru or Rima elements (Table 5.3). There is an Iwa insertion in the 5' upstream region of 7 (from a possible 20) predicted genes. Three of these insertions are within 1 kb of the translational start site.

Figure 5.5. Iwa structure.
A. Schematic representation of Iwa. Lengths in bp of TIRs (light brown arrows) and internal sequence (dark brown bar) of the *EAS* Iwa element are shown below. TA TSDs are shown to each side.
B. MFold predicted secondary structure of Iwa
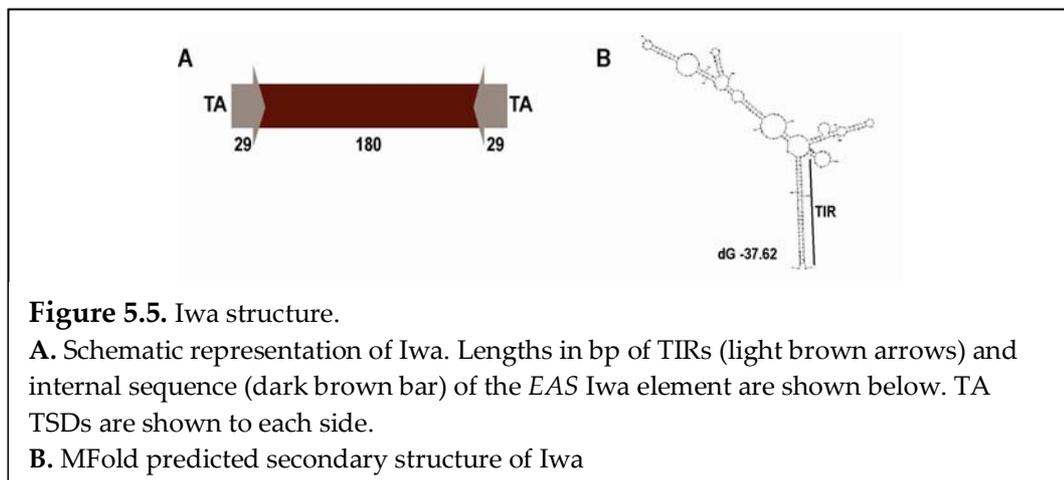
Table 5.3. Insertion site of Iwa elements

Contig	Position	Left Feature[a]	BLAST match	Distance	Right Feature[a]	BLAST match	Distance
118	70030 – 70266	A.18.6325 (5')	*Fg* hypothetical protein (70%, XP_386309) *Af* aspartyl aminopeptidase (59%, EAL85907)	1471	A.19.6325 (5')	*Mg* formate dehydrogenase (81%, AAW69358)	3258
1397	298 – 535	EFT-19		72	A.1.6506 (5')	*Nf* Phosphoglycerate mutase family protein (60%, XP_001261626)	248
948	27383 – 27601	A.9.9079 (5')	*Mg* hypothetical protein (31%, EDK00175)	1328	contig end	n/a	0
1281	920 – 1156	EFT-20		252	EFT-13	n/a	280
398	3959 – 4195	A.2.8172 (3')	*Fg* hyp (46%, XP_387553) *Nf* nucleosome assembly protein family (41%, EAW24343)	941	A.4.8172 (5')	*Mg* hypothetical protein (59%, EDK04331) *Ac* oligopeptide transporter (38%, EAW09074)	1483
961	1755 – 1991	EFT-21		137	A.3.9094 (3')	*Cp* Cot1 protein kinase (67%, CAH04535)	1345
34	29537 – 29780	EST[†] (5')	*Cg* hypothetical protein (33%, XP_001221518)	158	A.11.7837 (5')	*Fg* hyp, (50%, XP_389465) *Ma* cell wall glucanosyltransferase (49%, ABC65824)	617
663	1306 – 1549	Toru-1	n/a	7	A.1.8814 (3')	*Fg* hypothetical protein (75%, XP_386501) *Cp* glycosyl hydrolase (57%, ABA54914)	710
0	1801 – 2037	Rua	n/a	205	EFT-14	n/a	10
817	9617 – 9853	EFT-22	n/a	2	Waru-nui	n/a	146

[a]features are either FGENESH CDS predictions from the *E. festucae* E2368 Gbrowse, previously-identified epichloë genes, or repeat elements. Where an EST was annotated but there was no FGENESH prediction, the EST was labelled as the feature. When the feature associated with a Toru element is a predicted gene, it is annotated as in either the 5' or 3' region, and the top BLASTP match, along with the top match with some indication of function where available, is given.

[†]EST CLS_CLiFpEfSpn_52a2_1_m20cLibkit5LD_G10

Mg, Magnaporthe grisea; Cg, Chaetomium globosum; Nf, Neosartorya fischeri; Fg, Fusarium graminearum; Ac, Aspergillus clavatus; Af, Aspergillus fumigatus; Ma, Metarhizium anisopliae; Cp, Coccidioides posadasii

5.3 Autonomous DNA Transposons – Wha and Waru

DNA, or type 2, transposons have TIRs and TSDs, similar to MITEs, which are derived from them (Casacuberta and Santiago, 2003). However, the autonomous DNA transposons encode the transposase required for mobilisation (Fig. 5.6) and are typically found in low copy-number, due to their mechanism of transposition via a DNA intermediate (Casacuberta and Santiago, 2003). Transposons from three superfamilies of autonomous DNA transposons, Tc1/*mariner*, *hAT* and *mutator*, have been identified in fungi (Daboussi and Capy, 2003).

5.3.1 Wha

Analysis of "extra" sequence between *N. coenophialum dmaW*-1 and the Toru MITEs, compared with *dmaW*-2, *Neotyphodium* sp. Lp1 and *E. festucae* E2368, revealed a putative Mutator-like element relic, Wha (four) (Fig. 5.2). Wha has 95-bp TIRs and an 8-bp putative TSD. The 2413 bp between the putative TIRs was highly degenerate (75% AT) and shared 32% identity to a 312-aa putative protein (CHGG_06856) from *Chaetomium globosum*, with several closely related proteins in the *C. globosum* genome. When submitted for BLASTX analysis itself, this sequence had 31% identity with a transposase from the *Hop* Mutator-like element in *Fusarium oxysporum* (Chalvet *et al.*, 2003).
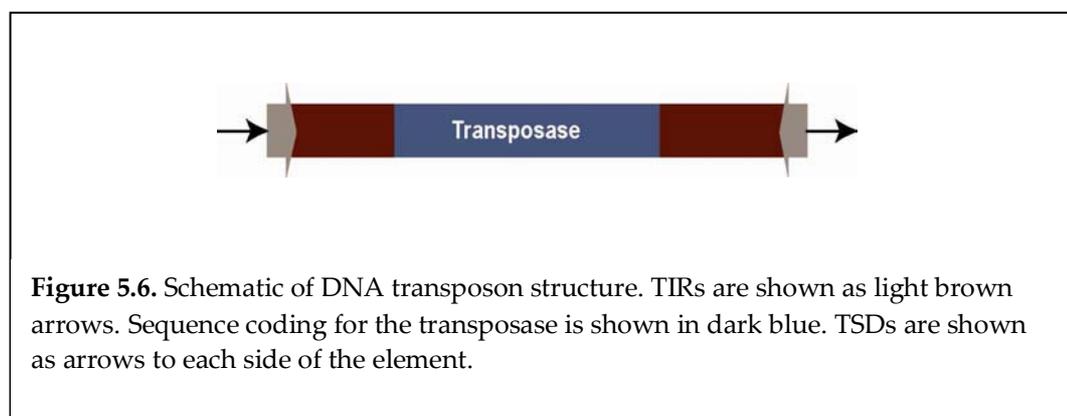
Figure 5.6. Schematic of DNA transposon structure. TIRs are shown as light brown arrows. Sequence coding for the transposase is shown in dark blue. TSDs are shown as arrows to each side of the element.

There are no full length Wha sequences in the *E. festucae* E2368 genome assembly, although the entire contig 1978 consists of an almost-full-length copy, with 38 bp of the TIRs truncated at each end. Seven other contigs contained varying length copies of one end of Wha at one end of the contig. Alignment of all of the Wha sequence fragments revealed considerable degeneration of the element, including many stop codons, SNPs and 8 deletions in one or more element of 1 – 16 bp and a larger deletion in contig 566 of 113 bp. Several of the smaller deletions were shared between two or more copies, however a shared deletion at one site did not necessarily correlate with a shared deletion at a different site, indicating that these may have been de novo deletions at the same site in different elements, rather than arising from amplification of a deleted element. Analysis of the sequence adjacent to the Wha fragments revealed that the sequences on contigs 68 and 680, and 1909 and 1609 respectively are likely to be ends of the same elements as they share the 8 bp immediately adjacent to the ends as direct repeats, which are likely to be TSDs (Fig. 5.7).

Analysis of the integration site of the elements revealed that the element found at the end of contig 566 was likely to have inserted within an ORF coding for a predicted membrane protein (Table 5.4). BLASTX analysis of the sequence adjacent to integration of Wha in other contigs revealed the remainder of this ORF on contig 803, suggesting that these were ends of the same element that had integrated within this ORF. Of all the integration sites examined for each of the different transposons in this study, this was the only evidence of an element disrupting an ORF.

Figure 5.7. TSDs are shared between some Wha end sequences. Alignment of left (**A**) and right (**B**) ends of Wha elements is shown. TSDs are underlined. Sequences are labelled with associated gene or contig.
Nc, *N. coenophialum*; Ef, *E. festucae* E2368

Table 5.4. Insertion sites of Wha elements

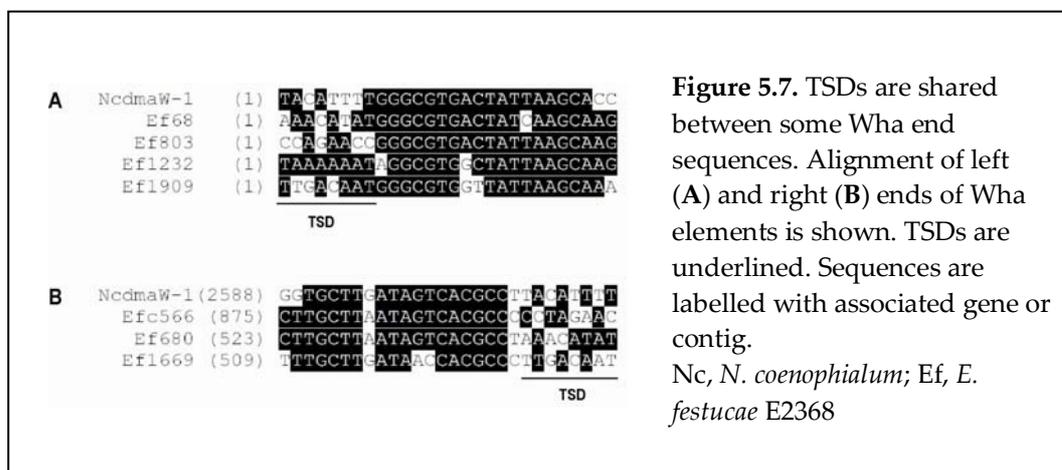| Contig (size)	Position	Adjacent feature[a]	BLAST match	Distance
1978 (2553)	1 – 2553	n/a	n/a	n/a
1232 (8750)	7911 – 8750	A.3.6370 (5')	*Fg* hypothetical protein (63%, XP_390027)	431
			Nf AAA family ATPase (47%, XP_001258398)	
1909 (3570)	2175 – 2839	none	n/a	n/a
68 (14750)	14157 – 14750	A.29.8830 (5')	*Fg* hypothetical protein (53%, XP_388120)	1277
			Ac cell wall glucanase (48%, EAW06867)	
680 (27703)	1 – 545	A.9.2541 (3')	*Fg* hypothetical protein (76%, XP_388121)	1284
			Ac palmitoyl CoA transferase subunit (57%, XP_001268294)	
566 (29911)	1 – 897*	A.9.8723 (ORF)	*Af* DUF895 membrane protein (66%, EAL85419)	1
803 (35664)	34950 – 35443*	unannotated	*Af* DUF895 membrane protein (71%, EAL85419)	1

[a]features are FGENESH CDS predictions from the *E. festucae* E2368 Gbrowse. These were annotated as in either the 5' or 3' region, and the top BLASTP match, along with the top match with some indication of function where available, is given.

*These Wha element ends are likely to represent each end of the same element, which has interrupted an ORF

Fg, Fusarium graminearum; Nf, Neosartorya fischerii; Ac, Aspergillus clavatus; Af, Aspergillus fumigatus

5.3.2 Waru

A fragment of a second putative autonomous type 2 element, Waru (eight) was identified by BLASTN analysis as a repeat sequence downstream of *lpsA* on contig 1183 of the *E. festucae* E2368 genome sequence assembly (Fig. 5.1). This sequence was interrupted by a Rua retroelement, however, a full-length element was identified on contig 213. Waru is 3076 bp in length and contains 34-bp TIRs and a putative 3-bp TSD (TTA for contig 213, TAA (one end only) downstream of *lpsA*). This putative T(T/A)A TSD suggests Waru belongs to the PIF/*harbinger* superfamily of transposons, transposases from which leave this signature duplication at the target site. This family of transposons had not previously been identified in fungi and this was reflected by BLASTX analysis, with the Waru sequence showing 28 – 37% identity over 115 – 184 amino acids of many transposases from various plant transposons. No fungal transposases were present in the list of BLAST matches. As for Wha, the Waru transposase sequence contained many stop codons and is thus unlikely to be functional.

A query of the *E. festucae* E2368 genome sequence showed 88 sequences matching Waru with an E value of 1e-05 or less, a high copy number for a type 2 transposon. Further analysis revealed that many of the copies had undergone large deletions, several leaving just a right or left end, which may have arisen from recombination

between elements. At least 13 – 25 full-length Waru copies are expected to be present in the genome based on analysis of the BLAST results. Of the internally deleted sequences containing both TIRs some appear to have been transposed as more than one copy is present in the genome. However, no evidence of a MITE derived from Waru was observed as none of the short internally deleted copies predominated and none were amplified to high-copy number.

5.4 Retrotransposons – Ono, Whitu and Tekau

Retrotransposons, or type 1 transposons, are the most widespread transposable element in eukaryotic genomes due to the amplification of RNA intermediates (Casacuberta and Santiago, 2003). They are divided into two broad divisions, the long terminal repeat (LTR) retroelements and non-LTR elements (Fig. 5.8), both of which are represented in fungi (Daboussi and Capy, 2003). LTR retrotransposons are flanked by long repeats of ~100 bp to several kb and contain two genes, *gag* and *pol*. The *gag* gene codes for a structural protein while *pol* codes for a polyprotein, which is cleaved into an integrase, protease, reverse-transcriptase and RNaseH, which are required for transposition (Casacuberta and Santiago, 2003). The order that these proteins are encoded on the *pol* gene determines whether an LTR retroelement belongs to the Ty1/*copia* or Ty3/*gypsy* families (Fig. 5.8).

Non-LTR elements are classified as either long or short interspersed nuclear elements (LINEs or SINEs), which are transcribed from internal promoters. LINEs contain *gag* and *pol* genes while SINEs use LINE transpositional machinery to mobilise (Casacuberta and Santiago, 2003). Although the non-autonomous SINEs use autonomous element proteins, they are not related in the way that MITEs are to DNA transposons and are usually derived from retroposed tRNA genes (Casacuberta and Santiago, 2003).
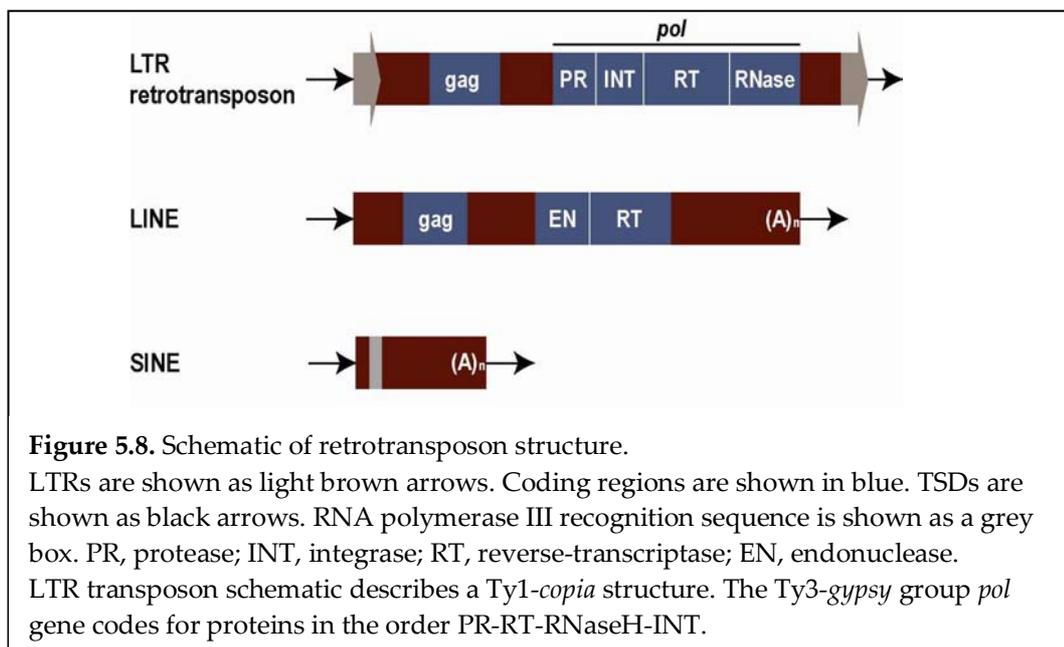
Figure 5.8. Schematic of retrotransposon structure.
LTRs are shown as light brown arrows. Coding regions are shown in blue. TSDs are shown as black arrows. RNA polymerase III recognition sequence is shown as a grey box. PR, protease; INT, integrase; RT, reverse-transcriptase; EN, endonuclease.
LTR transposon schematic describes a Ty1-*copia* structure. The Ty3-*gypsy* group *pol* gene codes for proteins in the order PR-RT-RNaseH-INT.

5.4.1 Ono

A repeat sequence 97 bp from Rima, intergenic between *lpsB* and *easE* (Fig. 5.1), was identified by MEME analysis as similar to a sequence upstream of *ltmM*. This element was named Ono (six) and was 241 bp long and 82% and 66% identical to sequences upstream of *ltmM* and *ltmE* in the *LTM* gene cluster respectively. Of the three Ono elements, only the *ltmM*-associated element had a putative TSD of 7 bp. This element had been identified previously as a 115-bp sequence shared between the *ltmM* and *ltmE* promoter regions and was described as possibly being a relic of a SINE (Young *et al.*, 2006). While Ono contains a possible RNA polymerase III recognition sequence, another key signature of SINEs, the presence of a polyA or di- or tetranucleotide repeat at the 3' end of the sequence, was not present in any of the three copies.

When the *E. festucae* E2368 genome sequence was queried with Ono, 101 matches were returned with an E value less than 1e-05. One of the contigs containing Ono (contig 12) had two copies in tandem, separated by 5 kb of AT-rich sequence and with a 5-bp direct repeat found immediately before and after the two copies respectively. When the sequence between the two copies was analysed by BLASTX, a weak match was returned to 545 amino acids from a retrotransposon polyprotein from *Staganospora nodorum* (accession CAB91877). This, combined with a probable TSD, revealed Ono to be a solo LTR, rather than a non-autonomous transposable element. These solo LTRs are presumably derived from recombination between the two LTRs of a single element, which would retain a TSD as seen upstream of *ltmM*, or between two different elements. The full-length element was thus labeled Ono-nui, (nui = large).

The Ono-nui sequence on contig 12 is the only full-length sequence in the genome build, the remaining matches are all either truncated at each end on small contigs, or found truncated at one end of a contig. Alignment of several Ono-nui fragments, along the length of Ono-nui, showed a considerable number of SNPs but very few single-nucleotide deletions or insertions and no larger deletions as seen in Wha elements.

5.4.2 Whitu

Whitu (seven) is an LTR retrotransposon identified by BLASTN as partially repeated sequences upstream of both *dmaW* and *easH* that were possibly LTRs, as they showed no similarity to other types of transposon. There were no contigs in the *E. festucae* E2368 genome that contained both putative LTRs meaning that if Whitu was indeed an LTR retrotransposon, no full-length sequences were present in the database. A full-length Whitu consensus sequence was therefore pieced together by placing overlapping fragments into contigs until a full-length element was re-constituted (Fig. 5.9). Each section of the Whitu contig contained at least two sequences apart from two single-sequenced areas of 100 bp and 340 bp. Several of the Whitu sequences used contained other elements inserted within them and these were removed from the consensus sequence.

