CLONING AND CHARACTERISATION OF TWO SUBTILISIN-LIKE PROTEASE GENES FROM NEOTYPHODIUM LOLII

A Thesis presented in partial fulfilment of the
requirements for the degree of
Master of Science in Molecular Genetics
at Massey University, Palmerston North,
New Zealand

Michelle Kay McGill

2000

ABSTRACT

PCR amplification of *Neotyphodium lolii* genomic DNA with degenerate primers detected two different sequences with homology to subtilisin-like proteases. These two PCR products were used to screen a *N. lolii* Lp19 genomic library.

The *prt1* gene was isolated by screening the genomic library with the GH30 PCR product. This gene encodes a putative peptide of 434 amino acids that is most similar to subtilisin-like proteases from *Aspergillus* sp. The *prt1* gene contained a single intron, which was in a position conserved with other fungal genes. 3'RACE was used to determine the polyadenylation site for the *prt1* gene.

Repetitive DNA was a feature of both the 3' untranslated region (UTR) and sequences downstream of the *prt1* gene. Within the 3' UTR, a complex microsatellite was found extending over 50 base pairs. Downstream of the gene, a minisatellite locus of 360 base pairs in size was found, consisting of 40 copies of a 9 base pair AT-rich repeat.

Expression of *prt1* was examined in cultures with various types of carbon and nitrogen sources. Although no conclusive results could be drawn, the type of carbon and nitrogen available did have some effect on *prt1* expression. Repression of *prt1* expression was only observed in media supplemented with sucrose and glutamate.

A 500 bp fragment from the *prt1* promoter was introduced into the vector pFunGus to create a translational fusion with gusA. This vector, pMM9, was transformed into *Penicillium paxilli*. Although transformation frequencies were low, the transformants obtained appeared to be stable for hygromycin resistance. Expression of GUS was observed in seven out of twelve of the stable transformants. This showed that the promoter fragment in pMM9 was sufficient for expression of GUS in a heterologous system.

The *prt2* gene was isolated by screening a genomic library with the GH3 PCR product. Partial sequence has been obtained for the *prt2* gene. The *prt2* gene contains at least three introns, the first of which is conserved with *prt1*. From the sequence obtained, *prt2* encodes a peptide with strong similarity to subtilisin-like proteases from *Metarhizium anisopliae*, a fungal pathogen of insects.

ACKNOWLEDGEMENTS

Firstly, I would like to express my thanks to my supervisor, Professor Barry Scott, for his excellent supervision during my project. I have deeply appreciated your support and encouragement throughout my thesis. Thanks to Dr. Grant Hotter for his initial interest and work on this project, especially for the primer design and probe preparation. To Mike Christensen, thank you for our discussions on the endophytes and their interactions with their hosts. Your enthusiasm and knowledge about endophyte associations has been an inspiration. Thanks also to Dr. Andrew Griffiths for the endophyte genomic DNA used for Southern blotting in Figures 3.2 and 4.2, and to Dr. Peter Farley for the conversations on proteases and their regulation. Thanks also to Dr. Gretchen McCaffrey for her supervision while Barry was away on sabbatical and her encouragement throughout my degree.

To Carolyn Young, how do people in other labs cope without you? I'd probably still be library screening if it wasn't for your help. Thanks for all your help and advice for the practical work, and especially for proof-reading most of this thesis and helping me to put it all together. To Lisa McMillan, thanks for your help with the proofreading, for answering all my questions and for help learning all those new experimental procedures. To my fellow lab-members past and present, Christina, Austen, Emily, Xiuwen, Renae, Raj, Rohan, Mike and Shuguang, and the other members of MGU (Bek, Beccy, Seth, Janet, Paula and Jonathan) thanks for your help and support. To Bec, thankyou for your encouragement during the writing of this thesis, especially when I felt it would never be finished.

To my parents Pat and Teresa, I could never, ever have achieved this without both of you. Thankyou so much for all your love and support, both financial and emotional, especially during the times I would have given up without your help. To my sister Roslyn, thanks for always being there for me, and thanks also to my brothers Steven and Desmond. Special thanks must go to my flatmates for their encouragement while I have been writing. Todd, Andrew, Larnie and Shan, thanks for being great friends. You made sure life at the flat was fun and that I didn't stress out too much. Thanks for the many laughs and songs, although I'm not sure if I want to remember all of the lyrics! Thanks also to Htar for her friendship, support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 FUNGAL ENDOPHYTES OF GRASSES

Fungal endophytes often grow in plant intercellular spaces. Endophytic fungi are thought to have evolved from pathogenic fungi to give interactions that benefit the host plant and the fungal endophyte (Carroll, 1988). Endophytes of grasses are usually Ascomycetes of the family Clavicipitaceae, tribe Balansiae (Clay, 1988). Clavicipitaceous fungi are pathogens of grasses, sedges, other ascomycete fungi and insects (Glenn et al., 1996). Entomopathogenic and mycoparasitic fungi are placed in the Oomycetoideae and Cordycipitoideae subfamilies. The Clavacipitoideae subfamily contains plant pathogenic fungi, including Tribe Balansiae, which contains fungal pathogens of grasses.

Balansiae fungi are either epibiotic or endophytic in growth. Epibiotic species form reproductive stromata on leaves or inflorescences, while endophytic species form systemic infections within host grasses as well as external reproductive stromata (Clay, 1988). Balansiae species live mainly as non-pathogenic endophytes within the plant. These fungi are widespread within the plant tissue, and infection persists throughout the lifetime of the host. The fungal hyphae grow between cells in root and stem tissue without damaging or penetrating host cells.