The full-length Whitu is 8939 bp in length with 119-bp LTRs. As for other autonomous elements studied, the Whitu sequence is highly AT-rich and there are many SNPs between the different sequences. No protein identity was able to be determined by BLASTX analysis, however, Whitu is very similar in both LTR (Fig 5.10) and total length to Tahi elements and is ~30% identical at the nucleotide level. Tahi and Whitu are thus likely to be part of a family of LTR retroelements in epichloë.
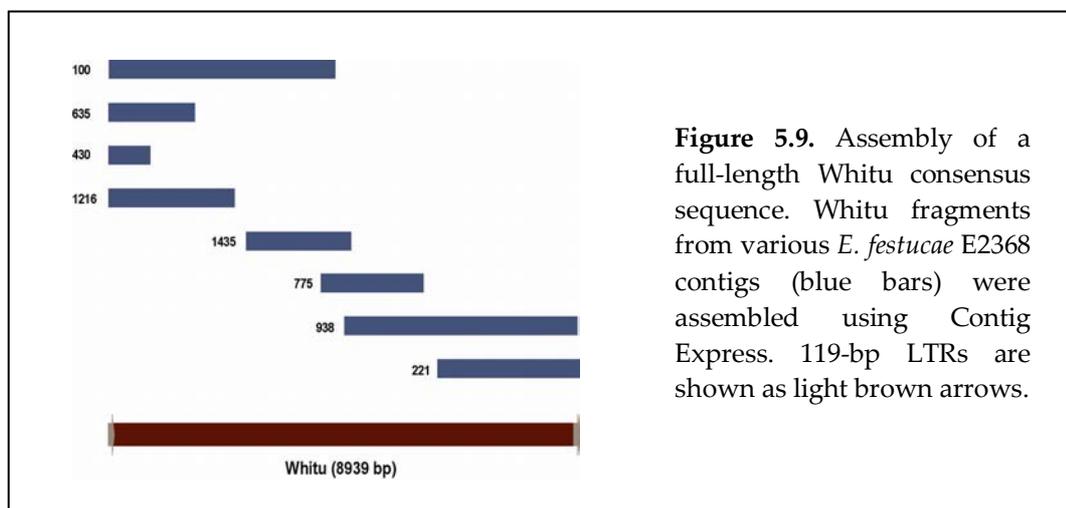
Figure 5.9. Assembly of a full-length Whitu consensus sequence. Whitu fragments from various *E. festucae* E2368 contigs (blue bars) were assembled using Contig Express. 119-bp LTRs are shown as light brown arrows.

Figure 5.10. Alignment of Whitu and Tahi LTRs. Whitu LTR sequences are labelled with the contig number on which they are found. The Tahi LTR sequence is from the *N. lolii LTM* cluster (accession AY742903)

5.4.3 Tekau

Tekau (ten) is predicted to be an LTR retrotransposon found downstream of *lpsB* in *EAS* cluster 1 (Fig. 5.1). There is a putative LTR with many repeats in the *E. festucae* E2368 genome and the AT-rich sequence downstream of this putative LTR contains a weak BLASTX match (3e-16) to a *pol* protein from the *skippy* retroelement from *F. oxysporum* (Anaya *et al.*, 1995), a member of the Ty3/*gypsy* superfamily. No contigs were identified containing two copies of the putative LTR, which would indicate a full-length copy of Tekau, and although there were many fragments matching different sections of the Tekau sequence, these were all very small. Contigs containing less than ~2 kb were not available for download from the genome database at the time of writing and thus a full-length Tekau sequence could not be generated from overlapping sequences as was possible for Whitu.

Upstream of the putative Tekau LTR were other sequences with multiple repeats in the *E. festucae* genome. As for Tekau, these were too degenerate to be able to be characterised in any detail.

5.5 Repeat Induced Point-Mutation (RIP) Analysis

All of the ten transposable elements associated with the *EAS* cluster appeared to be degenerate relics. The Tahi and Rua sequences associated with the *LTM* cluster were also observed to be degenerate and the AT-richness of these elements was proposed to be evidence of RIP (Young *et al.*, 2005), which results in C:G to T:A transitions. RIP is a fungal mechanism first discovered and best characterised in *Neuropora crassa*, which disables repeated genomic sequences (Cambareri *et al.*, (1989), reviewed in Selker, (2002)). The genome is scanned pre-meiosis for repeats, which are then mutated by converting cytosine residues to thymines, resulting in guanine to adenine transitions on the complementary strand. Cytosine and histone methylation is known to be important (Freitag *et al.*, 2002; Selker *et al.*, 2002) but the exact mechanism for conversion of the C to T mutations is not known. While RIP only occurs during sexual reproduction, evidence of RIP can be ascertained from previously RIPed sequences based on the sequence context of SNPs between copies (Margolin *et al.*, 1998). This analysis was performed on each of the transposons characterised in this chapter.

5.5.1 Epichloë Transposons Have Been Subject to RIP

In *N. crassa*, RIP occurs on cytosines preferentially at the CpA context (Grayburn and Selker, 1989). The ratio of TpA (resulting from CpA conversion) to ApT (a dinucleotide with the same composition) and CpA + TpG to ApC + GpT is thus indicative of whether or not RIP has occurred for a given sequence (Margolin *et al.*, 1998). These ratios were determined for a representative of each class of transposons identified in this study and also for the entire *EAS* cluster sequence with transposons removed (Table 5.5). (Galagan *et al.*, 2003) suggest that a ratio of TpA/ApT of greater than 2.0 or CpA + TpG/ApC + GpT less than 0.7 is indicative of RIP mutation in *N. crassa*. Each of the elements but Rima satisfy at least one of these criteria while the *EAS* cluster sequence without transposons does not. The degenerate transposons can thus be classified as having undergone RIP. For the RIP degenerated sequences found in *N. lolii*, an asexual species, the mutations are

presumed to have occurred in the sexual progenitor strain. In *N. crassa* RIP only occurs on repeat sequences of greater than ~400 bp, it would thus be unexpected for the MITEs Toru and Iwa to have undergone this process. However, it would seem from this analysis that in *E. festucae* RIP may function on smaller sequences than is possible in *N. crassa*.

5.5.2 De-RIP Wha and Ono-nui and *E. festucae* RIP Sequence Preference

Different ascomycetes in which RIP has been observed, either experimentally or by sequence analysis, appear to have C to T conversion at different sequence contexts in terms of the base 3' to the cytosine (Grayburn and Selker, 1989; Idnurm and Howlett, 2003; Clutterbuck, 2004). None of the transposon families identified contained copies in the *E. festucae* E2368 genome that were not degenerated. Two of the larger elements, Wha and Ono-nui, for which RIP can confidently be asserted to have taken place, were thus "deRIPed", similar to the analysis done by Attard *et al.*, (2005) with the *Leptosphaeria maculans Pholy* retrotransposon. Transposon sequences were aligned and wherever at least one of the sequences contained a C rather than T, or G rather than A, this base was inserted into the consensus sequence, resulting in a sequence that was "deRIPed" and likely to be more similar to the progenitor sequence.

Table 5.5. RIP indices for transposons associated with the *EAS* cluster

element	size	stop codons[a]	CpA*	TpG*	TpA*	TpA/ApT	CpA +TpG/ ApC+GpT
Toru-1a	141	9.2	0.67	0.80	2.05	2.02	0.56
Wha	2553	11.2	0.39	0.43	2.97	1.58	0.67
Rima	291	11.2	0.98	0.65	2.12	1.63	1.0
Ono-nui	5435	10.5	0.27	0.30	2.91	1.64	0.47
Whitu	8939	10.1	0.07	0.08	2.92	1.50	0.15
Waru	3075	10.5	0.37	0.32	2.91	1.50	0.55
Iwa	234	11.1	0.34	0.48	3.09	2.11	0.38
Tekau	4038	10.8	0.07	0.10	3.57	1.83	0.11
EAS cluster	36612	4.7	1.10	1.13	0.91	0.83	1.21

*ratio of observed to expected
[a]per 100 bp

Table 5.6. RIP sequence context preference summed for both strands[a]

element	CpA -> TpA	CpG -> TpG	CpC -> TpC	CpT -> TpT	non C ->T SNP
Wha	23	12	12	22	17
Ono-nui	96	40	21	87	20
combined	119	52	33	109	37

[a]Derived from comparison of Wha sequence from E. festuae E2368 contig 1978 and Ono-nui sequence from contig 12 with "deRIPed" sequences

Table 5.7. Ascomycete RIP sequence context preference

Species	3' base preference[†]
E. festucae	A>T>>G>C
N. crassa	A>>T>G>C
A. nidulans	G>A>>C≈T
Leptosphaeria maculans	G≈A
F. oxysporum	A>G
M. grisea	T>A>>C>G

[†]Preferences for base 3' of mutated cytosine from fungi other than *E. festucae* were taken from (Clutterbuck, 2004)

The deRIP consensus sequence for both Wha and Ono-nui contained fewer stop codons than degenerate copies; 307 reduced to 276 and 572 to 545 respectively. DeRIPing improved the BLASTX match for both elements to the top match (E values 1e-13 reduced to 3e-34 and 4e-26 to 2e-28 respectively) and included a direct match to the *F. oxysporum Hop* transposase (E value 2e-08) for the deRIP Wha, which was not observed in the list of matches for the degenerate Wha sequence. Although the BLAST matches were still weaker than might be expected for a non-degenerate element, the improvement in amino acid identity over longer stretches

of protein alignment and reduction in stop codons suggested that the deRIP sequences more accurately represented the original sequences for both elements.

Each sequence was compared with a representative degenerate copy and the base 3' of the mutated cytosine was determined on each strand (Table 5.6). Total C to T transitions for both elements showed 119 CpA to TpA, 109 CpT to TpT, 52 CpG to TpG and 33 CpC to TpC transitions (Table 5.6). This was a similar RIP site preference to *N. crassa* (Table 5.7), however, preference for CpG may be underestimated as mutation of the cytosine in either strand changes the context of the other.

6. Results

Comparative Analysis of the *EAS* Cluster

6.1 Taxonomic Distribution of *EAS* Genes

The ability to synthesise ergovaline is discontinuous among epichloë species (Clay and Schardl, 2002). In order to explore whether the presence of *eas* genes correlated with ergovaline biosynthesis, a selection of epichloë strains were tested for the presence of *eas* genes. Diagnostic PCR was performed with genomic DNA extracted from *E. festucae* Fl1, *E. typhina* E8 and several strains of *Neotyphodium* from the AgResearch endophyte collection (Table 6.1). These strains were tested for the presence of *dmaW*, *cloA*, *lpsA* and *lpsB*, each of which has been functionally characterised and is found on a different transposon-flanked cluster. The *tef1* gene was amplified from each sample as a control for template quality. Products were amplified for either all or none of the genes examined, indicating the presence of each cluster or no clusters.

The presence of *eas* genes correlated with the ability to produce ergovaline with one exception, *N. lolii* AR1, which contained each of the four genes but does not produce ergovaline. To determine whether the reason *N. lolii* AR1 does not produce ergovaline was due to the absence of *eas* genes other than those tested, this strain, along with the model strains *E. festucae* Fl1, *N. lolii* Lp19 and *Neotyphodium* sp. Lp1, were further examined for the presence of the remaining known *eas* genes. This analysis showed that each gene in the *EAS* cluster is present in these strains.

Table 6.1. Taxonomic distribution of *eas* genes

Strain	Parentage*	Ergovaline Production[a]	dmaW	cloA	lpsA	lpsB	easA	easC	easD	easE	easF	easG	easH
E. festucae Fl1	N/A	+	+	+	+	+	+	+	+	+	+	+	+
E. typhina E8	N/A	-	-	-	-	-	-	-	-	-	-	-	-
N. lolii AR1	Ef	-	+	+	+	+	+	+	+	+	+	+	+
N. lolii AR40	Ef	-	-	-	-	-							
N. lolii AR47	Ef	-	-	-	-	-							
N. lolii AR59	Ef	+	+	+	+	+							
N. lolii Lp11	Ef	+	+	+	+	+							
N. lolii Lp19	Ef	+	+	+	+	+	+	+	+	+	+	+	+
N. sp. Lp1 (LPTG-2)	Et × Ef	+	+	+	+	+	+	+	+	+	+	+	+
N. coenophialum AR512	Et × Ef × Eb	+	+	+	+	+							
N. coenophialum AR542	Et × Ef × Eb	-	-	-	-	-							
N. coenophialum AR565	Et × Ef × Eb	+	+	+	+	+							
N. coenophialum Tf5	Et × Ef × Eb	+	+	+	+	+							
N. sp. AR553 (FATG-2)	Ef × Eb	+	+	+	+	+							
N. sp. AR501 (FATG-3)	Et × Eb	-	-	-	-	-							
N. sp. AR507 (FATG-3)	Et × Eb	-	-	-	-	-							
N. sp. AR510 (FATG-3)	Et × Eb	-	-	-	-	-							
N. uncinatum Fp1	Et × Ebr	-	-	-	-	-							

*closest sexual ancestors to asexual species. Ef, *E. festucae*; Et, *E. typhina*; Eb, *E. baconii*; Ebr, *E. bromicola*

[a]determined by B. Tapper

6.2 Chemistry and Expression Analysis

E. festucae Fl1, *N. lolii* Lp19, *N. lolii* AR1 and *Neotyphodium* sp, Lp1 contained all of the genes predicted to be required for ergovaline synthesis. To determine whether any intermediate ergot alkaloids are produced in *N. lolii* AR1 and whether there are any differences in intermediate profiles between the strains, each strain was tested by tandem mass spectrometry. Each strain was inoculated into clonal host plant material (perennial ryegrass cv. Nui) and single pseudostems were analysed. *E. festucae* Fl1 tillers examined were uninfected, as frequently occurs with inoculations of clonal plant material (W. Simpson, pers. com.). Thus three strains were further analysed. This showed that there were no ergot alkaloids detectable in duplicate samples of *N. lolii* AR1. Analysis of peramine levels in each strain showed that AR1 contained levels in the normal range, comparable with the other two strains tested.

The predominant difference between the *N. lolii* Lp19 and *Neotyphodium* sp. Lp1 samples was the substantial difference in overall ergot alkaloid synthesis between the strains with *Neotyphodium* sp. Lp1 producing between ~3 – 5 fold increased levels (Fig. 6.1). A striking exception was in the synthesis of chanoclavine with 33 times higher levels detected in the *Neotyphodium* sp. Lp1 sample. Although standards were not available and the data are thus only semi-quantitative with respect to comparison between different intermediates, which have different proton affinities and abilities to ionise, there was also an exceptionally high amount of chanoclavine relative to other early intermediates within *Neotyphodium* sp. Lp1 (Fig 6.1) and to a lesser extent in *N. lolii* Lp19. Fluorescence HPLC has previously shown chanoclavine to be the most abundant ergot alkaloid in *Neotyphodium* sp. Lp1 symbiota (Panaccione 2005; Panaccione *et al.*, 2006). Comparisons with ergovaline and lysergic acid amides are not possible without authentic standards. The other notable difference between the two species was the ratio of ergovaline to ergovalinine, which is higher in *N. lolii* Lp19.

A study using the same plant material but with milled, pooled pseudostems performed by B. Tapper and A. Koulman (unpublished) showed a very similar profile of ergot alkaloids between the strains compared to this study, one exception being a relatively higher amount of ergine in the *N. lolii* Lp19 sample.

Figure 6.1. LC-MS/MS analysis of *Neotyphodium* strains. Estimates of accumulation of metabolites are in arbitrary units relative to ergotamine internal standard, corrected for sample dry weight, but uncorrected for differences in mass spectrometric response and isomerisation of the standard. (Note different scales for each graph.)
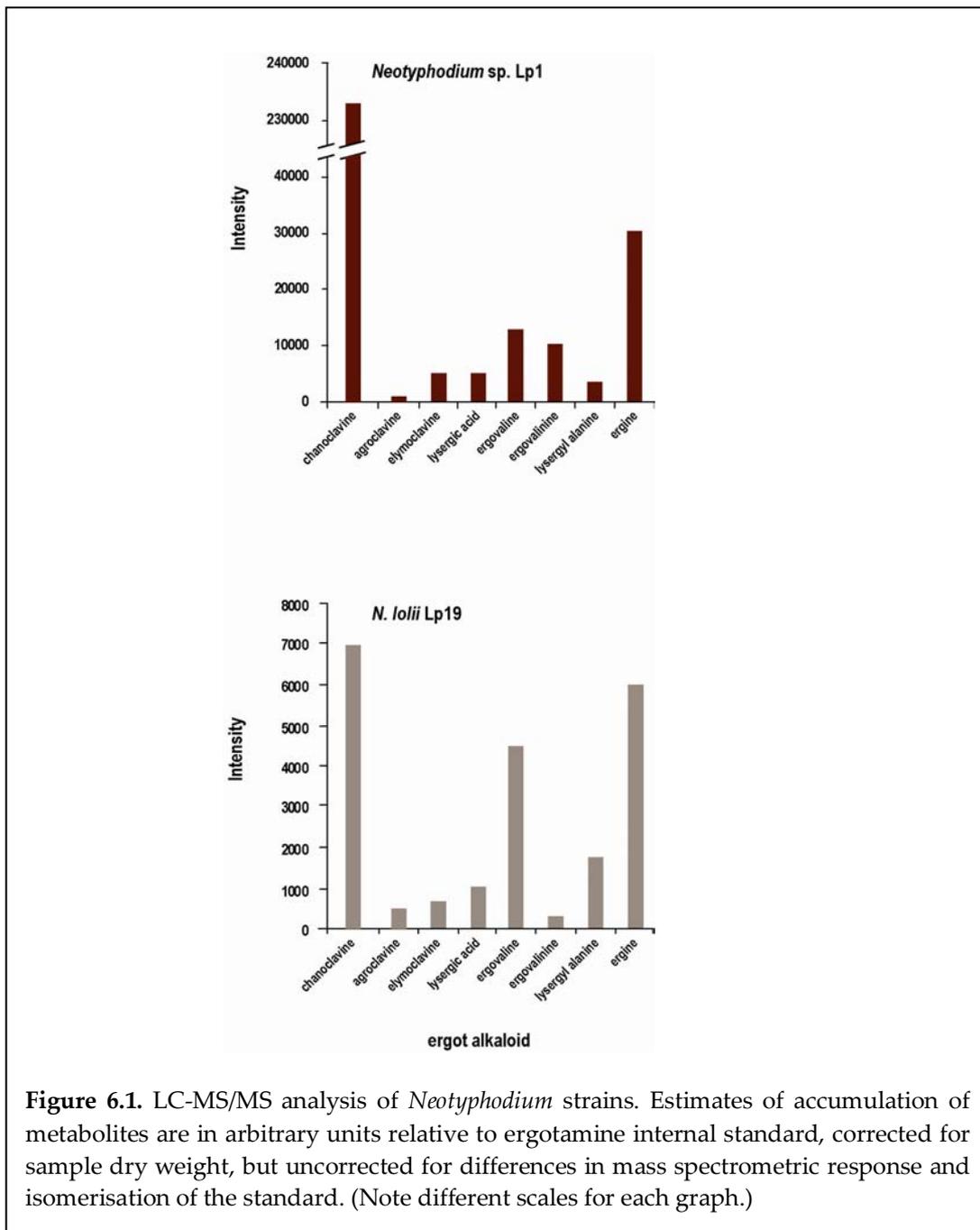
To determine whether the differences in the levels of ergot alkaloids between the strains was due to gene expression, RNA was extracted from pooled pseudostems removed from the clonal plant material at the same time as those for chemical analysis. Real time RT-PCR was performed and expression of 8 of the 11 genes (*dmaW, easC, easD, easE, easF, easG, lpsB* and *lpsA*) determined relative to *tubB* (Fig 6.2).

Overall, gene expression levels correlated with the levels of ergot alkaloids detected by LC-MS/MS. There was no expression of most of the tested *N. lolii* AR1 genes and only extremely weak expression of *easE*, *easG* and *lpsA*. Overall expression was higher in *Neotyphodium* sp. Lp1 than *E. festucae* Fl1, which was somewhat higher than *N. lolii* Lp19. NRPS genes tended to be expressed at a lower level than other genes. Ratios of expression between genes likely to be required for intermediate steps in the pathway were variable for each of the strains although *easG* gene expression was considerably higher than other genes and particularly in *Neotyphodium* sp. Lp1 where *easG* expression was 20 fold higher than *easF*, the next most highly expressed gene. This extremely high expression rate correlated with a very high level of chanoclavine in *Neotyphodium* sp. Lp1. In *N. lolii* Lp19, chanoclavine levels were high but not to the same extent, again correlating with the level of *easG* expression in this organism, which was not as high as in *Neotyphodium* sp. Lp1. While circumstantial, these data suggest that the *easG* gene product may have a key role in the early steps in the ergot alkaloid pathway leading to chanoclavine.
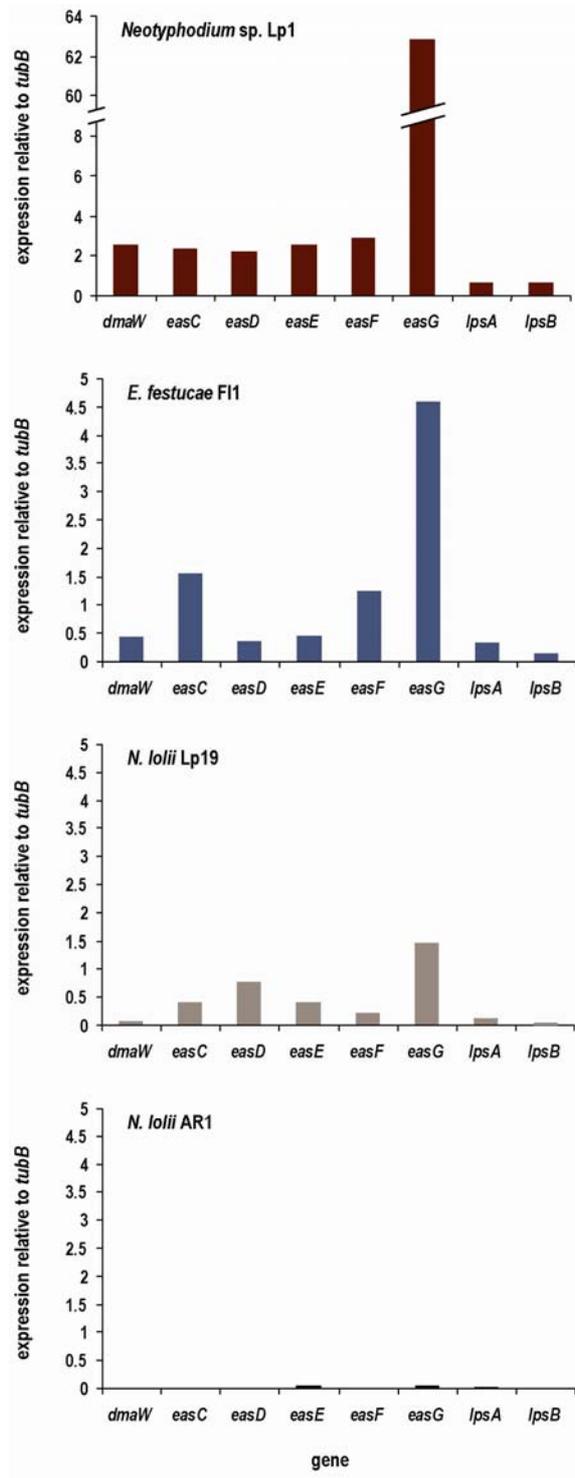
Figure 6.2. Expression of *eas* genes in different epichloë strains. Expression was determined relative to *tubB* as described in section . The expression of *easE* and *easG* in *N. lolii* AR1 is overrepresented as primer dimers were observed in the diagnostic melt curves (Appendix F). (Note different scales between *Neotyphodium* sp. Lp1 and other strains.)