Tribe Balansiae consists of five genera: *Balansia*, *Balansiopsis*, *Myriogenospora*, *Echinodothis* and *Epichloë*. *Epichloë* species infect cool season grasses such as the Festucoideae (Siegel and Schardl, 1991). Infections of grasses with *Epichloë* species are largely asymptomatic, but during the sexual stage of the fungal life cycle reproductive stromata prevent the host grass flowering (Figure 1.1). *Epichloë* species are heterothallic, with spores transferred to hyphae of individuals of the opposite mating type by female flies (Bultman et al., 1998).

The fungal *Neotyphodium* genus infects many of the same grass species as *Epichloë* (Clay, 1988). *Neotyphodium* species grow intercellularly in apical meristems and aerial tissue with no invasion or destruction of the host cells. Evidence for a close relationship between *Neotyphodium* species and Balansiae fungi comes from similarities in morphology and secondary metabolites, as well as similar host ranges. Analysis of nuclear DNA sequences such rRNA gene ITS regions (Schardl et al., 1991) and β-tubulin genes (Tsai et al., 1994), isozyme characterisation (Leuchtmann and Clay, 1990) and serology studies all support a

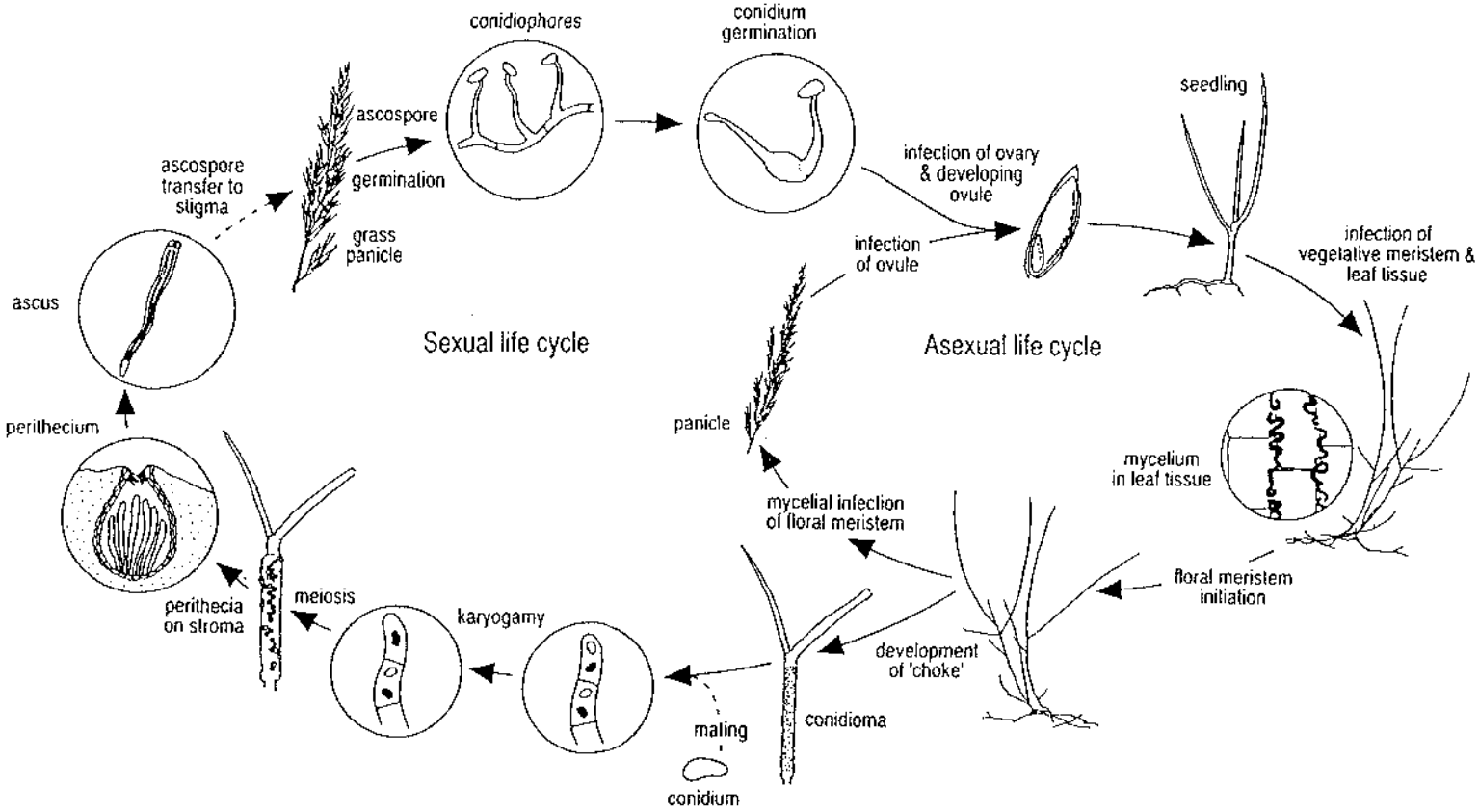
Figure 1.1 Life cycles of sexual and asexual grass endophytes

The life cycles represented show the sexual and asexual stages of the life cycles of grass endophytes. The sexual life cycle, which is represented by *E. festucae*, initiates when mycelia emerge from leaf intercellular spaces. Conidioma develop on the leaf surface, causing choke disease in immature florescences. This prevents the host from flowering, and renders the grass tiller infertile. Epichloe species are heterothallic, which means mating can only take place between conidia of two different mating types. Female flies (*Phorbia* sp.) mediate transfer of conidia between infected plants (Bultman et al., 1998). After mating has taken place, sexual stromata are formed. The stromata contain perithecia, which bear ascospores. The ascospores are transferred to the stigma of a grass flower, where they germinate. Conidiophores are formed, releasing conidia, which germinate and infect the ovary and developing ovule of the grass flower.