6.3 Gene Order in Different Epichloë *EAS* Clusters

The four strains *E. festucae* Fl1, *N. lolii* Lp19, *N. lolii* AR1 and *Neotyphodium* sp. Lp1 each contain the *EAS* cluster of genes, however, gene expression and ergot alkaloid production is markedly different between the four strains. To determine whether there were any differences in gene order between the strains, inter-gene diagnostic PCR was performed on genomic DNA from each strain using the same primers used for the taxonomic distribution study, resulting in overlapping PCR products covering *EAS* clusters 1 and 2 (Fig. 6.3).

Figure 6.3. Gene order in different epichloë *EAS* clusters. Primers used for PCR are shown as numbered arrows. Asterisks mark areas in which polymorphisms were observed in the size of the PCR product amplified with primers spanning these regions. 1, DamP81; 2, DamP83; 3, DamP84; 4, DamP85; 5, DamP86; 6, DamP87; 7, DamP88; 8, DamP89; 9, DamP90; 10, DamP92; 11, dmaWqF; 12, dmaWqR; 13, DamP106; 14, DamP107; 15, DamP111; 16, DamP112; 17, DamP109

This analysis showed that with the exception of *dmaW-cloA*, gene order was conserved between each cluster. The *dmaW* and *cloA* genes were earlier shown to be linked in *E. festucae* Fl1 and one locus of *N. coenophialum* but not *N. lolii* Lp19 or *Neotyphodium* sp. Lp1 (section 3.6). This was further examined in the range of epichloë strains examined for taxonomic distribution and showed that *cloA* was found directly downstream of *dmaW* only in *E. festucae* and *N. coenophialum*.

6.3.1 Polymorphisms in the *easA-easG* and *cloA-easC* Regions

While gene order was conserved between each of the strains, two intergenic regions, *easA-easG* and *cloA-easC* appeared to be polymorphic based on a difference in the size of PCR products amplified (Fig. 6.3). Sequence analysis of the PCR products was undertaken to determine the nature of the polymorphisms.

6.3.1.1 Recombination in the *easA-easG* Intergenic Region

At the *easA-easG* intergenic region, an expansion of the SSR between the two Toru-1 elements has occurred in *Neotyphodium* sp. Lp1, resulting in 15 copies of the repeat rather than the 8 found in *N. lolii* Lp19 (Fig. 6.4). In *N. lolii* AR1 and *E. festucae* Fl1 there appears to have been recombination between Toru elements leading to polymorphism at this locus compared with *N. lolii* Lp19. In *N. lolii* AR1 there is just one Toru-1 element, rather than the two found at this position in other strains, and the sequence between the two elements found in the other strains, including the SSR, has been deleted (Fig. 6.4). This appears to have arisen from recombination between the two Toru-1 elements as the left 81 bp of the *N. lolii* AR1 Toru-1 element are identical with the *N. lolii* Lp19 left Toru-1 element while the right 57 bp are identical with the right element.

The large increase in size of the *E. festucae* Fl1 locus was due to duplication of the first exon of *easG* and the upstream region as far as the first Toru-1 element (Fig. 6.4). The presence of 18 bp of a Toru TIR directly adjacent to the duplication suggests that this was also due to recombination between TIRs of Toru elements, most likely the extant Toru-1 element upstream of *easG* and a third Toru element which integrated into the first *easG* intron.
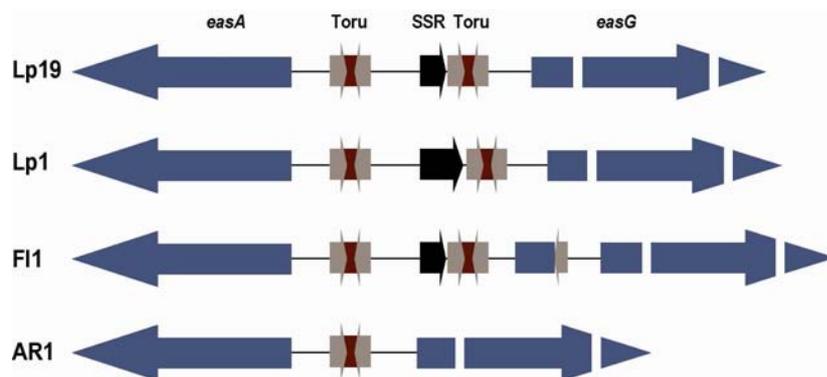
Figure 6.4. Schematic showing repeat-mediated polymorphism at the *easA-easG* locus. Genes are shown as blue arrows, introns are shown as gaps.
Lp19, *N. lolii* Lp19; Lp1, *Neotyphodium* sp. Lp1; Fl1, *E. festucae* Fl1; AR1, *N.lolii* AR1
(Diagram not shown to scale.)

6.3.2 The *cloA-easC* Intergenic Region

Sequence analysis of triplicate *N. lolii* AR1 *cloA-easC* PCR products revealed a 245-bp deletion 188 bp upstream of the *cloA* translational start site. Examination of the sequence within and surrounding the deletion did not reveal any repeat sequences indicative of recombination events. The SSR found by MEME analysis in the *cloA* upstream region of *E. festucae* E2368 (Section 4.2) was found to be reduced to 6 copies in the *N. lolii* AR1 promoter region.

6.3.3 *N. lolii* AR1 *dmaW* Sequence Analysis

The deletions of regions of the *N. lolii* AR1 *cloA*, *easA* and *easG* promoter regions could potentially explain the lack of expression of these genes. However no polymorphisms indicative of deletions or insertions in other promoter regions were observed. To determine whether the *dmaW* gene, the gene for the first committed step in ergovaline synthesis, and promoter region contained mutations, a PCR product was amplified from the 3' end of *dmaW* to the first upstream Toru-1 TIR with primers DamP117 and DamP118. Three cloned products were sequenced. This analysis showed 100% conservation of the *dmaW* coding sequence and introns compared with *Neotyphodium* sp. Lp1 and just three transition SNPs in the promoter region at positions -206, -320 and -1048 relative to the translational start

site. None of these SNPs are found in potential DNA-binding motifs identified in section 5.2 and although mutation of a *cis* regulatory element cannot be ruled out, the *N. lolii* AR1 *dmaW* gene seems likely to be functional. Possible reasons for the lack of expression of the AR1 genes are discussed in Chapter 7.

6.4 *EAS* Homologues in Aspergillus

With the genome sequence of eight species of *Aspergillus* available, these were examined for homologues of *A. fumigatus eas* genes. Each of the species contained genes with matches to each of the *eas* genes, in most cases several genes with E values of e-30 or less. Thus only the most highly conserved genes were examined further for linkage with other *eas* gene homologues. No other *Aspergillus* species examined contained an *EAS* cluster, including the very closely related *Neosartorya fischerii*.

Within the *A. fumigatus* genome itself, a cluster of genes with multiple *eas* homologues encoding typical secondary metabolism and hypothetical enzymes is found just 5 kb from a telomere of chromosome 8. This cluster contains genes that are predicted to encode enzymes for fumitremorgen synthesis, including the functionally characterised two-module NRPS for brevianamide synthesis, the first committed step (Maiya *et al.*, 2006) and prenyl transferase *ftmPt1* (Grundmann and Li, 2005). *N. fischerii* also appears to contain a fumitremorgen cluster with conserved gene order. Like the *EAS* cluster, the fumitremorgen cluster contains two prenyl transferase genes, both of which are highly similar to *dmaW* (E = 0.0) (Fig. 6.5). The cluster also contains highly similar sequences to the P450 monooxygenase gene *easM* (E = 0.0) and to the *E. festucae easH* dioxygenase (E = 5e - 34). The *E. festucae* gene was used to query the database as the *A. fumigatus easH* is a degenerate pseudogene. Interestingly, both the *EAS* and fumitremorgen clusters are bordered by an identical LINE retrotransposon, between the cluster and the telomere in the case of the *EAS* cluster and between the cluster and a PKS-containing gene cluster in the case of the fumitremorgen cluster. These data possibly support a duplication event for the progenitor of the *EAS* and *FTM* clusters.
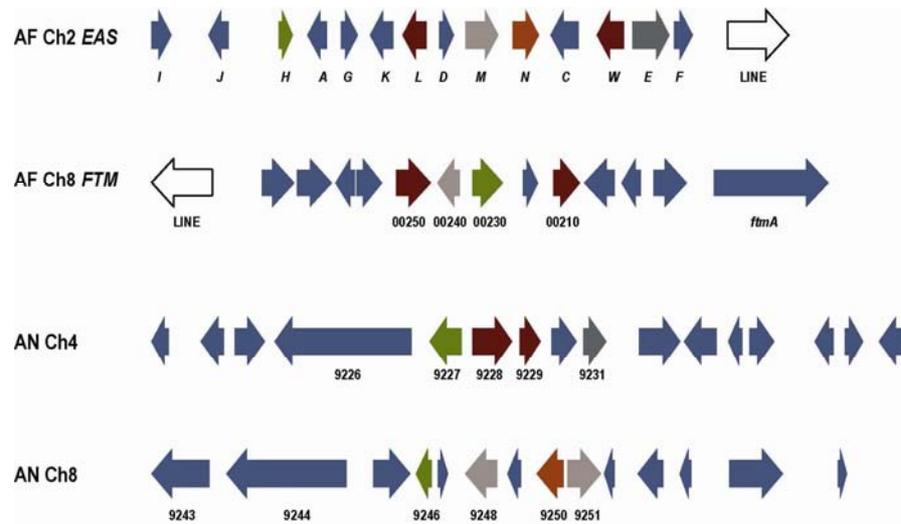
Figure 6.5. *EAS* gene homologues in aspergillus gene clusters. Blue arrows are unique to each cluster, other coloured arrows denote *eas* genes with homologues in at least one other cluster.

AF, *A. fumigatus*; AN, *A. nidulans*; Ch, chromosome

In *A. terreus* a more distantly related species, homologues for *easM* and *easH* are located next to each other in the same orientation as found in the *A. fumigatus* and *N. fischerii* fumitremorgen clusters. No other *EAS* gene homologues were found clustered with these genes however. In the aspergilli that are closely related to *A. terreus – A. flavus, A. oryzae* and *A. niger –* no linked *eas* gene homologues were able to be identified.

In *A. nidulans*, the species most distantly related to *A. fumigatus* and the other sequenced aspergilli (Galagan *et al.*, 2005), there are two putative gene clusters containing multiple *EAS* homologues (Fig. 6.5). The first cluster, located on contig 170, chromosome four, is located at the end of the chromosome. This cluster has not been characterised and thus the borders are not defined. However a zinc binuclear cluster transcription factor, a class of gene frequently found in association with biosynthetic gene clusters, is located just 928 bp from the telomeric repeats. Genes similar to *easE* (E = 0.0), *dmaW* (×2 E = 1.9e-27/2.9e-12) and to *E. festucae easH* (E = 5.6e – 33) are found present in this putative cluster. Other

genes encode putative *O*-methyltransferases, oxido-reductases and a two-module NRPS gene. Although also two modules, this putative NRPS has low similarity by BLAST analysis to that found in the fumitremorgen cluster and contains a methyltransferase domain.

The second *A. nidulans* gene cluster, located on contig172, on chromosome 8, contains homologues of *easM* (×2, both E = 0.0), *easN* (E = 0.0) and *E. festucae easH*. Unlike the other clusters this one does not contain *dmaW*-like genes. As for each of the other *EAS*-like clusters, an NRPS gene is found as part of the putative cluster although the predicted NRPS has an unusual domain structure of A-C-C-A. Adjacent to this gene is another gene with high similarity to various NRPS genes, however the predicted protein contains just a single A-domain. As for the other *A. nidulans* cluster the border was unable to be defined without experimental data but one side of the cluster is located adjacent to Dane2 degenerate retrotransposon sequences at the edge of the contig.

Taken together, the finding of two *EAS*-like clusters in both *A. fumigatus A. nidulans*, suggests an ancient duplication of a progenitor *EAS* cluster. Further bioinformatics analysis, outside the scope of this study, is required to determine at what point in ascomycete, Eurotiales, or *Aspergillus* evolution this occurred.

7. Discussion

7.1 *lpsB* - an NRPS Gene Required for Ergovaline Synthesis

An important element of the interaction between temperate grasses and epichloë endophytes is the fungal production of plant-protective natural products. Four classes of fungal alkaloids, peramine, lolines, indole diterpenes and ergot alkaloids in particular have received attention due to their well characterised anti-herbivore effects. Key genes or gene clusters for the synthesis of each of these compounds have been isolated (Panaccione *et al.*, 2001; Spiering *et al.*, 2002; Wang *et al.*, 2004; Spiering *et al.*, 2005; Tanaka *et al.*, 2005; Young *et al.*, 2005; Young *et al.*, 2006) and recent studies which have isolated NRPS and PKS genes have indicated that epichloë endophytes are likely to produce a great deal more potentially bioactive compounds (Tanaka *et al.*, 2005; Johnson *et al.*, 2007b).

The research described in this thesis arose from the isolation of an *in planta*-expressed NRPS gene fragment, ps12, with unknown function. Cloning of the gene associated with the ps12 fragment however revealed a single-module NRPS gene with high similarity to the *lpsB* gene from *C. purpurea*. The ps12-associated gene was thus likely to encode the LpsB NRPS required for synthesis of ergovaline rather than a novel secondary metabolite.

The predicted *N. lolii* LpsB enzyme is a single module with similar domain architecture to the orthologous *C. purpurea* enzyme, including an adenylation domain, thiolation domain, and condensation domain. The amino-terminal domain of LpsB also has weak similarity to a condensation domain, which may be evidence that LpsB has evolved from a longer NRPS. NRPS domains are identified by the presence of highly conserved motifs found in both fungal and bacterial enzymes (Marahiel *et al.*, 1997). The LpsB A- and T-domain motifs were readily identified, however the C-domain motifs were less clear, with the exception of the C3 motif, which contains a histidine residue that is required for catalysis (Bergendahl *et al.*, 2002) and is thus likely to be under greater selection pressure.

LpsB is unusual among fungal NRPSs in that it forms a multi-enzyme complex with LpsA (Riederer *et al.*, 1996). Multienzyme NRPSs are common in prokaryotes but the LpsA/LpsB system is the only described example in the fungal kingdom, where NRPSs are otherwise found on a single large multimodular polypeptide (Walton *et al.*, 2004). There is a condensation domain found at the carboxy end of the LpsB protein and also a partial condensation domain at the amino-proximal end of the predicted *N. lolii*, but not *C. purpurea*, LpsA enzyme (Damrongkool *et al.*, 2005). Whether the partial LpsA condensation domain plays a role in the condensation of the lysergyl-alanine peptide bond in the *N. lolii* enzyme system remains to be determined.

A feature of multienzyme NRPSs in prokaryotic systems is the presence of recently described COM domains, which are required for specific protein-protein recognition (Hahn and Stachelhaus, 2004). Alignment of the carboxy end of the predicted *N. lolii* and *C. purpurea* LpsB sequences with bacterial donor COM domains showed that the TPSD motif present at the junction between epimerisation and COM domains is absent, a result not totally unexpected given the C-terminus of LpsB is preceded by a condensation domain. The amino-terminal LpsA sequence from *N. lolii* and the LpsA-1 and LpsA-2 sequences from *C. purpurea* are also dissimilar to the bacterial acceptor COM domain consensus sequence and, interestingly, to each other. Different acceptor COM domains for LpsA-1 and LpsA-2, one perhaps with higher affinity for the LpsB donor COM domain, may partly explain why less of one ergopeptine product is synthesised in *C. purpurea* ergots. With no other fungal multienzyme NRPS systems yet described, functional analysis is required to identify the residues important for LpsB-LpsA protein-protein interaction.

Contrary to the interpretation of Correia *et al.* (2003), which concluded that the *C. purpurea* LpsB binding pocked was dissimilar to those of tryptophan activating modules, analysis here of the likely substrate-binding amino acids from the putative LpsB binding site revealed strong similarity to those from tryptophan-activating adenylation modules. This may indicate A-domain binding of the indole

ring moiety of lysergic acid and possibly indicates that the LpsB enzyme evolved from a tryptophan-activating module. Why the LpsB/LpsA NRPS system evolved at all when, as mentioned above, other characterised fungal NRPSs are found on single polypeptides is intriguing. It has been hypothesised that having the lysergic acid-activating module on a separate enzyme allowed formation of multiple lysergyl-containing peptides using the same module (Correia *et al.*, 2003). This seems to be the case in *C. purpurea* where there are two LpsA orthologues which each interact with LpsB to form different ergopeptines (Haarmann *et al.*, 2005; Haarmann *et al.* 2007). Lysergyl alanine was also proposed to be the product of a non-LpsA enzyme interacting with LpsB. This theory appeared to be incorrect for epichloë as both lysergyl alanine and ergine were not detected in an LpsA mutant of *Neotyphodium* sp. Lp1 (Panaccione *et al.*, 2003). However analysis of the *lpsA* mutant in this study by tandem mass spectrometry did detect very small amounts of lysergyl alanine and ergine. It thus seems likely that a third NRPS is indeed required to interact with LpsB to form lysergyl alanine, and possibly ergine. Why there would be such low levels in the *lpsA* mutant is not clear. It may be that the bulk of lysergyl alanine is indeed produced by LpsA or it may be that expression of the as yet unidentified NRPS is inhibited by some form of feedback. This idea is further discussed below in section 7.4.

To confirm the role of *lpsB* in the formation of ergovaline, targeted disruption was performed in *E. festucae*. Plants infected with this mutant were unable to synthesise ergovaline while complementation with a fragment containing the *lpsB* gene sequence restored this ability, confirming that *lpsB* is required for ergot alkaloid biosynthesis. Complementation did not restore ergovaline production to wild-type levels. This was possibly due to the length of sequence (379 bp) upstream of the *lpsB* start site, position effects of the ectopically transformed constructs, or both. Nevertheless, complemented strains were clearly able to synthesise ergovaline.

Synthesis of lysergyl alanine is also blocked by mutation of *lpsB*, and was complemented by pDF1. Although production of ergine by wild type *E. festucae* symbiota was not always detected, its absence from all plants infected with the

lpsB mutant strain and occurrence in some plants with the complemented strains suggest that *lpsB* is required for its synthesis. As with the *lpsA* mutant of *N. sp.* Lp1, accumulation of the major identified clavine alkaloids did not appear to be affected by mutation of *lpsB.* An increased concentration of an unidentified minor clavine was noted. Interestingly this appears to be an isomer of the clavine assigned the structure 6,7-secolysergine that increased in concentration in the *Neotyphodium.* sp. Lp1 *lpsA* mutant compared with wild-type (Panaccione *et al*., 2003). A compound previously identified in *E. typhina*, 6, 7-secoagroclavine (Porter *et al.*, 1981), has mass spectrometric properties consistent with this isomer although the structure of the unknown compound could not be determined from this analysis and a standard was not available for comparison.

In order to gain insight into the ecological role of ergot alkaloids, the *lpsB* mutant was used in a bioassay performed with black beetle, an important pasture pest in the warm northern regions of New Zealand. Results showed no significant difference in feeding deterrence from wild type or *lpsB* mutant *E. festucae*-infected ryegrass. This suggests that the predominant feeding deterrent in *E. festucae* symbiota may not be ergovaline or lysergic acid amides, which are absent in the *lpsB* mutant. This was surprising as ergopeptines, including ergovaline, are known feeding deterrents (Ball *et al*., 1997) and circumstantial evidence previously implicated ergovaline-producing endophytes in pasture protection against black beetle (Ball *et al*., 1994). Intermediate ergot alkaloids may be responsible for the feeding deterrence. Ball *et al.* (1997) showed that clavines festuclavine and lysergol did not have any effect on black beetle, however, other clavines were not tested. Black beetle deterrence has been observed from non-ergot alkaloid producing epichloë endophytes, although not as strongly as from those producing ergot alkaloids (Popay and Baltus, 2001; Popay *et al*., 2005). What compound is responsible in these strains is not known, although terpendoles, which are synthesised by *E. festucae* Fl1, may be important (A. Popay and B. Tapper, pers. com.). Black beetle bioassays comparing the *Neotyphodium* sp. Lp1 *dmaW* and *lpsA* mutants, which are blocked at early and late stages in the ergot alkaloid pathway, will be useful in determining whether intermediate ergot alkaloids or other

endophyte compounds are likely to be responsible for the deterrence effect seen from the *lpsB* mutant.