During the asexual phase of the endophyte life cycle, growth within the grass is asymptomatic. The endophyte is initially found within the apical meristem of vegetative plants. After the floral meristem develops, mycelia invade the ovaries and ovules, before eventual incorporation into the seed. If seeds from the infected plant germinate, the seedlings will almost certainly be infected with the some endophyte as the maternal parent plant. The asexual phase of the life cycle is also referred to as vertical transmission. Vertical transmission is a highly efficient means of endophyte transmission, resulting in nearly all the seeds of a host plant infected.

Epichloë species are able to use both the asexual and sexual stages of the endophyte life cycle. *Neotyphodium* species are no longer able to produce ascospores, so they are unable to complete the sexual phase of the life cycle. This means that Epichloe species are able to maintain their genetic diversity through the sexual cycle, but Neotyphodium species are effectively clones of their parental strain, leading to a lack in genetic diversity.

Dashed lines indicate spore transfer. Diagram prepared by Liz Grant, Department of Ecology, Massey University, New Zealand.


conidiophores
conidium germination
ascospore
germination
grass panicle
ascospore transfer to stigma
ascus
perithecium
perithecia on stroma
meiosis
karyogamy
mating
conidium
conidioma
development of 'choke'
mycelial infection of floral meristem
panicle
infection of ovule
infection of ovary & developing ovule
seedling
infection of vegetative meristem & leaf tissue
mycelium in leaf tissue
floral meristem initiation
Sexual life cycle
Asexual life cycle

close evolutionary relationship between *Epichloë* and *Neotyphodium* species. Unlike their Balansiae relatives that have a sexual cycle, *Neotyphodium* species are incapable of forming fruiting bodies such as sexual ascospores or asexual conidia within the host plant (Clay, 1988). The loss of the sexual cycle from *Neotyphodium* has had important implications in endophyte evolution.

1.1.1 Taxonomy and evolution of *Neotyphodium* endophytes

The genus *Neotyphodium* was previously known as *Acremonium*, section Albo-lanosa (Morgan-Jones and Gams, 1982). The *Acremonium* genus had a high degree of heterogeneity, containing fungi related to three different groups of Ascomycetes: the Hypocreaceae, the Clavicipitaceae and the Sordariales. Phylogenetic studies using 18S ribosomal DNA sequences have revealed the grass endophytes of section Albo-lanosa are more closely related to the clavicipitaceous fungi than most *Acremonium* species. Fungal species previously in *Acremonium* section Albo-lanosa were consequently reclassified in the *Neotyphodium* genus (Glenn et al., 1996).

Although some *Neotyphodium* species appear to have arisen directly from the loss of the sexual cycle in *Epichloë* species, as in *N. lolii* from *E. festucae*, most have a more complex evolutionary history. Most *Neotyphodium* species are interspecific hybrids of two or more endophyte species (Tsai et al., 1994). *Lolium perenne* taxonomic group 2 (LpTG-2), for example, is thought to have arisen by interspecific hybridisation between *E. typhina* and *N. lolii* (Schardl et al., 1994). Studies show that the LpTG-2 *tub*2-1 and rRNA genes are very similar to corresponding genes from *Epichloë typhina*, while the LpTG-2 *tub*2-2 gene was identical to the *N. lolii tub*2 gene (Tsai et al., 1994). The proposed origins of asexual endophyte groups are shown in Table 1.1.

Interspecific hybridisation may be important in maintaining genetic diversity by providing new genetic material which lacks harmful mutations that usually build up in an asexual organism (Schardl, 1996). Hybridisations involve the formation of heterokaryotic mycelium, which can form when mycelia from two different fungi fuse together (Chung and Schardl, 1997). For endophytes to undergo hybridisation, a single host plant must be infected by two different endophytes at the same time. Although multistrain infections are rare, they have been observed in some grass species (Meijer and Leuchtmann, 1999).

Table 1.1 Origin of asexual endophyte taxonomic groups

Endophyte group	Host grass	Likely ancestors
N. coenophialum (FaTG-1)	*Festuca arundinacea*	*E. baconii*[1], *E. festucae*[1], *N. uncinatum*
FaTG-2	*Festuca arundinacea*	*E. baconii, E. festucae*
FaTG-3	*Festuca arundinacea*	*E. baconii, E. typhina*
N. lolii (LpTG-1)	*Lolium perenne*	*E. festucae*
LpTG-2	*Lolium perenne*	*E. typhina, N. lolii*
N. uncinatum	*Festuca pratensis*	ancient hybrid origin[2]

This table is based on Table 2 from Schardl, 1996

[1] The *E. baconii* and *E. festucae* ancestry of *N. coenophialum* probably came from FaTG-2 endophytes.

[2] Schardl, 1997

1.2 THE MUTUALISTIC SYMBIOSIS BETWEEN GRASS AND ENDOPHYTE

1.2.1 Effects on the plant

The interactions between the host plant and the *Neotyphodium* endophyte have beneficial and detrimental effects for grasses (Clay, 1991). Endophyte infection costs the plant through loss of photosynthates and mineral nutrients, and can lead to decreased growth of the grass under stressful conditions (Cheplick et al., 1989). Changes in morphology due to endophyte infection could affect the competitive ability of the host. However, the plant also gains a number of benefits from endophyte infection. Secondary metabolites provide protection to the grass from herbivores and pathogens. Plant growth is often increased in endophyte-infected plants under normal conditions.