7.2 Ergot Alkaloid Genes are Clustered with *lpsB*

Genes for ergot alkaloid synthesis, *dmaW* and *lpsA*, had previously been identified (Panaccione *et al.*, 2001; Wang *et al.*, 2004; Damrongkool *et al.*, 2005), however, flanking sequences were not isolated. This was presumably due to the difficulty in cloning these regions of the genome, as was also observed with the *lpsB* gene, being absent from any large insert libraries screened. In order to determine whether ergot alkaloid genes were found as a gene cluster in epichloë, as is the case for *C. purpurea* and *A. fumigatus*, sequence surrounding *lpsB* was cloned and analysed for the presence of genes homologous to those in the *C. purpurea* ergotamine gene cluster.

The *EAS* cluster was very difficult to clone. Aside from the two clones that contained *lpsB*, just five further clones were retrieved from repeated screening of two λ libraries and one of these was chimeric. Attempts to clone sequence flanking *dmaW* and *lpsA* were similarly difficult with no clone isolated for *dmaW* and just one at the 5' end of *lpsA*. Attempts at long PCR between *dmaW, lpsA* and each end of the *lpsB* sequence were unsuccessful. Southern analysis did however show that each of the genes was linked on substantially different sized *Not*I fragments in *N. lolii* and *E. festucae*, indicative of the presence of additional transposon blocks in *N. lolii*. Subsequent analysis of the preliminary *E. festucae* E2368 genome revealed a second cluster of genes associated with *dmaW*. Combined *N. lolii* and *E. festucae* data thus revealed the *EAS* cluster to be found on three (*E. festucae*) or four (*N. lolii*) linked clusters with AT-rich, transposon-derived sequence separating the loci. This is reminiscent of the epichloë *LTM* gene cluster for indole diterpene synthesis, which was also difficult to clone and associated with large tracts of AT-rich, retrotransposon derived sequence which separates three clusters of *ltm* genes (Young *et al.*, 2005; Young *et al.*, 2006). These two gene clusters are among the most complex yet isolated from fungi. Only the *TOX2* locus for HC-toxin synthesis in *Cochliobolus carbonum* shows a similar level of complexity, genes for which are duplicated, sometimes twice, and spread over ~600 kb (Ahn *et al.*, 2002). Some

genes for other fungal biosynthetic pathways are not found in the core cluster, *tri1*, *tri16* and *tri101* for tricothecene synthesis in fusaria for example (Kimura *et al.*, 1998b; Meek *et al.*, 2003), and some contain transposons or transposon relics, such as the *A. fumigatus EAS* and *FTM* clusters, but none are reported as being interspersed with large tracts of transposon relics which seem to be a unique feature of these epichloë gene clusters. At least some of the *TOX* genes are flanked by transposons (Panaccione *et al.*, 1996) although whether these are a part of larger clusters is not known.

Combined *N. lolii* and *E. festucae* genome sequence revealed the presence of eight genes associated with either *lpsB* or *dmaW*. Each of the genes within the *EAS* cluster contained homologues of ergot alkaloid biosynthetic gene clusters found in *C. purpurea* (Tudzynski *et al.*, 1999; Correia *et al.*, 2003; Haarmann *et al.*, 2005) and *A. fumigatus* (Coyle and Panaccione, 2005; Unsold and Li, 2005). With the exception of *easH*, which is likely to be a pseudogene in *A. fumigatus* (Schardl *et al.*, 2006), the genes common to all three clusters are proposed to encode enzymes for the early steps in the ergot alkaloid biosynthetic pathway. Genes that are shared between *C. purpurea* and *N. lolii* but absent from *A. fumigatus* - *lpsA, lpsB* and *cloA* - have been shown to be required for steps leading to ergopeptines (Panaccione *et al.*, 2001; Correia *et al.*, 2003; Haarmann *et al.*, 2006).

Correia *et al.* (2003), Haarmann *et al.* (2005) and Schardl *et al.* (2006) have proposed functions for each of the genes in the *C. purpurea* and *A. fumigatus EAS* gene clusters. The sequence similarity of the epichloë *eas* genes with those of the other clusters suggests identical roles for the epichloë othologues. There are seven enzymatic steps in the pathway predicted to be shared between epichloë/claviceps and *A. fumigatus* and seven shared potentially functional genes. The first committed step in ergot alkaloid biosynthesis is the prenylation of tryptophan by DmaW to form dimethylallyl tryptophan (Fig. 1.4). The subsequent step to chanoclavine requires *N*-methylation and two oxidation reactions. The EasF enzyme is likely to be responsible for the *N*-methylation as it contains a weak match to a SAM-dependant methyltransferase domain. The *easC*-encoded catalase

is not immediately obvious as one of the seven biosynthetic enzymes, however, catalases have been shown to catalyse oxidative reactions (Zamocky and Koller, 1999) and Schardl *et al.* (2006) speculate that it may work in conjunction with the potentially peroxide-producing FAD-dependant oxidoreductase EasE in catalysing consecutive oxidations, most likely the overlapping steps from *N*-methyl DMAT to chanoclavine or from the diene of *N*-methyl DMAT to chanoclavine aldehyde. The correlation between extremely high *easG* expression and very high levels of chanoclavine observed in this study, particularly in *Neotyphodium* sp. Lp1, may suggest a role for this enzyme in chanoclavine synthesis, so the latter role for EasE/EasC seems most likely. Thus a working model for the pathway to chanoclavine aldehyde would be *N*-methylation of DMAT by EasF, diene formation by EasG, C5-C10 ring closure by EasE and subsequent aldehyde formation at C7 by EasC and H_2O_2-mediated oxidation.

The conversion of chanoclavine aldehyde to agroclavine requires epimerisation to the iminium ion, followed by a separate reduction reaction. Schardl *et al.* (2006) postulate that the EasA flavoenzyme is likely to catalyse the epimerisation reaction as the tightly bound FMN co-factor would allow release of a hydrogen atom to reduce the C8-C9 double bond and then to use the reduced substrate to reacquire a hydrogen atom and oxidise the same bond, which is required for closure of the D-ring. The chanoclavine cyclase activity identified in a *Claviceps* mycelial fraction required NAD(P)H (Schardl *et al.*, 2006), hence if other predictions are correct, EasD, a putative NAD(P)H dehydrogenase, fits with the model as the enzyme to catalyse the reduction of the iminium ion to agroclavine.

Subsequent steps in ergovaline synthesis are shared with *C. purpurea* but not *A. fumigatus*. The synthesis of elymoclavine from agroclavine and lysergic acid from elymoclavine appear to be P450 monooxygenase-mediated reactions. Just one P450 gene, *cloA*, is present in the *EAS* cluster however, and it has been shown in *C. purpurea* to catalyse the formation of lysergic acid from elymoclavine (Haarmann *et al.*, 2006). P450 monoxygenases frequently catalyse more than one step in a biosynthetic pathway (eg. gibberellin and indole diterpene pathways

(Tudzynski *et al.*, 2001a; Tudzynski *et al.*, 2002; Saikia *et al.*, 2007)), however, while CloA was not ruled out as being capable of using agroclavine as a substrate, the presence of elymoclavine in a *cloA* mutant suggests there is another enzyme capable of catalysing this step. This reaction may thus be catalysed by the EasH hydroxylase. This leaves one step in ergovaline synthesis unaccounted for, the formation of ergovaline from the LpsB/LpsA product lysergyl peptide lactam. This enzyme may be encoded outside of the cluster, as is seen in trichothecene synthesis where *tri1*, *tri16* and *tri101* genes are dispersed in the genome outside the core *TRI* cluster (Kimura *et al.*, 1998b; Meek *et al.*, 2003), or alternatively one of the *EAS* oxidases, perhaps CloA, is also able to catalyse this reaction. Feeding studies with lysergyl peptide lactam in a host background containing single candidate *eas* genes would help to answer this question, although the lysergyl peptide lactam may not be easy to obtain in sufficient quantity.

The predicted roles for *EAS* enzymes are a working model only and functional analysis is required to confirm roles for each of the enzymes in the pathway. The roles of *easC* and *easH* products in particular will be of considerable interest. For the early steps *A. fumigatus* is the logical organism in which to perform these analyses as it is considerably more genetically tractable than either *E. festucae* or *C. purpurea*.

7.3 Transposons Associated with the *EAS* Cluster

There are at least ten different transposons associated with the epichloë *EAS* gene cluster. Five LTR retrotransposon relics, Tahi, Rua, Ono-nui, Whitu and Tekau; two type two element relics, Wha and Waru; and three MITEs, Toru, Rima and Iwa. This striking number of different elements associated with a single gene cluster suggests a very large number are present in epichloë genomes. Environmental isolates of predominantly asexual fungal species, *Fusarium oxysporum* and *Magnaporthe grisea* in particular, have been shown to be rich in transposable elements (Daboussi and Capy, 2003). In the sequenced genome of *M. grisea*, 17 different transposable elements were reported (Dean *et al.*, 2005). During the course of characterising the ten *EAS*-associated repeats, a further seven repetitive sequences were identified but not further characterised. A very preliminary survey of the *E. festucae* genome has thus revealed the same numbers of transposons as are found in *M. grisea*. At least double this number does not seem unreasonable. Could the potentially very large number of transposons be derived from the plant that the fungus is so closely associated with? The similarity of the Waru element, the first described fungal *Pif/harbinger* type transposon, with elements from plants may support this hypothesis, although a considerably larger analysis of both epichloë and grass genomes is required before this could be confidently asserted.

The number of MITEs in the *E. festucae* genome – at least three – is surprising in that just two MITEs have been reported in fungi previously, one in *N. crassa* (*guest*) (Yeadon and Catcheside, 1995; Ramussen *et al.*, 2004) and one in *F. oxysporum* (*mimp*) (Hua-Van *et al.*, 2000; Dufresne *et al.*, 2007). The low number suggested that MITEs were not very successful in these organisms (Dufresne *et al.*, 2007) compared to other eukaryotic genomes where they are common, particularly in plants (Wessler *et al.*, 1995; Feschotte *et al.*, 2002). The degeneracy of the elements, Toru and Rima in particular, meant that these elements would not be likely to be found by repeat masking algorithms during genome sequence analysis. There is thus a possibility that MITEs are more prevalent in fungi than has been reported.

Analysis of a limited number of *M. grisea* contigs for high-copy inverted repeats did not reveal any MITE-like elements, however further bioinformatics analysis is required to determine whether MITEs are peculiarly prevalent in epichloë.

The copy number of Rima in the genome, approximately 10, was similar to that observed for *mimp*, the low copy number of which was put down to the small genome size of fungi and therefore the expectation that transposon numbers would be low due to the correlation between genome size and transposon content (Dufresne *et al*., 2007). The Toru and Iwa MITEs have considerably higher copy number though and are therefore more similar to MITE families of higher eukaryotes.

MITEs are non-autonomous transposable elements that require a parent autonomous element in order to be mobilised. Just a few parent elements have been definitively identified for MITEs although several families can be linked to a parent transposase based on similarity of TIRs (Feschotte *et al*., 2002). Examination of the genome sequence for a potential parent element of Toru, Rima or Iwa was not successful, however this analysis is difficult with the current restricted access. Future examination of the genome sequence, combined with PCR amplification using the TIR sequences, will determine whether parent elements exist for any of these elements or whether they, like the autonomous elements identified in this study, are immobile relics.

In plant genomes, MITEs appear to integrate preferentially into genic regions of genomes. *Tourist* and *Stowaway* were identified near more than 100 plant genes and a large number of MITEs are found in rice genes (Bureau and Wessler, 1992, 1994a, b; Bureau *et al*., 1996). Not all MITE families have a site preference for genic areas however and analysis of integration sites for the *F. oxysporum mimp* element showed it to be largely found adjacent to or nested within other repeat elements (Dufresne *et al*., 2007). To investigate whether this was true of the epichloë MITEs, integration sites were determined for a selection of copies of each element in the *E. festucae* E2368 genome. This analysis revealed that the Toru element in particular was frequently found within 1 kb of a gene, and particularly the 5' region. This has

potential implications for gene regulation. Transposons can alter gene expression by integration into *cis*-acting elements, by providing new regulatory elements, and through epigenetic effects on nearby genes (Cambareri *et al.*, 1996; Kidwell and Lisch, 1997; Lippman *et al.*, 2004; Cultrone *et al.*, 2007; Slotkin and Martienssen, 2007). A MITE element in rice was recently shown to have both a positive *cis*-regulatory effect and a negative epigenetic effect through DNA methylation (Yang *et al.*, 2005a).

The non-MITE elements identified were generally found at the edges of contigs. The finding that the elements at each edge of the *EAS* clusters were different, or contained different TSDs, suggests that these elements are clustered with other elements as has been observed at the *LTM* locus and in some other fungi (Thon *et al.*, 2006; Young *et al.*, 2006). Transposons may be clustered as a passive evolutionary consequence, that is, insertions into genic regions are selected against due to deleterious effects while insertions can build up in areas where there is little selection pressure. Alternatively there may be an evolutionary benefit to the clustering of transposons. Transposons have been called "drivers of genome evolution" (Kazazian, 2004). Transposon clusters have been proposed to be important for the maintenance of heterochromatin and in centromere function. Probably the clearest example of an evolutionary benefit from a build-up of transposons is the finding that drosophila telomeres actually consist of tandem arrays of *HET-A* and *TART* transposons (Pardue *et al.*, 1996). Large numbers of repeat sequences are also able to mediate recombination leading to deletions, inversions and duplications, which can have both positive and negative effects. Possible evidence of past recombination among mobile elements in epichloë is the presence of solo Ono LTRs. One of these LTRs is flanked by a TSD, indicating recombination between the LTRs of a single element, thought to be a common mechanism for disabling retroelements and reducing genome size (Vitte and Panaud, 2003). Two of the elements were not flanked by duplications and are thus likely to have arisen from recombination between different elements. For each of the retroelements examined, a much larger number of BLAST matches to LTRs were observed in the *E. festucae* E2368 genome than to internal regions of the

elements, suggesting recombination has been common within these elements. An example of recombination mediated by transposons was observed in the intergenic region between *easA* and *easG* where apparent recombination between Toru elements lead to a deletion of intergenic sequence in *N. lolii* AR1 and a duplication of the first exon of *easG* and 452 bp of upstream sequence in *E. festucae* Fl1 relative to the *N. lolii* Lp19 sequence. This clearly demonstrates the evolutionary capability of transposons in epichloë.

Aside from the emerging positive effects of transposable elements on host genomes, clearly mobile elements can have negative effects, such as disruption of coding sequences, as observed with the Wha element disruption of a putative membrane protein in *E. festucae* E2368, and altered expression from either disruption of *cis* regulatory elements or epigenetic silencing. To limit these effects, organisms have evolved genome defence mechanisms that mutate or silence non-host sequences. In bacteria, host restriction is a well-known genome defence mechanism. Gene silencing of repeat sequences by quelling/RNAi and methylation of repeat sequences are mechanisms shared by many eukaryotes (Almeida and Allshire, 2005; Freitag and Selker, 2005). A mechanism that appears to be limited to fungi is repeat induced point mutation (RIP). RIP was first identified in *N. crassa* and is characterised by C:G to T:A mutations in all copies of duplicated sequences, whether from transposons, gene duplications or transformation (Cambareri *et al.*, 1989). Originally RIP was thought to occur only in *N. crassa* however more recently evidence of RIP has been observed in aspergilli (Clutterbuck, 2004; Montiel *et al.*, 2006), *M. grisea* (Nakayashiki *et al.*, 1999), *F. oxysporum* (Hua-Van *et al.*, 1998), *Podospora anserina* (Graia *et al.*, 2001) and *L. maculans* (Idnurm and Howlett, 2003; Attard *et al.*, 2005). The AT-richness of the Tahi and Rua elements observed at the *LTM* locus of *N. lolii* and *E. festucae* suggested that RIP might also occur in epichloë (Young *et al.*, 2005). This was examined in more detail in this study. Using the method of analysis of Margolin *et al.* (1998) each *EAS*-associated element, with the exception of Rima, appears to have undergone RIP. RIP in *N. crassa* only occurs on repeat sequences larger than ~400 bp (Cambareri *et al.*, 1989) so ratios indicative of RIP were surprising for the Toru (~140 bp) and Iwa (~238 bp) elements. This may

mean that the RIP mechanism in epichloë is capable of recognising smaller repeats than that of *N. crassa*. There is some evidence that RIP mechanisms may be different in different fungal genera. The amount of RIP can be quite different in different organisms, *N. crassa* for example typically mutating up to 70% of cytosine residues in a repeated sequence (Freitag *et al.*, 2002) while mutated residues can be as low as less than 1% for certain repeats in *M. grisea* (Ikeda *et al.*, 2002). The sequence context in which a cytosine residue is preferentially mutated is also different between fungi, as shown in Table 5.7. Thus, although each of the fungi examined, including *E. festucae*, contain a copy of the *rid* gene, encoding a putative cytosine methyltransferase required for RIP in *N. crassa* (Freitag *et al.*, 2002), it seems other factors involved in the RIP process may be different in different fungi. Future experiments transforming *E. festucae* with various sized DNA fragments and passing through a sexual cycle will determine whether this species can indeed recognise small repeated sequences and whether *E. festucae* is still capable of performing RIP at all. Certainly the RIP observed in *N. lolii*, an asexual organism, is likely to have occurred long ago prior to the loss of sexuality. Although a possible RIP-like process in vegetative cells has been mooted (Daboussi and Capy, 2003), this seems unlikely in epichloë, as RIP would be lethal to the hybrid species that have frequently arisen during epichloë evolution.

So, if *E. festucae* is able to recognise and disable repeat sequences through RIP, how did the genome become host to such a large number of transposable elements? Genome versus genome BLAST analysis suggests that almost every contig end contains repeat sequences, suggesting the transposable element content could be largely unsequenced and a very large percentage of the genome. Strains of *E. festucae* rarely produce stromata on host plants and go through the sexual cycle. Thus there may be prolonged periods when transposons are able to invade the genome and replicate before the host genome goes through the sexual cycle and causes RIP to disable these sequences. This is supported by the finding that some repeat sequences such as Tekau, are very heavily mutated, while some, such as Waru, are less so, suggesting they have been through different numbers of sexual cycles. Also, analysis of the Tahi and Rua retrotransposons in different epichloe

species showed that these elements were absent from *E. typhina* genomes (Young, 2005), which frequently form stromata and go through a sexual cycle. Interesting future experimentation would be to examine the phylogenetic distribution of each of the elements identified in this study and also to identify potential active transposons, particularly in the asexual species, possibly by transposon trapping which has identified several active elements in *F. oxysporum* (Daboussi and Capy, 2003).

7.4 Perspectives on *EAS* Cluster Regulation

Ergot alkaloids are specifically produced during biotrophic growth of epichloë endophytes. Attendant with this, the *eas* genes are all expressed *in planta* but not under any axenic culture condition tested. Ergot alkaloid biosynthetic genes in *C. purpurea* cultures are repressed by high phosphate levels (Correia *et al.*, 2003; Haarmann *et al.*, 2005) and nitrogen and/or carbon deprivation induces a number of genes in plant-pathogenic fungi including the *G. fujikuroi* gibberellin genes (Mihlan *et al.*, 2003), the *M. grisea mpg1* hydrophobin gene (Talbot *et al.*, 1993) and the *Cladosporium fulvum avr9* (van den Ackerveken *et al.*, 1994). However, results here suggest that in *E. festucae* the *lpsB* gene at least is not derepressed in culture by phosphate, carbon or nitrogen catabolite starvation. Other than in *N. uncinatum*, in which some carbon sources support expression (Blankenship *et al.*, 2001), loline genes are not affected by carbon and nitrogen limitation and a comprehensive recent analysis showed that lolitrem genes are similarly unaffected by differences in carbon and nitrogen source or amount (K. May and M. Bryant, pers. com.). To further support the lack of a role for catabolite repression in *EAS* gene regulation, no enrichment for AreA/Nit2 or CreA (the DNA-binding proteins respectively mediating nitrogen and carbon catabolite repression) binding sites was observed in *eas* gene promoters. Although possible AreA/Nit2 binding sites were observed in some promoters, none were found in the closely spaced arrangement observed in high affinity AreA/Nit2 binding sites (Marzluf, 1997).

Oxidative stress and growth in darkness, conditions potentially faced *in planta* and known to induce expression of some fungal genes, also failed to induce expression of *lpsB*. These results suggest that specific plant conditions may be required for induction. Several genes that are specifically expressed in response to plant signals have been identified in plant-associated fungi (Wubben *et al.*, 1994; Basse *et al.*, 2002; Thara *et al.*, 2003; Bohnert *et al.*, 2004), however it was only recently that a specific plant factor responsible for expression of one of these genes was identified when Yang *et al.* (Yang *et al.*, 2005b) showed that the *pelD* gene of *Nectria haematococca* was expressed in response to homoserine and asparagine. The

inability of *E. festucae* Fl1 to induce *lpsB* expression in the presence of a plant extract, albeit after a relatively brief exposure, suggests that the inducing factor was not stable or alternatively that specific signalling between the epichloë endophyte and the grass host occurs. This is difficult to test, although current work with forward genetic screening (K. May, A. Tanaka and B. Scott pers. com.) and functional analysis of fungal secreted proteins (R. Johnson, S. Bassett and G. Bryan pers. com.) may shed some light on whether host-fungal signalling occurs.