1.2.1.1 Resistance to herbivores and plant pathogens

Neotyphodium endophytes produce a wide range of biologically active secondary metabolites, which confer resistance to herbivores, nematodes and pathogenic fungi. The metabolites produced include ergot alkaloids, indole-diterpenoids, lolines (pyrrolizidine alkaloids) and pyrrolpyrazine alkaloids, along with sterols and steroid metabolites and other compounds. Although most of the secondary metabolites from the symbiosis are of fungal origin, some may be produced by the host grass in response to endophyte infection.

Ergot alkaloids, which are formed from tryptophan and mevalonate, are highly toxic to mammalian herbivores. Consumption of grasses containing these compounds leads to reductions in mass gain, milk production, feeding and reproduction (Strickland et al., 1993). Ergot alkaloids have a high affinity for amine receptors such as dopamine, epinephrine and serotonin receptors, which may be related to toxicosis in the animal. Ergot alkaloids also have insecticidal activity against some insect species.

Lolitrems are indole-diterpenoids produced by *Neotyphodium lolii* and other endophyte species. Under certain environmental conditions consumption of lolitrems, which are tremorgenic neurotoxins, leads to ryegrass staggers. This neuromuscular disease, which affects sheep, cattle, deer and horses, leads to a loss of co-ordination due to tetanic muscle

spasms (Siegel and Bush, 1997) and can eventually cause death. Lolitrems also act as a feeding deterrent to Argentine stem weevil larvae.

The primary deterrent to insect feeding is a pyrrolpyrazine alkaloid called peramine. This compound was particularly effective against the Argentine stem weevil, which is a major pest in grass pastures in New Zealand (Rowan, 1993). Peramine is not known to have a significant effect on mammalian or plant species.

Lolines have broad range insecticidal activity, and may act as metabolic toxins or feeding deterrents depending on the insect species (Bush et al., 1997). Lolines are also active against aphids, which act as a vector for barley yellow dwarf virus (Siegel et al., 1990). Feeding studies with lolines have shown a range of responses in vertebrate herbivores (Strickland et al., 1993), but no definite conclusions could be made on their effect. Lolines may also contribute to allelopathic effects, which are detrimental to the growth of other plants. Experiments by Bush et al., 1993 showed that lolines inhibit the rate and amount of germination for several monocotyledonous and dicotyledonous species. Lolines are produced by endophytes when grown under certain culture conditions or in cultures containing autoclaved leaf material (Franks et al., 1997; Wilkinson et al., 1997). Mendelian genetic analysis has shown that a single genetic locus controlled loline expression in *E. festucae* (Wilkinson et al., in press).

Endophyte-infected grasses show enhanced resistance to nematodes and other fungal species (Kimmons et al., 1990; Siegel and Latch, 1991). In culture, *Neotyphodium* and *Epichloë* endophytes inhibit the growth of several species of phytopathogenic fungi (Siegel and Latch, 1991). Field studies show that endophyte-infected grasses were less affected by the fungal phytopathogens *Alternaria triticina and Cladosporium phlaei*, which cause leaf spots (Siegel and Bush, 1997), and *Puccina coronata*, which causes crown rust. Endophyte infected plants were more resistant to seedling disease caused by *Rhizoctonia zeae* (Gwinn and Gavin, 1992).

Sesquiterpenes, phenolic glycerides, fatty acids and ring aromatic sterols have also been found in endophyte infected plants (Siegel and Bush, 1997). While all of these compounds have antifungal activity, it is possible that some of these metabolites are phytoalexins, which are produced by plants as a defence against potential pathogens. Resveratrol is a phytoalexin and stress metabolite found at higher concentrations in endophyte-infected ryegrasses

(Powell et al., 1994). Resveratrol has antifungal activity, but it is also toxic to animals such as insects, fish and mice.

1.2.1.2 Effects of endophyte infection on plant growth and development

Endophyte infection is generally advantageous to the growth and development of its host plant. Infection leads to increased seed germination and seedling establishment, as well as increased tillering and dry matter yield, enhanced persistence and drought resistance (van Heeswijck and McDonald, 1992). The symbiosis between the plant and its endophyte leads to significant increases in leaf area, tiller numbers and leaf growth under controlled conditions.

In growth chamber studies, grass genotypes varied in their responses to endophyte infection. Final leaf length, tiller number, pseudostem mass and root mass varied among infected genotypes (West, 1994). Stem, leaf and root mass was increased in endophyte infected plants even in the absence of herbivores. This infers that protection from herbivores by secondary metabolites is not solely responsible for increased growth in infected grasses. Improved plant growth was generally observed at high nutrient levels. At low nutrient levels, infected grass seedlings showed inhibited growth, while growth of infected adult tall fescue plants was enhanced (Cheplick et al., 1989).

The increase in growth seen in endophyte-infected plants may be due to the activity of a phytohormone. Some endophyte species produce 3-indole acetic acid (IAA) in culture (De Battista et al., 1990). IAA is a naturally occurring auxin, which regulates cell growth, differentiation and apical dominance. This hormone could be capable of promoting the increased growth seen in endophyte-infected grasses. However, no difference in the levels of unbound IAA has been found between infected and uninfected tall fescue (Siegel and Schardl, 1991). IAA may be kept in several storage forms, but only bound IAA is biologically active. This could mean that the endophyte contributes to the overall levels of auxin in the plant.