The expression of each of the *eas* genes *in planta* but not in axenic culture suggests co-regulation of these genes. Indeed, this is a key feature of fungal biosynthetic gene clusters. The upstream regions of the *eas* genes can thus be expected to contain shared *cis*-acting elements. *Cis* regulatory sequences have been identified in the promoter regions of genes in biosynthetic gene clusters (Fernandes *et al.*, 1998; Hohn *et al.*, 1999; Pedley and Walton, 2001). Putative enriched sequences have also been identified in the upstream regions of *LTM* and *LOL* genes in epichloë (Young *et al.*, 2006; Kutil *et al.*, 2007). Analysis of the *EAS* promoter regions revealed five motifs that appeared to be present more than could be expected by chance. This number was reduced to four when control promoters were examined. Of the four genuinely enriched motifs, motif 2 is specifically enriched in *EAS* gene promoters with the exception of a binding site in the *hmgA* promoter. This does not rule this motif out as a regulatory element as it is not inconceivable for there to be some shared regulation between the *EAS* cluster and *hmgA*, the product of which, HMG CoA reductase, is involved in the synthesis of dimethylallyl diphosphate (DMAPP), a key metabolite in the ergot alkaloid pathway. The remaining three motifs (3 – 5) each had a cytosine-rich core sequence and best matched the same binding sites in the TRANSFAC transcription factor database, those for Adr1 and the stress responsive element (STRE)-binding Msn2/4p, possibly indicating these motifs represent a single degenerate DNA binding sequence. Binding sites for Adr1 and STRE were also identified in the *LOL* gene promoters (Kutil *et al.*, 2007), perhaps suggesting a level of shared regulation. Promoter deletion analyses are now required to confirm whether the identified motifs are indeed involved in *eas* gene regulation.

Chromatin remodelling as a method of co-ordinate gene regulation is a possible factor causing selection pressure for secondary metabolite genes to be clustered. The *EAS* cluster is located near to the telomere. These regions are frequently sites of heterochromatin formation in other organisms, a situation that is likely to be similar in filamentous fungi. This and the presence of transposons at the *EAS* locus, particularly AT-rich, RIP degenerated sequence, a potent inducer of DNA methylation in *N. crassa* (Tamaru and Selker, 2001), suggests that regulation at the epigenetic level may be important for *eas* genes. The identification in aspergilli of a putative histone methyltransferase, LaeA, that is a global secondary metabolite gene regulator (Bok and Keller, 2004), and the finding that transfer of a housekeeping gene to a location within the sterigmatocystin gene cluster partially silences expression (Bok *et al.*, 2006a), supports a gene silencing mechanism in fungal secondary metabolism. Strategies to clone *laeA* in *E. festucae* were not possible due to the lack of an obvious orthologue outside the aspergilli despite the claim that clear homologues were present in genomes other than aspergillus (Bok and Keller, 2004). The *A. nidulans laeA* gene was therefore expressed heterologously in *E. festucae* to determine whether this gene could induce epichloë alkaloid genes. Although the highest *laeA*-expressing strain did have the highest level of gene expression for the two *eas* genes tested, this was not true of the *ltm* gene and when other *laeA*-expressing and control strains were examined, there was no correlation between expression of *laeA* and secondary metabolism genes. While this was a small sample of the secondary metabolic genes present in *E. festucae*, this result, combined with genome synteny and phylogenetic data suggests that there is no *laeA* orthologue in *E. festucae*.

Is chromatin structure then important in *EAS* cluster regulation? For the reasons outlined above it seems likely. Possibly an LaeA-like protein functions in *E. festucae* but has diverged sufficiently from the aspergillus protein that it is no longer a functional orthologue. Or possibly, chromatin effects on gene clusters outside the aspergilli (or Eurotiales) are regulated differently. Some support for this comes from the recent report that the *dim2* DNA methyltransferase gene required for all DNA methylation in *N. crassa* (Kouzminova and Selker, 2001) is absent from

aspergillus genomes (Montiel *et al.*, 2006). Analysis of the *M. grisea*, *F. graminearum* and *E. festucae* genomes shows that this gene is present and suggests that at least one key aspect of epigenetic regulation, DNA methylation, is either absent or occurs via a different mechanism in the *Aspergillus* genus. In order to show whether epigenetics may be important for *EAS* cluster regulation, future studies should look for DNA methylation of the transposon sequences and promoters at the *EAS* locus by either methylation sensitive and insensitive restriction digests and Southern analysis or bisulphite sequencing, a method more likely to be informative both in culture grown mycelia and *in planta* as it is PCR based. Methylation of histone H3 at lysine 9 (H3K9), a heterochromatin marker, could also be tested by chromatin immunoprecipitation. A complementary approach would be to delete candidate genes such as the *dim2* DNA methyltransferase gene and *dim5* histone methyltransferase (Tamaru and Selker, 2001) and analyse *EAS* gene expression.

An interesting finding in this study was that the *N. lolii* AR1 genome contains each of the known *eas* genes, although ergovaline has not been detected in this strain. Further analysis showed that no ergot alkaloids of any kind were produced in *N. lolii* AR1-infected symbiota and that there was no, or extremely low, gene expression. Possible explanations for this include mutation of each of the promoters, or mutation of a *trans* regulatory factor for these genes. Analysis of the sequence of the *N. lolii* AR1 *dmaW* gene and promoter revealed an extremely high level of conservation compared with *Neotyphodium* sp. Lp1 with just three SNPs in the promoter region and 100% conservation of the coding sequence. Thus this gene at least appears to be potentially functional. Analysis of the intergenic regions of *N. lolii* AR1 revealed complete conservation of gene order with the other three strains analysed but deletions within the *cloA* and *easA-easG* promoter regions that would be likely to affect gene expression. This leads to another hypothesis; that these genes are subject to feedback regulation. Feedback was clearly shown to occur in the cercosporin biosynthetic gene cluster of *Cercospora nicotianae*, when deletions of two different genes in the cluster led to complete repression of expression of other genes in the pathway (Chen *et al.*, 2007). Feedback has been suggested previously

164

for ergot alkaloid metabolism when intermediates were observed to accumulate at lower than expected levels in the *Neotyphodium* sp. Lp1 *lpsA* mutant (Panaccione *et al.*, 2003). This could also explain the huge reduction of lysergyl alanine and ergine synthesis in the *lpsA* mutant. A mechanism for any feedback regulation at the transcriptional level is not obvious. Traditional feedback regulation is at the enzymatic level through allosteric inhibition from metabolites. Regulation via mRNA riboswitches (aptamers), common in bacteria and recently identified in fungi (Cheah *et al.*, 2007), which bind metabolites, causing a change in secondary structure and subsequent inhibition of translation, is also at the posttranscriptional level. Any transcriptional feedback mechanism is thus likely to be novel. Analysis of *eas* gene expression in the available mutant strains and transformation of *N. lolii* AR1 with intact (relative to other strains) *easA-easG* and *cloA* regions will shed light on whether this phenomenon is indeed occurring. An intriguing possibility is that the SSR found between *easA* and *easG* in ergot alkaloid producing species and deleted in *N. lolii* AR1 plays some role in regulation of one or both of these genes. Further support for this is found in the expansion of the SSR in *Neotyphodium* sp. Lp1, in which *easG* expression is extremely high compared to other genes.

Theories around epichloë endophyte natural product gene regulation remain speculative. Identifying the regulatory mechanisms that activate the expression of the *eas* genes *in planta* and determining whether there is some commonality with mechanisms for activating other plant-induced endophyte genes will be of considerable future interest.

7.5 *EAS* Cluster Evolution

In the study of the taxonomic distribution of the *eas* genes in epichloë strains, the cluster was found to be either completely absent or present. This pattern was also observed over a broad selection of strains tested for the presence of *lpsB* and *lpsA*, where neither gene was found without the other (Johnson *et al*., 2007b). The presence or absence of the cluster is strain- rather than species-dependent, indicating that the cluster has been frequently lost during evolution. This result was not expected as a similar, larger analysis with *ltm* genes showed that some strains contained different mixtures of genes, as well as several in which the whole cluster was present or absent (Young, 2005). The *LTM* cluster 3 in particular (*ltmE* and *ltmJ*) was frequently absent with the other two clusters remaining. Similarly, in fusaria different species contain different *tri* gene sets for tricothecene synthesis (Kimura *et al*., 1998a; Quarta *et al*., 2006). With both the *ltm* and *tri* genes the deletion of different genes can lead to a different chemotype, possibly with different bioactivity, for the producing organism, which may retain or give a new selective advantage in a given niche. Different ergot alkaloids have been shown to have different biological effects, albeit in a laboratory setting. Ergopeptines were shown to be active against black beetle (Ball *et al*., 1997), fall armyworm (Clay and Cheplick, 1989), mammals (Panaccione *et al*., 2006) and against the nematode *Pratylenchus scribneri* (Panaccione, 2005), while clavines are more cytostatic and active against bacteria (Panaccione, 2005) but also with some effects on mammals (Panaccione *et al*., 2006). If activity against insects and nematodes were most important for plant-epichloë symbiota, as intuitively may be expected, there would be selection pressure to keep the entire gene cluster to enable the production of ergovaline. In ecological niches where insect or nematode pressure is low, or where there is redundancy from other alkaloids for protection against a given herbivore, as work in this study may indicate (discussed in section 7.1), selection would be lost for the entire cluster, perhaps explaining its frequent loss. Further analysis of epichloë strains deleted in different *eas* genes is required to determine potential ecological roles for the various alkaloids which would shed further light on this question.

At the family level, the known ability to produce ergot alkaloids is restricted to some members of the Clavicipitaceae in the order Hypocreales and to some *Penicillium* species and *A. fumigatus* from the quite distant order Eurotiales (Schardl *et al.*, 2006). Why are genes for ergot alkaloid synthesis found so discontinuously in the fungal tree of life? There are two possibilities. One is that the gene cluster was present in a very early shared ancestor of Hypocreales and Eurotiales and has been frequently lost throughout ascomycete evolution. The second is that the cluster was horizontally transferred from one order to the other. Panaccione (2005) states that the codon usage and GC content of the *C. purpurea* and *A. fumigatus EAS* clusters are not consistent with a horizontal transfer event. The former thus seems more likely, although the conservation of gene order between the shared genes in the *A. fumigatus* cluster and the *N. lolii EAS* cluster 1 may provide some support for a transfer event, given both clusters are found in what are likely to be dynamic regions of the genome.

Both the *N. lolii* and *A. fumigatus* gene clusters are located in regions linked to the telomere – a common location for biosynthetic gene clusters as detected by fungal genome sequencing projects (Galagan *et al.*, 2005; Nierman *et al.*, 2005; Rehmeyer *et al.*, 2006). Other genes with "disposable" functions are often found subtelomerically, including gene families involved in antigenic variation in *Plasmodium* (Gardner *et al.*, 2002) and *Trypanosoma spp.* (Horn and Barry, 2005), and adhesin variability in *C. glabrata* (Castano *et al.*, 2005) and *S. cerevisiae* (Halme *et al.*, 2004). Subtelomeric regions of the genome evolve more rapidly than other regions of the chromosome, show high rates of interchromosomal recombination and gene duplication, and are frequently sites for relocation of one copy of a sequence duplicated elsewhere in the genome (Kellis *et al.*, 2003). The abundance of transposon relic sequences adds to the evolutionary potential of the *eas* cluster. In *Magnaporthe oryzae* transposable element clusters correlate with increased recombination rate, loss of synteny, gene duplication and sequence divergence from orthologous genes (Thon *et al.*, 2006). Adaptive gene relocation to a subtelomeric location was recently shown to be responsible for the formation of the *DAL* gene cluster for allantoin utilisation in yeast (Wong and Wolfe, 2005).

Although it is not known whether this is a general mode of gene cluster evolution, the subtelomeric location of the *EAS* cluster and many other fungal secondary metabolite biosynthetic gene clusters supports this hypothesis. Adaptive gene relocation is an attractive mechanism to explain different biosynthetic capabilities associated with the evolution of the ergopeptine-producing (contain *lpsA*, *lpsB* and *cloA*), and clavine-producing (contain *easI – easN*) fungi, subsequent to either the divergence of these lineages or transfer of the cluster.

Analysis of the *EAS* gene homologues in sequenced aspergillus genomes revealed that the cluster was possibly subject to an ancient duplication. Within the *A. fumigatus* genome, the gene cluster for fumitremorgen synthesis contains four genes homologous to genes in the *EAS* cluster. The similarity of the two clusters is highlighted by the *FTM* cluster being mistakenly labelled as an ergot alkaloid biosynthetic cluster previous to functional analysis (Sheppard *et al.*, 2005). Both of the clusters are found very near to an identical LINE element and to telomeres, perhaps indicating that the clusters arose from telomere recombination. If the *EAS* cluster is found in distant families due to gene loss in other organisms then it seems likely that it was an *EAS* cluster that was duplicated and that the duplicated cluster was neo-functionalised in *A. fumigatus* to synthesise fumitremorgens. In *A. nidulans* both clusters appear to have diverged and functional analysis is required to determine the product of these clusters. It seems likely that the progenitor *EAS* cluster contained an NRPS gene as each of the *EAS*-like clusters, barring the *EAS* cluster itself, have an associated NRPS; as of course do the clavicipitacean *EAS* clusters. Further support for this is that of all apergillus NRPS modules, the *C. purpurea* LpsB enzyme (and by extension the *N. lolii* enzyme) is most closely related to the first module of the *FtmA* protein found in the fumitremorgen cluster (Cramer *et al.*, 2006). A full analysis of the *EAS* genes within the aspergilli and other sequenced fungi was outside the scope of this thesis. Future analysis though seems likely to yield further interesting insights into *EAS* cluster evolution.

7.6 Summary

There is a long history of research into ergot alkaloids. This study contributes to this large body of work by advancing the knowledge of ergot alkaloid molecular genetics in epichloë. The full complement of genes found in *C. purpurea* have now been isolated from epichloë and shown to be linked within a complex subtelomeric gene cluster. The function of the LpsB NRPS in ergovaline synthesis was confirmed by mutational analysis. Evidence was provided against the hypothesis that ergovaline was primarily responsible for pasture protection against black beetle. It is now possible to perform similar analyses of each *eas* gene in epichloë, which will enable further elucidation of ergot alkaloid biochemistry and the ecological roles of intermediate alkaloids.

Sequence analysis revealed the presence of at least ten different transposable element relics interspersed between *EAS* clusters. Particularly interesting was the presence of at least three different MITEs in the *E. festucae* genome – more than is found to date in all other fungi examined. The recombination events observed at the *easA-easG* locus, where both deletion and duplication events appear to have been mediated by Toru-1 elements, highlights the potential impact of these elements on epichloë evolution.

Analysis of the *EAS* cluster in four different epichloë strains showed almost complete conservation of gene order but markedly different production of ergot alkaloids and *eas* gene expression. Key future research should test whether the regulatory mechanisms proposed are indeed involved in *eas* gene regulation.

This study provides a genetic foundation for elucidating further biochemical steps in the ergovaline pathway, the ecological role of individual ergot alkaloid compounds and understanding the evolution of *eas* genes and their regulation *in planta*.

8. Bibliography

Abe, Y., Suzuki, T., Ono, C., Iwamoto, K., Hosobuchi, M., and Yoshikawa, H. (2002). Molecular cloning and characterization of an ML-236B (compactin) biosynthetic gene cluster in Penicillium citrinum. Mol Genet Genomics **267,** 636-646.

Aharonowitz, Y., Cohen, G., and Martin, J.F. (1992). Penicillin and cephalosporin biosynthetic genes: Structure, organization, regulation, and evolution. Annu Rev Microbiol **46,** 461-495.

Ahn, J.-H., Cheng, Y.-Q., and Walton, J.D. (2002). An Extended Physical Map of the TOX2 Locus of Cochliobolus carbonum Required for Biosynthesis of HC-Toxin. Fungal Genet Biol **35,** 31-38.

Ahn, J.H., and Walton, J.D. (1998). Regulation of cyclic peptide biosynthesis and pathogenicity in Cochliobolus carbonum by TOXEp, a novel protein with a bZIP basic RNA-binding motif and four ankyrin repeats. Mol Gen Genet **260,** 462-469.

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

Almeida, R., and Allshire, R.C. (2005). RNA silencing and genome regulation. Trends Cell Biol **15,** 251-258.

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

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25,** 3389-3402.

Anaya, N., Isabel, M., and Roncero, G. (1995). *Skippy,* a retrotransposon from the fungal plant pathogen *Fusarium oxysporum*. Mol Gen Genet **249,** 637-647.

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

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

Bacon, C. W. (1988). Procedure for isolating the endophyte from tall fescue and screening isolates for ergot alkaloids. Appl Environ Microbiol **54,** 2615-2618

Bacon, C.W. (1995). Toxic endophyte-infected tall fescue and range grasses: historic perspectives. J Anim Sci **73,** 861-870.

Bacon, C.W., Porter, J.K., Robbins, J.D., and Luttrell, E.S. (1977). *Epichloe typhina* from toxic tall fescue grasses. Appl Environ Microbiol **34,** 576-581.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology (Menlo Park, California: AAAI Press), pp. 28-36.

Ball, O.J.-P., Christensen, M.J., and Prestidge, R.A. (1994). Effect of selected isolates of Acremonium endophytes on adult black beetle (*Heteronychus arator*) feeding. Proceedings of the 47th New Zealand Plant Protection Conference, 227-231.

Ball, O.J.P., Miles, C.O., and Prestidge, R.A. (1997). Ergopeptine Alkaloids and *Neotyphodium lolii*-Mediated Resistance in Perennial Ryegrass Against Adult *Heteronychus arator* (Coleoptera: Scarabaeidae). J Econ Entomol **90,** 1382-1391.

Basse, C.W., Kolb, S., and Kahmann, R. (2002). A maize-specifically expressed gene cluster in *Ustilago maydis*. Mol Microbiol **43,** 75-93.

Belshaw, P.J., Walsh, C.T., and Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. Science **284,** 486-489.

Bergendahl, V., Linne, U., and Marahiel, M.A. (2002). Mutational analysis of the C-domain in nonribosomal peptide synthesis. Eur J Biochem **269,** 620-629.

Billman-Jacobe, H., McConville, M.J., Haites, R.E., Kovacevic, S., and Coppel, R.L. (1999). Identification of a peptide synthetase involved in the biosynthesis of glycopeptidolipids of Mycobacterium smegmatis. Mol Microbiol **33,** 1244-1253.

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

Bohnert, H.U., Fudal, I., Dioh, W., Tharreau, D., Notteghem, J.L., and Lebrun, M.H. (2004). A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice. Plant Cell **16,** 2499-2513.

Bok, J.W., and Keller, N.P. (2004). LaeA, a regulator of secondary metabolism in *Aspergillus spp.* Eukaryot Cell **3,** 527-535.

Bok, J.W., Noordermeer, D., Kale, S.P., and Keller, N.P. (2006a). Secondary metabolic gene cluster silencing in *Aspergillus nidulans*. Mol Microbiol **61,** 1636-1645.

Bok, J.W., Hoffmeister, D., Maggio-Hall, L.A., Murillo, R., Glasner, J.D., and Keller, N.P. (2006b). Genomic Mining for *Aspergillus* Natural Products. Chem Biol **13,** 31.

Bradshaw, R.E., Bhatnagar, D., Ganley, R.J., Gillman, C.J., Monahan, B.J., and Seconi, J.M. (2002). *Dothistroma pini*, a forest pathogen, contains homologs of aflatoxin biosynthetic pathway genes. Appl Environ Microbiol **68,** 2885-2892.

Brakhage, A.A., Al-Abdallah, Q., Tuncher, A., and Sprote, P. (2005). Evolution of β-lactam biosynthesis genes and recruitment of trans-acting factors. Phytochemistry **66,** 1200-1210.

Brandsch, R., Hinkkanen, A.E., Mauch, L., Nagursky, H., and Decker, K. (1987). 6-Hydroxy-D-nicotine oxidase of *Arthrobacter oxidans*. Gene structure of the flavoenzyme and its relationship to 6-hydroxy-L-nicotine oxidase. Eur J Biochem **167,** 315-320.

Brown, D.W., Proctor, R.H., Dyer, R.B., and Plattner, R.D. (2003). Characterization of a *Fusarium* 2-Gene Cluster involved in trichothecene C-8 modification. J Agric Food Chem **51,** 7936-7944.

Brown, D.W., Dyer, R.B., McCormick, S.P., Kendra, D.F., and Plattner, R.D. (2004). Functional demarcation of the *Fusarium* core trichothecene gene cluster. Fungal Genet Biol **41,** 454-462.

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

Brown, M.P., Brown-Jenco, C.S., and Payne, G.A. (1999). Genetic and molecular analysis of aflatoxin biosynthesis. Fungal Genet Biol **26,** 81-98.

Bullock, W.O., Fernandez, J.M., and Short, J.M. (1987). XL1-Blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. BioTechniques **5,** 376-378.

Bultman, T.L., White, J.F.J., Bowdish, T.I., Welch, A.M., and Johnston, J. (1995). Mutualistic transfer of *Epichloe* spermatia by Phorbia flies. Mycologia **87,** 182-187.

Bureau, T.E., and Wessler, S.R. (1992). *Tourist*: A large family of small inverted repeat elements frequently associated with maize genes. Plant Cell **4,** 1283-1294.

Bureau, T.E., and Wessler, S.R. (1994a). *Stowaway*: A new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell **6,** 907-916.

Bureau, T.E., and Wessler, S.R. (1994b). Mobile inverted-repeat elements of the Tourist family are associated with the genes of many cereal grasses. Proc Natl Acad Sci U S A **91,** 1411-1415.

Bureau, T.E., Ronald, P.C., and Wessler, S.R. (1996). A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. Proc Natl Acad Sci U S A **93,** 8524-8529.

Burhans, D., Ramachandran, L., Wang, J., Liang, P., Patterton, H., Breitenbach, M., and Burhans, W. (2006). Non-random clustering of stress-related genes during evolution of the *S. cerevisiae* genome. BMC Evol Biol **6,** 58.

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

Byrd, A.D., Schardl, C.L., Songlin, P.J., Mogen, K.L., and Siegel, M.R. (1990). The β-tubulin gene of *Epichloë typhina* from perennial ryegrass (*Lolium perenne*). Curr Genet **18,** 347-354.

Calvo, A.M., Wilson, R.A., Bok, J.W., and Keller, N.P. (2002). Relationship between secondary metabolism and fungal development. Microbiol Mol Biol Rev **66,** 447-459.

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

Cambareri, E.B., Foss, H.M., Rountree, M.R., Selker, E.U., and Kinsey, J.A. (1996). Epigenetic control of a transposon-inactivated gene in *Neurospora* is dependent on DNA methylation. Genetics **143,** 137-146.

Carr, L.G., Skatrud, P.L., and Scheetz Ii, M.E. (1986). Cloning and expression of the isopenicillin N synthetase gene from *Penicillium chrysogenum*. Gene **48,** 257-266.

Casacuberta, J.M., and Santiago, N. (2003). Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene **311,** 1-11.

Casas-Flores, S., Rios-Momberg, M., Rosales-Saavedra, T., Martinez-Hernandez, P., Olmedo-Monfil, V., and Herrera-Estrella, A. (2006). Cross talk between a fungal blue-light perception system and the cyclic AMP signaling pathway. Eukaryot Cell **5,** 499-506.

Castano, I., Pan, S.J., Zupancic, M., Hennequin, C., Dujon, B., and Cormack, B.P. (2005). Telomere length control and transcriptional regulation of subtelomeric adhesins in *Candida glabrata*. Mol Microbiol **55,** 1246-1258.

Challis, G.L., Ravel, J., and Townsend, C.A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem Biol **7,** 211-224.

Chalvet, F., Grimaldi, C., Kaper, F., Langin, T., and Daboussi, M.-J. (2003). *Hop*, an Active Mutator-like Element in the Genome of the Fungus *Fusarium oxysporum*. Mol Biol Evol **20,** 1362-1375.

Cheah, M.T., Wachter, A., Sudarsan, N., and Breaker, R.R. (2007). Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature **447,** 497-500.

Chen, H., Lee, M.H., Daub, M.E., and Chung, K.R. (2007). Molecular analysis of the cercosporin biosynthetic gene cluster in *Cercospora nicotianae*. Mol Microbiol **64,** 755-770.

Chiu, H.-T., Hubbard, B.K., Shah, A.N., Eide, J., Fredenburg, R.A., Walsh, C.T., and Khosla, C. (2001). Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. Proc Natl Acad Sci U S A **98,** 8548-8553.

Christensen, M. J., Bennet, R. J., Ansari, H. A., Koga, H., Johnson, R. D., Bryan, G. T., Simpson, W. R., Koolard, J. P., Nickless, E. M., Voisey, C. R. (2007). Epichloe endophytes grow by intercalary hyphal extension in elongating grass leaves. Fungal Genet Biol **In press**

Christensen, M.J., Bennet, R.J., and Schmid, J. (2002). Growth of *Epichloe/Neotyphodium* and p-endophytes in leaves of *Lolium* and *Festuca* grasses. Mycol Res **106,** 93-106.

Christensen, M.J., Leuchtmann, A., Rowan, D.D., and Tapper, B.A. (1993). Taxonomy of *Acremonium* endophytes of tall fescue (*Festuca arundinacea*), meadow fescue (F. pratensis) and perennial ryegrass (*Lolium perenne*). Mycol Res **97,** 1083-1092.

Chung, K.-R. and Schardl, C. L. (1997). Sexual cycle and horizontal transmission of the grass symbiont, *Epichloe typhina*. Mycol res **101**, 295-301.

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

Clay, K., and Schardl, C.L. (2002). Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. Am Nat **160,** S99-S127.

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

Cohen, B.A., Mitra, R.D., Hughes, J.D., and Church, G.M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet **26,** 183-186.

Conti, E., Franks, N.P., and Brick, P. (1996). Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. Structure **4,** 287-298.

Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. EMBO J **16,** 4174-4183.

Correia, T., Grammel, N., Ortel, I., Keller, U., and Tudzynski, P. (2003). Molecular cloning and analysis of the ergopeptine assembly system in the ergot fungus *Claviceps purpurea*. Chem Biol **10,** 1281-1292.

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

Cramer Jr, R.A., Stajich, J.E., Yamanaka, Y., Dietrich, F.S., Steinbach, W.J., and Perfect, J.R. (2006). Phylogenomic analysis of non-ribosomal peptide synthetases in the genus *Aspergillus*. Gene **383,** 24-32.

Cultrone, A., Dominguez, Y.R., Drevet, C., Scazzocchio, C., and Fernandez-Martin, R. (2007). The tightly regulated promoter of the *xanA* gene of *Aspergillus nidulans* is included in a helitron. Mol Microbiol **63,** 1577-1587.

Cunningham, I., J. (1948). Tall fescue grass is poison for cattle. NZ J Agric **77,** 519-.

Cutler, N.S., Heitman, J., and Cardenas, M.E. (1999). TOR kinase homologs function in a signal transduction pathway that is conserved from yeast to mammals. Mol Cell Endocrinol **155,** 135-142.

Daboussi, M.J., and Capy, P. (2003). Transposable elements in filamentous fungi. Annu Rev Microbiol **57,** 275-299.

Damrongkool, P., Sedlock, A.B., Young, C.A., Johnson, R.D., Goetz, K.E., Scott, B., Schardl, C.L., and Panaccione, D.G. (2005). Structural analysis of a peptide synthetase gene required for ergopeptine production in the endophytic fungus *Neotyphodium lolii*. DNA Seq. **16,** 379-385.

Dawkins, R. (1976). The selfish gene. (Oxford: Oxford University Press).

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.R., Pan, H. et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature **434,** 980-986.

Demeke, T., Kidane, Y., and Wuhib, E. (1979). Ergotism - A report on an epidemic, 1977-78. Ethiop Med J **17,** 107-113.

Ding, Y., Bojja, R.S., and Du, L. (2004). *Fum3p*, a 2-Ketoglutarate-Dependent Dioxygenase Required for C-5 Hydroxylation of Fumonisins in *Fusarium verticillioides*, pp. 1931-1934.

Dittrich, H., and Kutchan, T.M. (1991). Molecular cloning, expression, and induction of berberine bridge enzyme, an enzyme essential to the formation of benzophenanthridine alkaloids in the response of plants to pathogenic attack, pp. 9969-9973.

Dobson, J.M. (1997). Molecular characterisation of the HMG CoA reductase gene from *Neotyphodium lolii*. MSc thesis, Institute of Molecular BioSciences (Massey University).

Dowzer, C.E.A., and Kelly, J.M. (1991). Analysis of the *creA* gene, a regulatory of carbon catabolite repression in *Aspergillus nidulans*. Mol Cell Biol **11,** 5701-5709.

Dufresne, M., Hua-Van, A., Abd el Wahab, H., M'Barek, S.B., Vasnier, C., Teysset, L., Kema, G.H.J., and Daboussi, M.-J. (2007). Transposition of a fungal miniature inverted-repeat transposable element through the action of a Tc1-like transposase. Genetics **175,** 441-452.

Dutnall, R.N., Tafrov, S.T., Sternglanz, R., and Ramakrishnan, V. (1998). Structure of the histone acetyltransferase Hat1: a paradigm for the GCN5-related N-acetyltransferase superfamily. Cell **94,** 427-438.

Eadie, M.J. (2003). Convulsive ergotism: Epidemics of the serotonin syndrome? Lancet Neurol **2,** 429-434.

Ehmann, D.E., Gehring, A.M., and Walsh, C.T. (1999). Lysine biosynthesis in *Saccharomyces cerevisiae*: mechanism of alpha-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. Biochemistry **38,** 6171-6177.

Ehmann, D.E., Trauger, J.W., Stachelhaus, T., and Walsh, C.T. (2000). Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. Chem Biol **7,** 765-772.

Elmi, A.A., West, C.P., Robbins, R.T., and Kirkpatrick, T.L. (1995). Endophyte effects on reproduction of a root-knot nematode (*Meloidogyne marylandi*) and osmotic adjustment in tall fescue. Grass Forage Sci **55,** 166-172.

Fernandes, M., Keller, N.P., and Adams, T.H. (1998). Sequence-specific binding by Aspergillus nidulans AflR, a C6 zinc cluster protein regulating mycotoxin biosynthesis. Mol Microbiol **28,** 1355-1365.

Feschotte, C., Zhang, X., and Wessler, S.R. (2002). Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. In Mobile DNA II, N.L. Craig, R. Craigie, M. Gellert, and A.M. Lambowitz, eds (Washington DC: American Society for Microbiology Press), pp. 1147-1158.

Fierro, F., Barredo, J.L., Diez, B., Gutierrez, S., Fernandez, F.J., and Martin, J.F. (1995). The penicillin gene cluster is amplified in tandem repeats linked by conserved hexanucleotide sequences. Proc Natl Acad Sci U S A **92,** 6200-6204.

Finking, R., and Marahiel, M.A. (2004). Biosynthesis of nonribosomal peptides. Annu Rev Microbiol.

Fischbach, M.A., and Walsh, C.T. (2006). Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. Chem Rev **106,** 3468-3496.

Fleetwood, D.J., Scott, B., Lane, G.A., Tanaka, A., and Johnson, R.D. (2007). A complex ergovaline gene cluster in epichloë endophytes of grasses. Appl Environ Microbiol **73,** 2571-2579.

Fletcher, L.R., and Harvey, I.C. (1981). An association of a Lolium endophyte with ryegrass staggers. N Z Vet J **29,** 185-186.

Floss, H.G. (1976). Biosynthesis of ergot alkaloids and related compounds. Tetrahedron **32,** 873-912.

Freitag, M., and Selker, E.U. (2005). Controlling DNA methylation: many roads to one modification. Curr Opin Genet Dev **15,** 191-199.

Freitag, M., Williams, R.L., Kothe, G.O., and Selker, E.U. (2002). A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. Proc Natl Acad Sci U S A **99,** 8802-8807.

Galagan, J.E., Calvo, S.E., Cuomo, C., Ma, L.J., Wortman, J.R., Batzoglou, S., Lee, S.I., Basturkmen, M., Spevak, C.C., Clutterbuck, J. et al. (2005). Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature **438,** 1105-1115.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., FitzHugh, W., Ma, L.J., Smirnov, S., Purcell, S. et al. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. Nature **422,** 859-868.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S. et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature **419,** 498-511.

Gebler, J.C., and Poulter, C.D. (1992). Purification and characterization of dimethylallyl tryptophan synthase from *Claviceps purpurea*. Arch Biochem Biophys **296,** 308-313.

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

Grayburn, W.S., and Selker, E.U. (1989). A natural case of RIP: Degeneration of the DNA sequence in an ancestral tandem duplication. Mol Cell Biol **9,** 4416-4421.

Gronover, C.S., Schorn, C., and Tudzynski, B. (2004). Identification of *Botrytis cinerea* genes up-regulated during infection and controlled by the Gα subunit BCG1 using suppression subtractive hybridization (SSH). Mol Plant-Microbe Interact **17,** 537-546.

Grundmann, A., and Li, S.M. (2005). Overproduction, purification and characterization of FtmPT1, a brevianamide F prenyltransferase from *Aspergillus fumigatus*. Microbiology **151,** 2199-2207.

Gwinn, K.D., Collins-Shepard, M.H., and Reddick, B.B. (1991). Tissue print-immunoblot, an accurate method for the detection of *Acremonium coenophialum* in tall fescue. Phytopathology **81,** 747-748.

Haarmann, T., Lorenz, N., and Tudzynski, P. (2007). Use of a nonhomologous end joining deficient strain (Δku70) of the ergot fungus *Claviceps purpurea* for identification of a nonribosomal peptide synthetase gene involved in ergotamine biosynthesis. Fungal Genet Biol **In Press**.

Haarmann, T., Ortel, I., Tudzynski, P., and Keller, U. (2006). Identification of the cytochrome P450 monooxygenase that bridges the clavine and ergoline alkaloid pathways. Chembiochem **7,** 645-652.

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

Haas, H. (2003). Molecular genetics of fungal siderophore biosynthesis and uptake: the role of siderophores in iron uptake and storage. Appl Microbiol Biotechnol **62,** 316-330.

Haas, H., and Marzluf, G.A. (1995). NRE, the major nitrogen regulatory protein of *Penicillium chrysogenum*, binds specifically to elements in the intergenic promoter regions of nitrate assimilation and penicillin biosynthetic gene clusters. Curr Genet **28,** 177-183.

Haese, A., Schubert, M., Herrmann, M., and Zocher, R. (1993). Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing *N*-methyldepsipeptide formation in *Fusarium scirpi*. Mol Microbiol **7,** 905-914.

Hahn, M., and Stachelhaus, T. (2004). Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. Proc Natl Acad Sci U S A **101,** 15585-15590.

Halme, A., Bumgarner, S., Styles, C., and Fink, G.R. (2004). Genetic and epigenetic regulation of the *FLO* gene family generates cell-surface variation in yeast. Cell **116,** 405-415.

He, C., Rusu, A.G., Poplawski, A.M., Irwin, J.A.G., and Manners, J.M. (1998). Transfer of a supernumerary chromosome between vegetatively incompatible biotypes of the fungus *Colletotrichum gloeosporioides*. Genetics **150,** 1459-1466.

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

Hoffmeister, D., and Keller, N.P. (2007). Natural products of filamentous fungi: Enzymes, genes, and their regulation. Nat Prod Rep **24,** 393-416.

Hohn, T.M., Krishna, R., and Proctor, R.H. (1999). Characterization of a transcriptional activator controlling trichothecene toxin biosynthesis. Fungal Genet Biol **26,** 224-235.

Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., Wilkinson, B., Sidebottom, P.J., Rudd, B.A.M., Hayes, M.A., Smith, C.P., and Micklefield, J. (2002). Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. Chem Biol **9,** 1175-1187.

Holmes, D.S., and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. Anal Biochem **114,** 193-197.

Horn, D., and Barry, J.D. (2005). The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. Chromosome Res **13,** 525-533.

Hua-Van, A., Henricourt, F., Capy, P., Daboussi, M.J., and Langin, T. (1998). Three highly divergent subfamilies of the impala transposable element coexist in the genome of the fungus *Fusarium oxysporum*. Mol Gen Genet **259,** 354-362.

Hua-Van, A., Daviere, J.-M., Kaper, F., Langin, T., and Daboussi, M.-J. (2000). Genome organization in *Fusarium oxysporum*: clusters of class II transposons. Curr Genet **37,** 339.

Hull, E.P., Green, P.M., Arst Jr, H.N., and Scazzocchio, C. (1989). Cloning and physical characterization of the L-proline catabolism gene cluster of *Aspergillus nidulans*. Mol Microbiol **3,** 553-559.

Hurst, L.D., Pal, C., and Lercher, M.J. (2004). The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet **5,** 299-310.

Hynes, M.J. (1975). Studies on the role of the *areA* gene in the regulation of nitrogen catabolism in *Aspergillus nidulans*. Aust J Biol Sci **28,** 301-313.

Idnurm, A., and Howlett, B.J. (2003). Analysis of loss of pathogenicity mutants reveals that repeat-induced point mutations can occur in the Dothideomycete *Leptosphaeria maculans*. Fungal Genet Biol **39,** 31-37.

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

Jekosch, K., and Kuck, U. (2000). Glucose dependent transcriptional expression of the *cre1* gene in *Acremonium chrysogenum* strains showing different levels of cephalosporin C production. Curr Genet **37,** 388-395.

Johnson, L.J., Johnson, R.D., Schardl, C.L., and Panaccione, D.G. (2003). Identification of differentially expressed genes in the mutualistic association of tall fescue with *Neotyphodium coenophialum*. Physiol Mol Plant Pathol **63,** 305-317.

Johnson, L.J., Steringa, M., Koulman, A., Christensen, M., Johnson, R.D., Voisey, C.R., Bryan, G., Lamont, I., and Rasmussen, S. (2007a). Biosynthesis of an extracellular siderophore is essential for maintenance of mutualistic endophyte-grass symbioses. In Proceedings of the 6th international symposium on fungal endophytes of grasses, A.J. Popay and E.R. Thom, eds (Dunedin, NZ: NZ Grassland Association), pp. 177-179.

Johnson, R., Voisey, C., Johnson, L., Pratt, J., Fleetwood, D., Khan, A., and Bryan, G. (2007b). Distribution of NRPS gene families within the *Neotyphodium/Epichloë* complex. Fungal Genet Biol **in press**.

Johnson, R.D., Johnson, L., Itoh, Y., Kodama, M., Otani, H., and Kohmoto, K. (2000). Cloning and characterization of a cyclic peptide synthetase gene from *Alternaria alternata* apple pathotype whose product is involved in AM-toxin synthesis and pathogenicity. Mol Plant Microbe Interact **13,** 742-753.

Jornvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J., and Ghosh, D. (1995). Short-chain dehydrogenases/reductases (SDR). Biochemistry **34,** 6003-6013.

Kazazian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. Science **303,** 1626-1632.

Keating, T.A., Ehmann, D.E., Kohli, R.M., Marshall, C.G., Trauger, J.W., and Walsh, C.T. (2001). Chain terminatin steps in nonribosomal peptide synthetase assembly lines: directed acyl-*S*-enzyme breakdown in antibiotic and siderophore biosynthesis. Chembiochem **2,** 99-107.

Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., and Wingender, E. (2003). MATCH: a tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res **31,** 3576-3579.

Keller, N.P., and Hohn, T.M. (1997). Metabolic Pathway Gene Clusters in Filamentous Fungi. Fungal Genet Biol **21,** 17-29.

Keller, N.P., Turner, G., and Bennett, J.W. (2005). Fungal secondary metabolism - From biochemistry to genomics. Nat Rev Microbiol **3,** 937-947.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423,** 241-254.

Kennedy, J., Auclair, K., Kendrew, S.G., Park, C., Vederas, J.C., and Hutchinson, C.R. (1999). Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. Science **284,** 1368-1372.

Keszenman-Pereyra, D., Lawrence, S., Twfieg, M.E., Price, J., and Turner, G. (2003). The *npgA/cfwA* gene encodes a putative 4'-phosphopantetheinyl transferase which is essential for penicillin biosynthesis in *Aspergillus nidulans*. Curr Genet **43,** 186-190.

Kidwell, M.G., and Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci U S A **94,** 7704.

Kimura, M., Matsumoto, G., Shingu, Y., Yoneyama, K., and Yamaguchi, I. (1998a). The mystery of the trichothecene 3-O-acetyltransferase gene. Analysis of the region around *Tri101* and characterization of its homologue from *Fusarium sporotrichioides*. FEBS Lett **435,** 163-168.

Kimura, M., Kaneko, I., Komiyama, M., Takatsuki, A., Koshino, H., Yoneyama, K., and Yamaguchi, I. (1998b). Trichothecene 3-O-acetyltransferase protects both the producing organism and transformed yeast from related mycotoxins. Cloning and characterization of *Tri101*. J Biol Chem **273,** 1654-1661.

Kodama, M., Rose, M.S., Yang, G., Yun, S.H., Yoder, O.C., and Turgeon, B.G. (1999). The translocation-associated *Tox1* locus of *Cochliobolus heterostrophus* is two genetic elements on two different chromosomes. Genetics **151,** 585-596.

Kohli, R.M., Trauger, J.W., Schwarzer, D., Marahiel, M.A., and Walsh, C.T. (2001). Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. Biochemistry **40,** 7099-7108.

Kouzminova, E., and Selker, E.U. (2001). *dim-2* encodes a DNA methyltransferase responsible for all known cytosine methylation in *Neurospora*. EMBO J. **20,** 4309-4323.

Kroken, S., Glass, N.L., Taylor, J.W., Yoder, O.C., and Turgeon, B.G. (2003). Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. Proc Natl Acad Sci U S A **100,** 15670-15675.

Krupinski, V.M., Robbers, J.E., and Floss, H.G. (1976). Physiological study of ergot: induction of alkaloid synthesis by tryptophan at the enzymatic level. J Bacteriol **125,** 158-165.

Kuldau, G.A., Liu, J., White, J.F., Jr., Siegel, M.R., and Schardl, C.L. (1997). Molecular systematics of Clavicipitaceae supporting monophyly of genus *Epichloë* and form genus *Ephelis*. Mycologia **89,** 431-441.

Kupfer, D.M., Drabenstot, S.D., Buchanan, K.L., Lai, H., Zhu, H., Dyer, D.W., Roe, B.A., and Murphy, J.W. (2004). Introns and Splicing Elements of Five Diverse Fungi. Eukaryot Cell **3,** 1088-1100.

Kutil, B.L., Greenwald, C., Liu, G., Spiering, M.J., Schardl, C.L., and Wilkinson, H.H. (2007). Comparison of loline alkaloid gene clusters across fungal endophytes: Predicting the co-regulatory sequence motifs and the evolutionary history. Fungal Genet Biol **In Press**.

Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T. (1996). A new enzyme superfamily - the phosphopantetheinyl transferases. Chem Biol **3,** 923-936.

Landan, G., Cohen, G., Aharonowitz, Y., Shuali, Y., Graur, D., and Shiffman, D. (1990). Evolution of isopenicillin N synthase genes may have involved horizontal gene transfer. Mol Biol Evol **7,** 399-406.

Latch, G.C.M., and Christensen, M.J. (1985). Artificial infection of grasses with endophytes. Ann Appl Biol **107,** 17-24.

Latch, G.C.M., Hunt, W.F., and Musgrave, D.R. (1985). Endophytic fungi affect growth of perennial ryegrass. N Z J Agric Res **28,** 165-168.

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

Lercher, M.J., Urrutia, A.O., and Hurst, L.D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet **31,** 180-183.

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

Lev, S., Hadar, R., Amedeo, P., Baker, S.E., Yoder, O.C., and Horwitz, B.A. (2005). Activation of an AP1-like transcription factor of the maize pathogen *Cochliobolus heterostrophus* in response to oxidative stress and plant signals. Eukaryot Cell **4,** 443-454.

Linne, U., and Marahiel, M.A. (2000). Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. Biochemistry **39,** 10439-10447.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., Carrington, J.C., Doerge, R.W., Colot, V., and Martienssen, R. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature **430,** 471-476.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods **25,** 402.

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

Maiya, S., Grundmann, A., Li, S.M., and Turner, G. (2006). The fumitremorgin gene cluster of *Aspergillus fumigatus*: Identification of a gene encoding brevianamide F synthetase. ChemBioChem **7,** 1062-1069.

Mantle, P.G., and Nisbet, L.J. (1976). Differentiation of *Claviceps purpurea* in axenic culture. J. Gen. Microbiol. **93,** 321-334.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. Chemical Reviews **97,** 2651-2673.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z. et al. (2005). CDD: a Conserved Domain Database for protein classification. Nucleic Acids Res **33,** D192-196.

Margolin, B.S., Garrett-Engele, P.W., Stevens, J.N., Fritz, D.Y., Garrett-Engele, C., Metzenberg, R.L., and Selker, E.U. (1998). A Methylated *Neurospora* 5S rRNA Pseudogene Contains a Transposable Element Inactivated by Repeat-Induced Point Mutation. Genetics **149,** 1787-1797.

Martin, J.F. (1998). New aspects of genes and enzymes for β-lactam antibiotic biosynthesis. Appl Microbiol Biotechnol **50,** 1-15.

Marzluf, G.A. (1997). Genetic regulation of nitrogen metabolism in the fungi. Microbiol Mol Biol Rev **61,** 17-32.

Mate, M.J., Zamocky, M., Nykyri, L.M., Herzog, C., Alzari, P.M., Betzel, C., Koller, F., and Fita, I. (1999). Structure of catalase-A from *Saccharomyces cerevisiae*. J Mol Biol **286,** 135-149.

Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V. et al. (2003). TRANSFAC(R): transcriptional regulation, from patterns to profiles, Nucleic Acids Res **31**: 374-378.

Mayr, E. (1999). Understanding evolution. Trends Ecol Evol **14,** 372-373.

McMillan, L.K., Carr, R.L., Young, C.A., Astin, J.W., Lowe, R.G., Parker, E.J., Jameson, G.B., Finch, S.C., Miles, C.O., McManus, O.B., Schmalhofer, W.A., Garcia, M.L., Kaczorowski, G.J., Goetz, M., Tkacz, J.S., and Scott, B. (2003). Molecular analysis of two cytochrome P450 monooxygenase genes required for paxilline biosynthesis in *Penicillium paxilli*, and effects of paxilline intermediates on mammalian maxi-K ion channels. Mol Genet Genomics **270,** 9-23.

Meek, I.B., Peplow, A.W., Ake Jr, C., Phillips, T.D., and Beremand, M.N. (2003). *Tri1* encodes the cytochrome P450 monooxygenase for C-8 hydroxylation during trichothecene biosynthesis in *Fusarium sporotrichioides* and resides upstream of another new *Tri* gene. Appl Environ Microbiol **69,** 1607-1613.

Mihlan, M., Homann, V., Liu, T.W.D., and Tudzynski, B. (2003). AREA directly mediates nitrogen regulation of gibberellin biosynthesis in *Gibberella fujikuroi*, but its activity is not affected by NMR. Mol Microbiol **47,** 975-991.

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

Moon, C.D., Tapper, B.A., and Scott, B. (1999). Identification of epichloe endophytes in planta by a microsatellite-based PCR fingerprinting assay with automated analysis. Appl Environ Microbiol **65,** 1268-1279.

Moon, C.D., Scott, B., Schardl, C.L., and Christensen, M.J. (2000). The evolutionary origins of *Epichloë* endophytes from annual ryegrasses. Mycologia **92,** 1103-1118.

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

Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2002). Ways of assembling complex natural products on modular nonribosomal peptide synthetases. Chembiochem **3,** 490-504.

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

Namiki, F., Matsunaga, M., Okuda, M., Inoue, I., Nishi, K., Fujita, Y., and Tsuge, T. (2001). Mutation of an arginine biosynthesis gene causes reduced pathogenicity in *Fusarium oxysporum* f. sp. *melonis*. Mol Plant-Microbe Interact **14,** 580-584.

Narasaiah, K., Sashidhar, R., and Subramanyam, C. (2006). Biochemical analysis of oxidative stress in the production of aflatoxin and its precursor intermediates. Mycopathologia **162,** 179-189.

Nierman, W.C., Pain, A., Anderson, M.J., Wortman, J.R., Kim, H.S., Arroyo, J., Berriman, M., Abe, K., Archer, D.B., Bermejo, C. *et al.* (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. Nature **438,** 1151-1156.

Oide, S., Moeder, W., Krasnoff, S., Gibson, D., Haas, H., Yoshioka, K., and Turgeon, B.G. (2006). *NPS6*, encoding a nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycetes. Plant Cell **18,** 2836-2853.

Panaccione, D.G. (2005). Origins and significance of ergot alkaloid diversity in fungi. FEMS Microbiol Lett **251,** 9-17.

Panaccione, D.G., and Schardl, C.L. (2002). Molecular genetics of ergot alkaloid biosynthesis. In Clavicipitalean fungi: evolutionary biology, chemistry, biocontrol and cultural impacts, J.F. White, Jnr. , C.W. Bacon, N.L. Hywel-Jones, and J.W. Spatafora, eds (New York: Marcel Dekker Inc.), pp. 399-424.

Panaccione, D.G., Pitkin, J.W., Walton, J.D., and Annis, S.L. (1996). Transposon-like sequences at the *TOX2* locus of the plant-pathogenic fungus *Cochliobolus carbonum*. Gene **176,** 103-109.

Panaccione, D.G., Tapper, B.A., Lane, G.A., Davies, E., and Fraser, K. (2003). Biochemical outcome of blocking the ergot alkaloid pathway of a grass endophyte. J Agric Food Chem **51,** 6429-6437.

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

Panaccione, D.G., Cipoletti, J.R., Sedlock, A.B., Blemings, K.P., Schardl, C.L., Machado, C., and Seidel, G.E. (2006). Effects of ergot alkaloids on food preference and satiety in rabbits, as assessed with gene-knockout endophytes in perennial ryegrass (*Lolium perenne*). J Agric Food Chem. **54,** 4582-4587.

Pardue, M.L., Danilevskaya, O.N., Lowenhaupt, K., Slot, F., and Traverse, K.L. (1996). Drosophila telomeres: New views on chromosome evolution. Trends Genet **12,** 48-52.

Parish, J.A., McCann, M.A., Watson, R.H., Hoveland, C.S., Hawkins, L.L., Hill, N.S., and Bouton, J.H. (2003a). Use of nonergot alkaloid-producing endophytes for alleviating tall fescue toxicosis in sheep. J Anim Sci **81,** 1316-1322.

Parish, J.A., McCann, M.A., Watson, R.H., Paiva, N.N., Hoveland, C.S., Parks, A.H., Upchurch, B.L., Hill, N.S., and Bouton, J.H. (2003b). Use of nonergot alkaloid-producing endophytes for alleviating tall fescue toxicosis in stocker cattle. J Anim Sci **81,** 2856-2868.

Pavela-Vrancic, M., Dieckmann, R., and Von Döhren, H. (2004). ATPase activity of non-ribosomal peptide synthetases. Biochim Biophys Acta **1696,** 83-91.

Pazirandeh, M., Chirala, S.S., Huang, W.Y., and Wakil, S.J. (1989). Characterization of recombinant thioesterase and acyl carrier protein domains of chicken fatty acid synthase expressed in *Escherichia coli*. J Biol Chem **264,** 18195-18201.

Pedley, K.F., and Walton, J.D. (2001). Regulation of cyclic peptide biosynthesis in a plant pathogenic fungus by a novel transcription factor. Proc Natl Acad Sci U S A **98,** 14174-14179.

Penalva, M.A., Moya, A., Dopazo, J., and Ramon, D. (1990). Sequences of isopenicillin N synthetase genes suggest horizontal gene transfer from prokaryocytes to eukarocytes. Proceedings of the Royal Society - Biological Sciences (Series B) **241,** 164-169.

Perrin, R.M., Fedorova, N.D., Bok, J.W., Cramer, R.A., Wortman, J.R., Kim, H.S., Nierman, W.C., and Keller, N.P. (2007). Transcriptional regulation of chemical diversity in *Aspergillus fumigatus* by LaeA. PLoS Pathogens **3,** e50.

Philipson, M.N., and Christey, M.C. (1986). The relationship of host and endophyte during flowering, seed formation, and germination on *Lolium perenne*. N Z J Bot **24,** 125-134.

Pignocchi, C., and Foyer, C.H. (2003). Apoplastic ascorbate metabolism and its role in the regulation of cell signalling. Curr Opin Plant Biol **6,** 379-389.

Pontecorvo, G., Roper, J.A., Hemmons, L.M., MacDonald, K.D., and Bufton, A.W. (1953). The genetics of *Aspergillus nidulans*. Adv Genet **5,** 141-238.

Popay, A.J., and Baltus, J.G. (2001). Black beetle damage to perennial ryegrass infeted with AR1 endophyte. Proc NZ Grass Assoc **63,** 267-271.

Popay, A.J., Jensen, J.G., and Cooper, B.M. (2005). The effect of non-toxic endophytes in tall fescue on two major insect pests. Proc NZ Grass Assoc **67,** 169-173.

Porter, J. K., Bacon, C. W., Robbins, J. D., and Betowski, D. (1981). Ergot alkaloid identification in clavicipitaceae systemic fungi of pasture grasses. J Agric Food Chem **29,** 653-657.

Pospiech, A., Bietenhader, J., and Schupp, T. (1996). Two multifunctional peptide synthetases and an O-methyltransferase are involved in the biosynthesis of the DNA-binding antibiotic and antitumour agent saframycin Mx1 from *Myxococcus xanthus*. Microbiology **142 (Pt 4),** 741-746.

Proctor, R.H., Hohn, T.M., McCormick, S.P., and Desjardins, A.E. (1995). *Tri6* encodes an unusual zinc finger protein involved in regulation of trichothecene biosynthesis in *Fusarium sporotrichioides*. Appl Environ Microbiol **61,** 1923-1930.

Proctor, R.H., Desjardins, A.E., Plattner, R.D., and Hohn, T.M. (1999). A polyketide synthase gene required for biosynthesis of fumonisin mycotoxins in *Gibberella fujikuroi* mating population A. Fungal Genet Biol **27,** 100-112.

Proctor, R.H., Brown, D.W., Plattner, R.D., and Desjardins, A.E. (2003). Co-expression of 15 contiguous genes delineates a fumonisin biosynthetic gene cluster in Gibberella moniliformis. Fungal Genet Biol **38,** 237-249.

Punt, P.J., Oliver, R.P., Dingemanse, M.A., Pouwels, P.H., and van den Hondel, C.A. (1987). Transformation of *Aspergillus* based on the hygromycin B resistance marker from *Escherichia coli*. Gene **56,** 117-124.

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

Quarta, A., Mita, G., Haidukowski, M., Logrieco, A., Mule, G., and Visconti, A. (2006). Multiplex PCR assay for the identification of nivalenol, 3- and 15-acetyl-deoxynivalenol chemotypes in *Fusarium*. FEMS Microbiol Lett **259,** 7-13.

Quigley, F.R., and Floss, H.G. (1981). Mechanism of amino acid α-hydroxylation and formation of the lysergyl moiety in ergotamine biosynthesis. J Org Chem **46,** 464-466.

Ramussen, J.P., Taylor, A.H., Ma, L.J., Purcell, S., Kempken, F., and Catcheside, D.E.A. (2004). *Guest*, a transposable element belonging to the Tc1/*mariner* superfamily is an ancient invader of *Neurospora* genomes. Fungal Genet Biol **41,** 52-61.

Rawlings, B.J. (1997). Biosynthesis of polyketides. Nat Prod Rep **14,** 523-556.

Rehmeyer, C., Li, W., Kusaba, M., Kim, Y.-S., Brown, D., Staben, C., Dean, R., and Farman, M. (2006). Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae*. Nucleic Acids Res **34,** 4685-4701.

Reithner, B., Brunner, K., Schuhmacher, R., Peissl, I., Seidl, V., Krska, R., and Zeilinger, S. (2005). The G protein α subunit Tga1 of *Trichoderma atroviride* is involved in chitinase formation and differential production of antifungal metabolites. Fungal Genet Biol **42,** 749-760.

Reuter, K., Mofid, M.R., Marahiel, M.A., and Ficner, R. (1999). Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily. EMBO J. **18,** 6823-6831.

Rice, P., Longden, L., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet **16,** 276-277.

Riederer, B., Han, M., and Keller, U. (1996). D-lysergyl peptide synthetase from the ergot fungus *Claviceps purpurea*. J Biol Chem **271,** 27524-27530.

Rosen, S., Yu, J.H., and Adams, T.H. (1999). The *Aspergillus nidulans sfaD* gene encodes a G protein α subunit that is required for normal growth and repression of sporulation. EMBO J. **18,** 5592-5600.

Rowan, D.D. (1993). Lolitrems, peramine and paxilline: mycotoxins of the ryegrass/endophyte interaction. Agric Ecosyst Environ **44,** 1-4.

Saikia, S., Parker, E.J., Koulman, A., and Scott, B. (2007). Defining paxilline biosynthesis in *Penicillium paxilli*: functional characterisation of two cytochrome P450 monooxygenases. J Biol Chem **282,** 16829-16837.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual. (Cold Spring Harbour NY: Cold Spring Harbor Laboratory Press).

Schardl, C.L. (1996). *Epichloë* species: fungal symbionts of grasses. Annu Rev Phytopathol **34,** 109-130.

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

Schardl, C.L., and Clay, K. (1997). Evolution of mutalistic endophytes from plant pathogens. In The Mycota V part B (Berlin Heidelburg: Springer-Verlag), pp. 221-238.

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

Schardl, C.L., Leuchtmann, A., and Spiering, M.J. (2004). Symbioses of grasses with seedborne fungal endophytes. Annu Rev Plant Biol **55,** 315-340.

Schardl, C.L., Panaccione, D.G., and Tudzynski, P. (2006). Ergot alkaloids - biology and molecular biology. In Alkaloids: Chemistry and Biology, G.A. Cordell, ed (San Diego: Academic Press), pp. 45-86.

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

Schiestl, F.P., Steinebrunner, F., Schulz, C., von Reuß, S., Francke, W., Weymuth, C., and Leuchtmann, A. (2006). Evolution of 'pollinator'- attracting signals in fungi. Biol Letters **2,** 401-404.

Schmoll, M., Franchi, L., and Kubicek, C.P. (2005). Envoy, a PAS/LOV domain protein of *Hypocrea jecorina* (anamorph *Trichoderma reesei*), modulates cellulase gene transcription in response to light. Eukaryot Cell **4,** 1998-2007.

Schoustra, S.E., Debets, A.J.M., Slakhorst, M., and Hoekstra, R.F. (2007). Mitotic recombination accelerates adaptation in the fungus *Aspergillus nidulans.* PLoS Genetics **3,** e68.

Scott-Craig, J.S., Panaccione, D.G., Pocard, J.A., and Walton, J.D. (1992). The cyclic peptide synthetase catalyzing HC-toxin production in the filamentous fungus *Cochliobolus carbonum* is encoded by a 15.7-kilobase open reading frame. J Biol Chem **267,** 26044-26049.

Scott, B. (2001). *Epichloë* endophytes: fungal symbionts of grasses. Curr Opin Microbiol **4,** 393-398.

Scott, B. (2003a). Functional analysis of the perennial ryegrass - *Epichloë* endophyte interaction. In Molecular breeding of forage and turf, A. Hopkins, Z.Y. Wang, R. Mian, M. Sledge, and R. Barker, eds (The Netherlands: Kluwer Academic.) pp. 133-144

Scott, B. (2003b). The evolution of gene clusters in plant-associated filamentous fungi. NZ BioScience, 22-26.

Scott, D.B., Jameson, G.B., and Parker, E.J. (2004). Isoprenoids: gene clusters and chemical puzzles. In Advances in fungal biotechnology for industry, agriculture, and medicine, J.S. Tkacz and L. Lange, eds (Kluwer Academic/Plenum Publishers.) pp. 163-200

Selker, E.U. (2002). Repeat-induced gene silencing in fungi. Adv Genet **46,** 439-450.

Selker, E.U., Freitag, M., Kothe, G.O., Margolin, B.S., Rountree, M.R., Allis, C.D., and Tamaru, H. (2002). Induction and maintenance of nonsymmetrical DNA methylation in *Neurospora*. Proc Natl Acad Sci U S A **99 Suppl 4,** 16485-16490.

Seo, J.A., Proctor, R.H., and Plattner, R.D. (2001). Characterization of four clustered and coregulated genes associated with fumonisin biosynthesis in *Fusarium verticillioides*. Fungal Genet Biol **34,** 155-165.

Sheppard, D.C., Doedt, T., Chiang, L.Y., Kim, H.S., Chen, D., Nierman, W.C., and Filler, S.G. (2005). The *Aspergillus fumigatus* StuA protein governs the up-regulation of a discrete transcriptional program during the acquisition of developmental competence. Mol Biol Cell **16,** 5866-5879.

Shimizu, K., and Keller, N.P. (2001). Genetic involvement of a cAMP-dependent protein kinase in a G protein signaling pathway regulating morphological and chemical transitions in *Aspergillus nidulans*. Genetics **157,** 591-600.

Siewers, V., Kokkelink, L., Smedsgaard, J., and Tudzynski, P. (2006). Identification of an abscisic acid gene cluster in the grey mold *Botrytis cinerea*. Appl. Environ. Microbiol. **72,** 4619-4626.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet **8,** 272-285.

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

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

Stachelhaus, T., Huser, A., and Marahiel, M.A. (1996). Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. Chemistry and Biology **3,** 913-921.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol **6,** 493-505.

Staunton, J., and Weissman, K.J. (2001). Polyketide biosynthesis: a millennium review. Nat Prod Rep **18,** 380-416.

Stein, T., Vater, J., Kruft, V., Otto, A., Wittmann-Liebold, B., Franke, P., Panico, M., McDowell, R., and Morris, H.R. (1996). The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. J Biol Chem **271,** 15428-15435.

Tag, A., Hicks, J., Garifullina, G., Ake C, Jr., Phillips, T.D., Beremand, M., and Keller, N. (2000). G-protein signalling mediates differential production of toxic secondary metabolites. Mol Microbiol **38,** 658-665.

Takemoto, D., Tanaka, A., and Scott, B. (2006). A p67Phox-like regulator is recruited to control hyphal branching in a fungal-grass mutualistic symbiosis. Plant Cell **18,** 2807-2821.

Talbot, N.J., Ebbole, D.J., and Hamer, J.E. (1993). Identification and characterization of MPG1, a gene involved in pathogenicity from the rice blast fungus *Magnaporthe grisea*. Plant Cell **5,** 1575-1590.

Tamaru, H., and Selker, E.U. (2001). A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. Nature **414,** 277-283.

Tanaka, A., Tapper, B.A., Popay, A., Parker, E.J., and 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. Mol Microbiol **57,** 1036-1050.

Tanaka, A., Christensen, M.J., Takemoto, D., Park, P., and Scott, B. (2006). Reactive oxygen species play a role in regulating a fungus-perennial ryegrass mutualistic interaction. Plant Cell **18,** 1052-1066.

Tapper, B.A., and Latch, G.C.M. (1999). Selection against toxin production in endophyte-infected perennial ryegrass. Grassland Research and Practise **7,** 107-111.

Teichert, S., Wottawa, M., Schonig, B., and Tudzynski, B. (2006). Role of the *Fusarium fujikuroi* TOR kinase in nitrogen regulation and secondary metabolism. Eukaryot Cell **5,** 1807-1819.

Thara, V.K., Fellers, J.P., and Zhou, J.-M. (2003). *In planta* induced genes of *Puccinia triticina*. Mol Plant Pathol **4,** 51-56.

Thon, M.R., Pan, H., Diener, S., Papalas, J., Taro, A., Mitchell, T.K., and Dean, R.A. (2006). The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. Genome Biology **7,** doi:10.1186/gb-2006-1187-1182-r1116.