1.2.1.3 Improved drought tolerance

Endophyte infected grasses show increased drought tolerance compared to uninfected plants. In experiments with endophyte-free and endophyte-infected tall fescue, infected grasses

were much more tolerant of water stress (Arachevaleta et al., 1989). Endophyte-infected plants exhibited increased leaf rolling, which prevents water loss, and greater regrowth following alleviation of drought conditions. Increased survival of endophyte-infected grasses during severe drought may be due to the endophyte affecting water relations in its host. The endophyte increases osmotic adjustment in infected plants, which results in accumulation of solutes in plant cells (Siegel and Schardl, 1991; West, 1994). Mannitol, which is an important solute for osmotic adjustment, accumulates in endophyte infected plants under water stress, but not in uninfected plants (Richardson et al., 1992). The endophyte could produce mannitol itself, or send some signal to the plant resulting in mannitol production.

Reports on the effect of the endophyte on stomatal conductance and photosynthesis are inconclusive (West, 1994). The endophyte could prevent water loss by early stomatal closure. This could be mediated by the production of the phytohormone abscisic acid or by secretion of anti-transpirant molecules by the fungus into the apoplast. Like many other effects of the endophyte on its host, the effects are largely dependent on the genotype of both the endophyte and its host.

1.2.2 Effects on the endophyte

The fungus also gains a number of benefits from the symbiosis between itself and its host. The plant provides nutrition in the form of photosynthates and inorganic ions. The fungus is also protected from biotic and abiotic stresses (Clay, 1991). As asexual endophytes, *Neotyphodium* species are also dependent on seed dissemination.

1.2.2.1 Seed dissemination of endophytes

Neotyphodium endophytes do not have a sexual stage to their life cycle, which means they must have some other mechanism of perpetuating themselves. *Neotyphodium* endophytes manage to propagate themselves by colonising seeds, or by growing within their host plant. Endophyte hyphae grow in aleurone layers of the seed, surrounding the endosperm and growing between the endosperm and scutellum (Philipson and Christey, 1986). Fungal hyphae grow between the host cell walls and the plasma membrane, usually without penetrating the embryo. This method of seed dissemination employed by *Neotyphodium* endophytes is called vertical transmission.

Vertical transmission of the endophyte within the seed is of advantage to both the endophyte and its host. The endophyte gains a highly efficient means of transmission, with nearly all the seeds of a host plant infected (Schardl, 1996). This means that the endophyte is inherited maternally in a manner reminiscent of mitochondrial inheritance. The fungus is dependent on the host for its survival, so it may explain selection for protective compounds that appear to be of no direct benefit to the endophyte. These protective compounds prevent seed predation by insects and other pathogens. Infected seeds show increased levels of germination compared with uninfected seeds (Siegel and Schardl, 1991).

Vertical transmission reflects the fact that the endophyte is unable to produce ascospores and conidia in the host (Clay, 1988). When the endophyte is disseminated by vertical transmission, fungal stromata that render grass inflorescences infertile are not present. This means the infected host plant is fertile, as well as showing enhanced vigour and resistance to mammalian, insect and microbial pests. However, loss of the sexual cycle has led to less genetic variability in endophyte populations (Bucheli and Leuchtmann, 1996).

The systemic growth of the endophyte within the plant and the spread of the endophyte into seeds decreases the probability of a mutation persisting that is not of some benefit to the endophyte or its host. Interspecific hybridisation (Section 1.1.1) is another important proposed mechanism for preventing an accumulation of harmful mutations in asexual endophytes.

1.2.2.2 Fungal nutrition and effects on host metabolism

As stated in Section 1.1, the endophyte grows within the intercellular spaces of its host without penetrating the cell wall or the cytoplasm of host cells. This means the endophyte must obtain all its nutrients from the plant apoplast (Lam et al., 1994) which is poor in nutrients. The endophyte would require amino acids, simple sugars, vitamins and vitamin precursors for growth (Bacon and White, 2000).

Studies of *Neotyphodium coenophialum* have shown this endophyte can utilise a wide range of carbon and nitrogen sources. This endophyte utilised fructose, glucose, mannose, sucrose, trehalose, raffinose, sorbitol and mannitol as carbon sources, but not galactose, sorbose, rhamnose, pectin or cellulose (Kulkarni and Nielsen, 1986). Trehalose and mannitol were excellent sources for supporting fungal growth. Kulkarni and Nielsen also discovered that

the endophyte could utilise ammonium, arginine, asparagine, cysteine, glutamine, proline and serine as nitrogen sources. Nitrate, alanine, glycine, histidine, leucine, lysine, methionine, phenylalanine, tryptophan and urea were not utilised as nitrogen sources. The endophyte also required the vitamin thiamine for growth.

Endophytes are mainly found in sink tissues within the plant, such as leaf sheaths and piths. Sinks are regions that import more carbohydrates than they export (Lam et al., 1994). As sinks are regions where phloem unloading takes place, the most common carbohydrates available to endophytes would be sucrose and its constituent sugars glucose and fructose. There is evidence for two methods of sucrose uptake in endophytes. The first involves a sucrose transporter, while the second involves the break down of sucrose into glucose and fructose by a cell wall invertase, before uptake by separate glucose and fructose transporters.

In order to maintain the source-sink relationship, endophytes probably convert glucose and fructose into fungal sugar alcohols and trehalose (Bacon and White, 2000). This maintains a concentration gradient between the source and the sink. Sugar alcohols are more reduced, and thus a more efficient form of energy storage. These metabolites also lower water potential in the endophyte, which favours diffusion of water from the grass into the fungus.