Tilburn, J., Sarkar, S., Widdick, D.A., Espeso, A., Orejas, M., Mungroo, J., Penalva, M.A., and Arst Jr, H.N. (1995). The *Aspergillus* PacC zinc finger transcription factor mediates regulation of both acid- and alkaline-expressed genes by ambient pH. EMBO J. **14,** 779-790.

Tkacz, J.S. (2000). Polyketide and peptide products of endophytic fungi: variations on two biosynthetic themes of secondary metabolism. In Microbial endophytes, C.W. Bacon and J.J.F. White, eds (New York: Marcel Dekker, Inc.), pp. 263-294.

Tsai, H.-F., Liu, J.-S., Staben, C., Christensen, M.J., Latch, G.C.M., Siegel, M.R., and Schardl, C.L. (1994). Evolutionary diversification of fungal endophytes of tall fescue grass by hybridization with *Epichloë* species. Proceedings of the National Academy of Sciences (USA) **91,** 2542-2546.

Tsai, H.F., Wang, H., Gebler, J.C., Poulter, C.D., and Schardl, C.L. (1995). The *Claviceps purpurea* gene encoding dimethylallyltryptophan synthase, the committed step for ergot alkaloid biosynthesis. Biochem Biophys Res Commun **216,** 119-125.

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

Tudzynski, B., and Holter, K. (1998). Gibberellin biosynthetic pathway in *Gibberella fujikuroi*: evidence for a gene cluster. Fungal Genet Biol **25,** 157-170.

Tudzynski, B., Hedden, P., Carrera, E., and Gaskin, P. (2001a). The P450-4 gene of *Gibberella fujikuroi* encodes ent-kaurene oxidase in the gibberellin biosynthesis pathway. Appl Environ Microbiol **67,** 3514-3522.

Tudzynski, B., Rojas, M.C., Gaskin, P., and Hedden, P. (2002). The gibberellin 20-oxidase of *Gibberella fujikuroi* Is a multifunctional monooxygenase. Journal Biological Chemistry **277,** 21246-21253.

Tudzynski, P., and Scheffer, J. (2004). *Claviceps purpurea*: Molecular aspects of a unique pathogenic lifestyle. Mol Plant Pathol **5,** 377-388.

Tudzynski, P., Correia, T., and Keller, U. (2001b). Biotechnology and genetics of ergot alkaloids. Appl Microbiol Biotechnol **57,** 593-605.

Tudzynski, P., Holter, K., Correia, T., Arntz, C., Grammel, N., and Keller, U. (1999). Evidence for an ergot alkaloid gene cluster in *Claviceps purpurea*. Mol Gen Genet **261,** 133-141.

Unsold, I.A., and Li, S.M. (2005). Overproduction, purification and characterization of FgaPT2, a dimethylallyltryptophan synthase from *Aspergillus fumigatus*. Microbiology **151,** 1499-1505.

Unsold, I.A., and Li, S.M. (2006). Reverse prenyltransferase in the biosynthesis of fumigaclavine C in *Aspergillus fumigatus*: gene expression, purification, and characterization of fumigaclavine C synthase FGAPT1. Chembiochem **7,** 158-164.

van den Ackerveken, G.F.J.M., Dunn, R.M., Cozijnsen, A.J., Vossen, J.P.M.J., van den Broek, H.W.J., and de Wit, P.J.G.M. (1994). Nitrogen limitation induces expression of the avirulence gene *avr9* in the tomato pathogen *Cladosporium fulvum*. Mol Gen Genet **243,** 277-285.

Vitte, C., and Panaud, O. (2003). Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. Mol Biol Evol **20,** 528-540.

von Döhren, H., Keller, U., Vater, J., and Zocher, R. (1997). Multifunctional peptide synthetases. Chem Rev **97,** 2675-2705.

Walsh, C.T. (2004). Polyketide and nonribosomal peptide antibiotics: modularity and versatility, pp. 1805-1810.

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

Walton, J.D. (2006). HC-toxin. Phytochemistry **67,** 1406-1413.

Walton, J.D., Panaccione, D.G., and Hallen, H.E. (2004). Peptide synthesis without ribosomes. In Advances in Fungal Biotechnology for Industry, Agriculture and Medicine, J.S. Tkacz and L. Lange, eds (Kluwer Academic/Plenum Publishers), pp. 127-162.

Walzel, B., Riederer, B., and Keller, U. (1997). Mechanism of alkaloid cyclopeptide formation in the ergot fungus *Claviceps purpurea*. Chem Biol **4,** 223-230.

Wang, J., Machado, C., Panaccione, D.G., Tsai, H.F., and Schardl, C.L. (2004). The determinant step in ergot alkaloid biosynthesis by an endophyte of perennial ryegrass. Fungal Genet Biol **41,** 189-198.

Wessler, S.R., Bureau, T.E., and White, S.E. (1995). LTR-retrotransposons and MITEs: Important players in the evolution of plant genomes. Current Opinion in Genetics and Development **5,** 814-821.

White, J.F.J., Morrow, A.C., Morgan-Jones, G., and Chambless, D.A. (1991). Endophyte-host associations in forage grasses. XIV. Primary stromata formation and seed transmission in *Epichloë typhina*: developmental and regulatory aspects. Mycologia **83,** 72-81.

Wiest, A., Grzegorski, D., Xu, B.W., Goulard, C., Rebuffat, S., Ebbole, D.J., Bodo, B., and Kenerley, C. (2002). Identification of peptaibols from *Trichoderma virens* and cloning of a peptaibol synthetase. J Biol Chem **277,** 20862-20868.

Wilkinson, H.H., Siegel, M.R., Blankenship, J.D., Mallory, A.C., Bush, L.P., and Schardl, C.L. (2000). Contribution of fungal loline alkaloids to protection from aphids in a grass-endophyte mutualism. Mol Plant-Microbe Interact **13,** 1027-1033.

Williams, R.E., and Bruce, N.C. (2002). 'New uses for an old enzyme' - The old yellow enzyme family of flavoenzymes. Microbiology **148,** 1607-1614.

Woakes, E. (1868). On ergot of rye in the treatment of neuralgia. British Med J **2,** 360-361.

Wong, S., and Wolfe, K.H. (2005). Birth of a metabolic gene cluster in yeast by adaptive gene relocation. Nat Genet **37,** 777-782.

Wu, S.-C., Kauffmann, S., Darvill, A.G., and Albersheim, P. (1995). Purification, cloning and characterization of two xylanases from *Magnaporthe grisea*, the rice blast fungus. Mol Plant-Microbe Interact **8,** 506-514.

Wubben, J.P., Joosten, M.H.A.J., and de Wit, P.J.G.M. (1994). Expression and localization of two in planta induced extracellular proteins of the fungal tomato pathogen *Cladosporium fulvum*. Mol Plant-Microbe Interact **7,** 516-524.

Yang, G., Lee, Y.-H., Jiang, Y., Shi, X., Kertbundit, S., and Hall, T.C. (2005a). A two-edged role for the transposable element *Kiddo* in the rice *ubiquitin2* Promoter. Plant Cell **17,** 1559-1568.

Yang, Z., Rogers, L.M., Song, Y., Guo, W., and Kolattukudy, P.E. (2005b). Homoserine and asparagine are host signals that trigger in planta expression of a pathogenesis gene in *Nectria haematococca*. Proc Natl Acad Sci U S A **102,** 4197-4202.

Yeadon, P.J., and Catcheside, D.E.A. (1995). *Guest*: A 98 bp inverted repeat transposable element in *Neurospora crassa*. Mol Gen Genet **247,** 105-109.

Young, C., Itoh, Y., Johnson, R., Garthwaite, I., Miles Christopher, O., Munday-Finch Sarah, C., and Scott, B. (1998). Paxilline-negative mutants of *Penicillium paxilli* generated by heterologous and homologous plasmid integration. Curr Genet **33,** 368-377.

Young, C.A. (2005). The indole diterpene gene cluster from the ryegrass endophyte, Neotyphodium lolii, is required for the biosynthesis of lolitrem B, a bioprotectve alkaloid. PhD thesis, Institute of Molecular BioSciences (Palmerston North: Massey University).

Young, C.A., McMillan, L., Telfer, E., and Scott, B. (2001). Molecular cloning and genetic analysis of an indole-diterpene gene cluster from *Penicillium paxilli*. Mol Microbiol **39,** 754-764.

Young, C.A., Bryant, M.K., Christensen, M.J., Tapper, B.A., Bryan, G.T., and Scott, B. (2005). Molecular cloning and genetic analysis of a symbiosis-expressed gene cluster for lolitrem biosynthesis from a mutualistic endophyte of perennial ryegrass. Mol Genet Genomics **274,** 13-29.

Young, C.A., Felitti, S., Shields, K., Spangenberg, G., Johnson, R.D., Bryan, G.T., Saikia, S., and Scott, B. (2006). A complex gene cluster for indole-diterpene biosynthesis in the grass endophyte *Neotyphodium lolii*. Fungal Genet Biol **43,** 679-693.

Yu, J., Chang, P.K., Cary, J.W., Wright, M., Bhatnagar, D., Cleveland, T.E., Payne, G.A., and Linz, J.E. (1995). Comparative mapping of aflatoxin pathway gene clusters in *Aspergillus parasiticus* and *Aspergillus flavus*. Appl Environ Microbiol **61,** 2365-2371.

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

Yu, J.H., Butchko, R.A., Fernandes, M., Keller, N.P., Leonard, T.J., and Adams, T.H. (1996). Conservation of structure and function of the aflatoxin regulatory gene *aflR* from *Aspergillus nidulans* and *A. flavus*. Curr Genet **29,** 549-555.

Zamocky, M., and Koller, F. (1999). Understanding the structure and function of catalases: Clues from molecular evolution and in vitro mutagenesis. Progress in Biophys Mol Biol **72,** 19-66.

Zdobnov, E.M., and Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17,** 847-848.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res **31,** 3406-3415.

Appendices

Appendix A – Sequence Data

The sequence of *N. lolii* Lp19 *EAS* cluster 1 and the 5′ end of cluster 3 (*lpsA*) is appended in FASTA and Vector NTI formats on a CD attached to the back cover of this thesis. *EAS* cluster 1 sequence is also deposited at the DDBJ/EMBL/GenBank databases (accession no. EF125025).

Appendix B – Intron Splice Site Analysis

Table A1. Intron splice and branch sites for *N. lolii EAS* cluster 1 genes

Intron	size (bp)	5' Donor SS	3' acceptor SS	branch site
lpsB-1	75	AA|GUUAGU	AUUAG	GCUAAC
lpsB-2A*	63	GA|GUAUGG	UGUAG	CCUAAC
lpsB-2B*	83	CG|GUGAGU	UGUAG	CCUAAC
easE-1	75	AG|GUAAGU	GAUAG	GCUGAC
easE-2	60	AU|GUAUGA	UUUAG	GCUUAC
easF-1	69	GC|GUGAGA	UUUAG	GCUAAU
easG-1	78	GG|GUAUGA	UUUAG	GCUAAC
easG-2	53	AG|GUAUGA	GAUAG	AUUAAC
consensus[a]		NG|GURAGU	UNYAG	RCURAY

5' GU, 3' AG and predicted branch point of introns are underlined, vertical line indicates 5' exon/intron junction

*alternatively spliced intron; [a]fungal consensus sequences from (Kupfer *et al.*, 2004); SS = splice site

Appendix C – LC-MS/MS Data

Microsoft Excel files containing LC-MS/MS data from *lpsB* targeted deletion (Section 3.3.2) and comparative analysis (Section 6.2) experiments are appended on a CD attached to the back cover of this thesis.

Appendix D – Black Beetle Feeding Data

Table A2. Black beetle feeding of perennial ryegrass infected with *E. festucae* Fl1 wild type, DFM3 Δ*lpsB* mutant or endophyte-free

| Cage	Treatment	total tillers	undamaged	minimal damage	moderate damage	severe damage	%tillers damaged
2	EGH5013 (E-)	15	9	0	2	4	40
4	EGH5014 (E-)	24	17	0	2	5	29.16667
5	EGH5015 (E-)	15	7	0	1	7	53.33333
6	EGH5016 (E-)	12	0	0	0	12	100
7	EGH5017 (E-)	19	0	0	1	7	42.10526
2	EGH4990 (Fl1)	17	16	1	0	0	5.882353
4	EGH4991 (Fl1)	13	10	3	0	0	23.07692
5	EGH4992 (Fl1)	10	10	0	0	0	0
6	EGH4993 (Fl1)	22	18	0	0	4	18.18182
7	EGH4994 (Fl1)	7	7	0	0	0	0
2	EGH4982 (Δ*lpsB*)	11	11	0	0	0	0
4	EGH4983 (Δ*lpsB*)	16	16	0	0	0	0
5	EGH4984 (Δ*lpsB*)	16	15	1	0	0	6.25
6	EGH4985 (Δ*lpsB*)	12	5	0	0	7	58.33333
7	EGH4986 (Δ*lpsB*)	15	10	2	3	0	33.33333

Appendix E – MITE Alignments
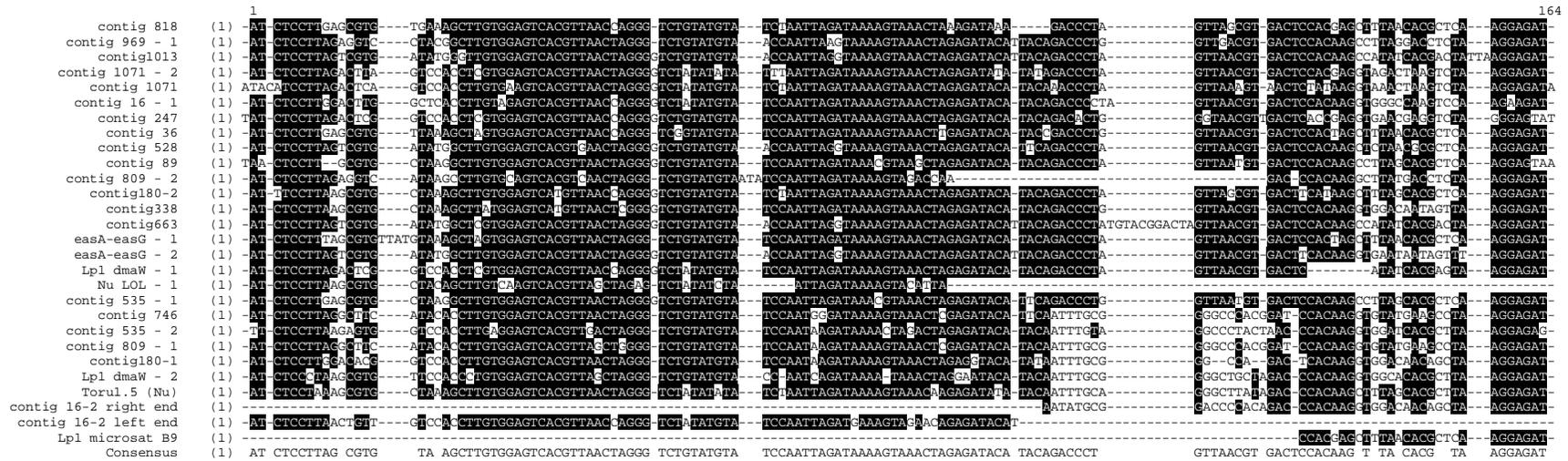
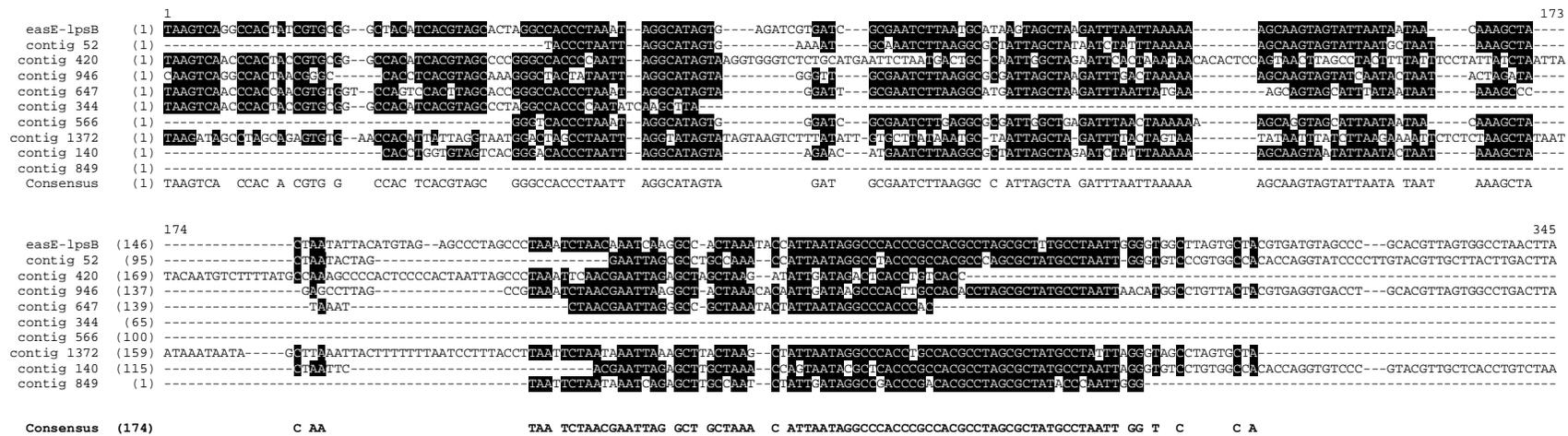**Figure A1**. Multiple sequence alignment of Toru-1 elements

Figure A2. Multiple sequence alignment of Rima elements

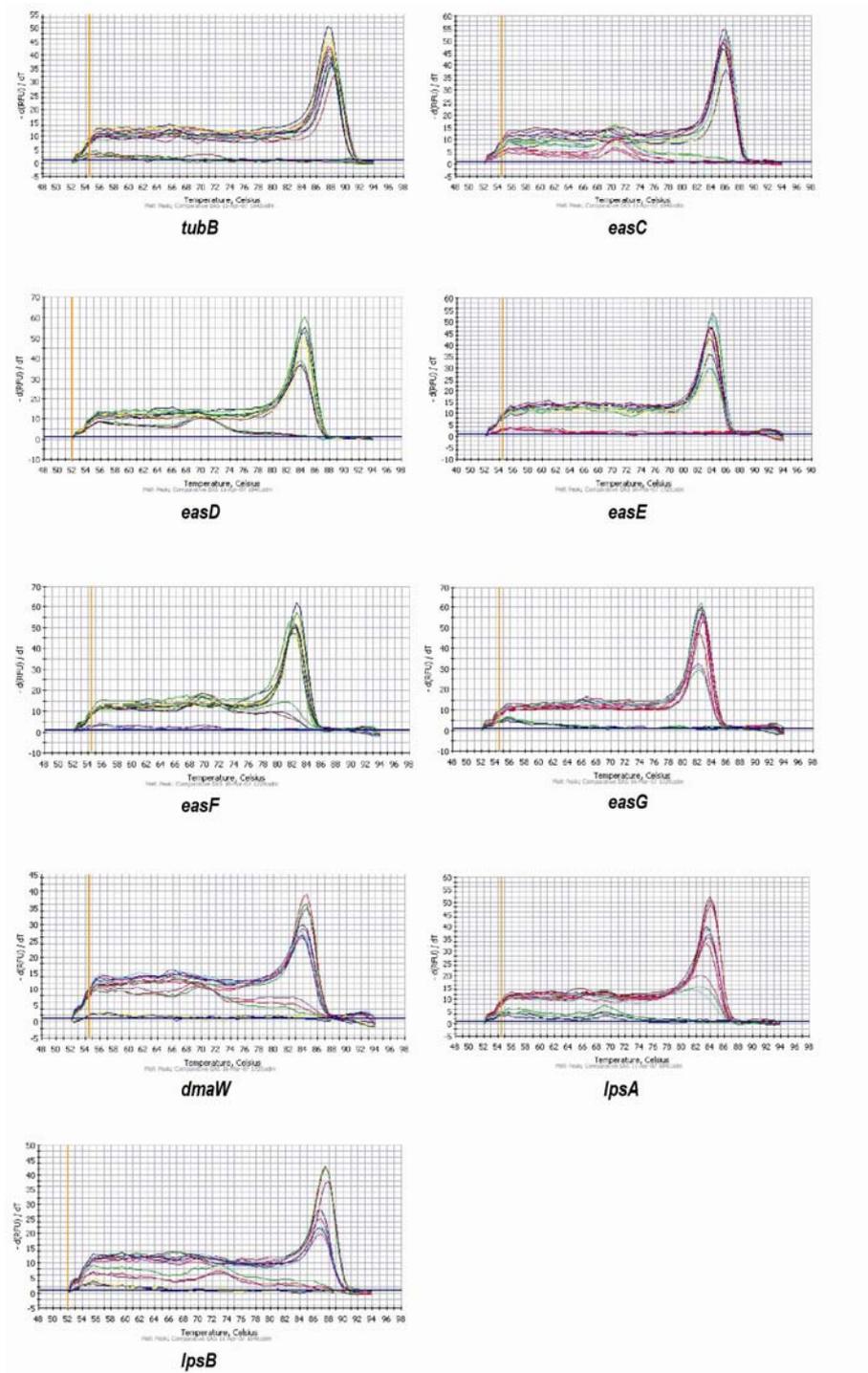# Appendix F – Real Time PCR Melt Curve Analysis

Figure A3. Melt curve analysis of real time PCR described in Section 6.2

Appendix G – Publication

A publication arising from this thesis, published in April 2007 in *Applied and Environmental Microbiology*, is appended overleaf.