Endophyte infection also has a significant effect on nitrogen metabolism within the plant (Bacon and White, 2000). When infected and non-infected grasses were examined for $^{14}CO_2$ assimilation, there was an 85% increase in incorporation into amino acids in the leaf blades of endophyte infected grasses. However, no differences were observed in the leaf sheath. The activity of glutamine synthetase was significantly higher in the leaf blade of infected plants. Glutamine synthetase is a crucial enzyme in plant nitrogen metabolism. The concentration of NH_4^+ in the leaf sheath was doubled in endophyte-infected plants, but was unaffected in the leaf blade.

1.2.2.3 Protection from the external environment

The interior of the plant provides a protective environment for mutualistic fungi such as *Neotyphodium*. It is provided with buffers against changes in water and nutrient availability, with most of its nutrients provided by its host. The fungus is also protected from biotic factors such as mycoparasites and abiotic stresses such as osmotic stress and nutrient deprivation.

1.3 SPECIFICITY OF ENDOPHYTE-HOST INTERACTIONS

1.3.1 Host specificity of endophytes

Host range reports indicate that particular endophytes are restricted to certain host species or genera. There is host specific genetic differentiation in geographically distant endophyte populations, which may involve the reproductive isolation of seed born endophytes (Leuchtmann and Clay, 1997). The host specificity of endophytes also depends on the genotypes of both the host and the endophyte. Endophyte and host genotypes affect traits such as water stress tolerance and secondary metabolite production by the fungus *in planta*. The genetic interactions that take place between the endophyte and its host are poorly understood.

1.3.2 Host resistance to endophyte infection

Host grasses are resistant to infection by the majority of potential endophytes. This is part of a non-specific resistance from the plant towards potential pathogens. Resistance to endophyte infection may involve induced and constitutive defence mechanisms (Leuchtmann and Clay, 1997). These mechanisms are non-specific towards pathogens (Hotter, 1996). Physical barriers and unsuitable nutritional compounds within the host may play a role in constitutive defences. Plants also produce a large array of antimicrobial proteins and chemicals that could provide resistance to endophyte infection (Broekaert et al., 1997).

Although the interaction between an endophyte and its natural host is compatible, it has been observed that endophyte infection increases expression of antifungal proteins such as chitinase (Roberts et al., 1992). The cell wall of *Neotyphodium lolii* contains very little chitin (J. Schmid, personal communication) so it is possible chitinases may not have much effect on the cell walls of some endophytes. Infected plants may also produce phytoalexins that are toxic to some fungal species (Powell et al., 1994).

Natural host-endophyte interactions show no physical symptoms of a defence response by the plant (Koga et al., 1993). To survive within its host, the endophyte must have developed mechanisms of overcoming host defence reactions. Degradation of antifungal proteins or the breakdown of antifungal chemicals could achieve this. Pathogenic fungi use enzymes or

toxins to enable them to circumvent the usual non-specific resistance of plants (Leuchtmann and Clay, 1997).

When endophytes are inoculated into a new host, most successful introductions occur when isolates are transferred between closely related grass species (Koga et al., 1993; Leuchtmann, 1994). This evidence supports the hypothesis that there is specificity in the interaction between the host grass and its endophyte. The failure of the endophyte to infect non-host grasses is probably due to the non-specific resistance plants have to most non-pathogenic fungi.

Two types of incompatibility are commonly seen in artificial host-endophyte interactions; incompatibility can affect either the endophyte or the host cells. In perennial ryegrass (*Lolium perenne*) infected with *N. coenophialum*, some isolates of fungi used to inoculate this grass appeared dead, with distorted hyphae lacking cytoplasm (Koga et al., 1993). The hyphae resembled hyphae from plant pathogenic fungi that have penetrated resistant hosts. The hyphae were distorted and collapsed, with degenerated cytoplasm and electron dense cell walls. The hyphae also lacked a fibrous outer layer, which is considered characteristic of the interaction between *Neotyphodium* endophytes and their hosts. Although no hypersensitive reactions were observed in host cells surrounding the hyphae, the extracellular matrix found between these fungi had become electron dense. This may have reduced permeability and nutrient availability to the fungus, leading to incompatibility between the host and endophyte.

The second type of incompatibility reaction affects host cells. In an artificial association between perennial ryegrass with *N. uncinatum* (an endophyte of meadow fescue) or tall fescue inoculated with LpTG-2 (an endophyte of perennial ryegrass), stunted tillers with necrosis in the apical meristem were observed (Christensen, 1995). Some tillers were also free of the endophyte, indicating the associations formed were incompatible. The necrosis observed in the apical meristem was similar to the effects of the hypersensitive response, in which localised cell death to prevent a pathogen spreading within a plant (Jackson and Taylor, 1996).

1.3.3 The physical interface between the endophyte and its host

A variety of structures are found at the interface between the endophyte and its host. Crystalline inclusions and tubular complexes are often observed in the association between

N. lolii and perennial ryegrass. The interaction between *E. typhina* and perennial ryegrass features tubular complexes and bilayered cell walls (Siegel et al., 1987). Empty vacuole or vacuoles filled with electron-dense material are often seen in grass-endophyte symbioses. These unusual ultrastructural features have never been found in fungi growing in culture, which infers these elements form only when the fungus is growing within its host. Possible functions of the ultrastructural elements are unclear, but they could play roles in toxin production or in nutrient exchange.

1.4 FUNGAL EXTRACELLULAR PROTEASES

Most Ascomycete fungi and their asexual relatives secrete extracellular proteases. Different classes of proteases are adapted to different environmental conditions, with factors such as pH and nutrient availability affecting protease gene expression. The different classes of secreted proteases include aspartic, cysteine, metallo- and serine proteases. Secreted proteases are produced by pathogens, involved in industrial processes and cause large-scale losses such as post harvest decay of grains caused by *Aspergillus* species.

1.4.1 Protein localisation and post-translational modifications of extracellular enzymes

In eukaryotes, all secreted proteins must pass through the endoplasmic reticulum, before transport onwards to other organelles (Alberts et al., 1994). Secreted proteins are produced as inactive preproproteins, which prevents possible damage to intracellular components. A signal peptide at the amino terminus of the protein targets ribosomes translating these proteins to the endoplasmic reticulum. The signal peptide is largely hydrophobic in nature, and typically consists of 20-30 amino acid residues. During translocation of the protein across the endoplasmic reticulum membrane, the signal peptide is cleaved.

Post-translational modifications, such as glycosylation, take place in the endoplasmic reticulum and Golgi body. These modifications may be necessary for activity or correct folding of proteins. In the Golgi body, proteins are packaged into vesicles before secretion to the cell exterior. After secretion, the propeptide is cleaved by autolysis to give an active enzyme.

1.5 SUBTILISIN-LIKE PROTEASES

Subtilisin-like proteases are a class of serine proteases that have a catalytic triad of aspartic acid, histidine and serine residues in a protein scaffold of α helices and β sheets (Siezen and Leunissen, 1997). These proteases are found in archaebacteria, eubacteria, yeasts, fungi and higher eukaryotes. Most subtilisin-like proteases are initially synthesised as prepropeptides before processing through the eukaryotic secretion system (Section 1.3.1). The proteinase K subfamily of subtilisin-like proteases comprise a group of proteins with high sequence similarity, and are only found in yeasts, fungi and Gram negative bacteria (Siezen and Leunissen, 1997).

1.5.1 Roles of proteases in fungal nutrition and pathogenesis

1.5.1.1 Roles of fungal subtilisin-like proteases in pathogenesis

Pathogenic fungi produce a wide range of extracellular enzymes such as chitinases, cellulases and proteases. These enzymes are critical components of the interaction between the pathogen and its host. Proteases may be involved as virulence factors (Bidochka and Khachatourians, 1994), or in nutrition (Segers et al., 1999). Proteases can also break down physical barriers to infection (Bonants et al., 1995; St Leger et al., 1992), degrade antifungal proteins produced by the host and prevent secondary infection by other organisms (Tunlid et al., 1994).

In entomopathogenic fungi, it is proposed extracellular proteases play a significant role in infection by degrading the insect cuticle. The cuticle is primarily made of chitin and various other proteins. Proteases from the fungus *Metarhizium anisopliae* can degrade the insect cuticle by secreting high concentrations of the enzyme at the site where the fungus penetrates (St Leger et al., 1992). In the migratory grasshopper (*Melanopus sanguinipes*), extracellular proteases from *Beauveria bassisana* or *M. anisopliae* degraded mainly acidic cuticular proteins, and also high molecular weight basic proteins (Bidochka and Khachatourians, 1994). Fungal strains deficient in protease production or treated with protease inhibitors were less virulent, suggesting subtilisin-like proteases produced by these fungi were virulence factors.

Fungal pathogens of nematodes also produce subtilisin like proteases that are involved in pathogenicity. *Paecilomyces lilacinus* produces a subtilisin-like extracellular protease with gene expression induced by colloidal chitin, vitellin and intact eggs from the root-knot nematode *Meloidogyne hapla*, but repressed by glucose (Bonants et al., 1995). This serine protease degrades vitellin, which is a constituent of the nematode eggshell. *Verticillium suchlasporium*, another fungal pathogen of nematodes, produces a similar protease which could degrade some nematode egg proteins (Lopez-Llorca, 1990). This protease is expressed in the fungal appressoria, which are specialised structures involved in penetration of physical barriers. *Arthrobotrys oligospora*, a nematode-trapping fungus, produces related subtilisin like proteases that can degrade proteins from the nematode cuticle (Tunlid et al., 1994).

The role of subtilisin-like proteases in plant pathogenic fungi is not well understood. Experiments by St Leger et al., 1997 showed the plant pathogenic fungi *V. albo-atrum* and *V. dahliae* expressed very low levels of subtilisin-like proteases, but high levels of trypsin-like proteases. However *Magnaporthe poae* (also known as *Magnaporthe grisea*), a pathogen of the roots of Poaceae grasses produces high levels of a subtilisin-like protease. There is evidence for a correlation between increased subtilisin-like protease production and worsening disease symptoms (Sreedhar et al., 1999), which could be due to increased protease expression or an increase of fungal mass within the roots. A subtilisin-like protease from *Fusarium oxysporum* f. sp. lycopersici is expressed during infection of tomato plants (unpublished, GenBank accession AAC27316).

1.5.1.2 Regulation of subtilisin-like proteases

Many fungal subtilisin-like proteases are regulated by nutrient levels, which could indicate a role in nutrition. Expression of these proteases is repressed when there are high levels of utilisable carbon and nitrogen sources. When utilisable carbon and nitrogen levels are low, protease expression is induced. In some cases, these proteases are also induced by macromolecules from host tissues. pH regulation is also critical for subtilisin-like protease expression, as these proteases have optimal activity in alkaline conditions.

Carbon, nitrogen and pH regulation is under the control of global regulatory systems in filamentous fungi. Carbon catabolite repression is controlled by a zinc finger transcription factor called CreA, which was first discovered in *Aspergillus nidulans*. CreA is a negative regulatory protein, which represses expression in the presence of glucose (Felenbok and

Kelly, 1996). The products of the creB, creC and creD genes have also been identified as playing a role in carbon regulation: CreB and CreC through indirect means, and CreD possibly through interaction with CreA.

A zinc finger protein known as AreA (*Aspergillus nidulans*) or Nit2 (*Neurospora crassa*) controls gene expression in response to nitrogen levels. In *A. nidulans* and *N. crassa*, inorganic nitrate is not metabolised until compounds such as ammonia, glutamate and glutamine have been depleted (Premakumar et al., 1979). The AreA protein is a positive acting regulatory protein, which induces expression of genes in response to nitrogen limitation (Fu and Marzluf, 1990). Genes encoding homologous proteins have been found in *Penicillium chrysogenum* (Haas et al., 1995), *Magnaporthe grisea* (Froeliger and Carpenter, 1996), *Gibberella fujikuroi* (Tudzynski et al., 1999) and *Metarhizium anisopliae* (Screen et al., 1998).

A global regulatory protein called PacC controls pH regulation in fungi. PacC, like CreA, Nit2 and AreA, is a zinc-finger transcription factor. The activity of PacC is controlled by proteolytic activation. Under alkaline ambient pH conditions, proteolytic processing of PacC is favoured (Orejas et al., 1995). The processed PacC transcription factor activates the expression of genes whose products function at alkaline pH and represses genes whose products function at acidic pH. A signal transduction cascade involving PalA, B, C, F, H and I is needed for PacC processing *in vivo* (Negrete-Urtasun et al., 1999). Mutations in any of the *pal* genes mimic the effects of acidic ambient pH.

Regulation of subtilisin-like proteases in respect to carbon and nitrogen levels and pH is crucial to ensure available resources are not wasted. pH regulation also ensures that subtilisin-like proteases are only expressed when the ambient pH permits their activity. As subtilisin-like proteases have optimal activity under alkaline pH conditions, the cell would not benefit from expressing these proteases when the ambient pH was acidic. As for many other genes regulated by nutrient levels, there is no benefit to the cell in expressing protease genes if easily utilised nutrients are present at high concentrations in the environment.

1.5.1.3 Roles for subtilisin-like proteases in fungal nutrition

Under conditions where nutrients are plentiful, protease gene expression is usually repressed. In the entomopathogenic fungi *M. anisopliae*, *V. lecanii*, *B. bassiana*, *Tolypocladium niveum*

and *P. farinosus*, expression of subtilisin-like protease is rapidly induced during nutrient starvation (St Leger et al., 1991). PII, a subtilisin-like protease from *A. oligospora* is expressed when the fungus is starved of carbon and nitrogen (Åhman et al., 1996). Expression is stimulated by adding soluble and insoluble proteins to cultures, but repressed in the presence to easily metabolised nitrogen sources or glucose.

In the mycoparasitic fungus *Trichoderma harzianum* subtilisin-like proteases are induced by autoclaved fungal mycelia, fungal cell walls and chitin but repressed in the presence of glucose (Geremia et al., 1993). P32, a protease from the nematode pathogen *V. suchlasporium*, was expressed when gelatine was the only available carbon and nitrogen source (Lopez-Llorca, 1990).

1.5.2 Endophyte subtilisin-like proteases

Like many ascomycete fungi, *Epichloë* endophytes also contained a subtilisin-like protease. Reddy and co-workers (1996) cloned a subtilisin-like protease gene (At1) from *Epichloë typhina*. Previous studies by Lindstrom and co-workers (1993) found At1 formed 1 to 2 % of the total leaf sheath protein in *Poa ampla* Merr. cv infected by *E. typhina*. At1 was also produced by *E. typhina* in culture, which confirmed At1 was of endophyte origin. The high levels of At1 in the host-endophyte interaction suggests it must form a considerable amount of the transcript produced by the fungus, and could play an important role in the interaction between the endophyte and its host.

The At1 protease gene shared significant homology with proteases from pathogenic fungi such as *Paecilomyces lilacinus*, a nematode pathogenic fungus, and *Metarhizium anisopliae*, an entomopathogen (Reddy et al., 1996). In pathogenic fungi, the main roles of subtilisin-like proteases are in host penetration and fungal nutrition.

Reddy and colleagues suggested a number of possible roles for the At1 protease in host-endophyte symbioses. At1 is induced by nutrient deprivation (Lindstrom and Belanger, 1994), so it could degrade proteins to provide peptides and amino acids as carbon and nitrogen sources. At1 could also reduce host structural barriers, allowing increased hyphal perennation through the plant. This seems unlikely, as there is little evidence of cell wall break down in endophyte-host interactions (M. Christensen, personal communication). The third option is that At1 could counteract host defence responses by degrading antifungal proteins produced by the plant.

At1 expression was detected in other endophyte-infected *Poa* species at levels comparable to those seen in the *P. ampla/E. typhina* interaction. In perennial ryegrass/*N. lolii* and *Festuca rubra/E. festucae* associations, copies of a similar gene were present, but levels of detectable At1 mRNA and protein were much reduced. This could be due to lower expression of subtilisin-like proteases by the endophytes in these symbioses, or due to a lower fungal mass within the infected plant (Reddy et al., 1996).

Previous work by G. Hotter (unpublished, 1997) using degenerate primers based on the At1 protease gene cloned by Reddy and co-workers amplified two different PCR products with homology to subtilisin-like proteases from *N. lolii* Lp19 genomic DNA (Appendix 2). One of these products, GH30, showed significant homology to *Aspergillus* subtilisin-like proteases. The other product, GH3, showed homology to subtilisin-like proteases from entomopathogenic fungi, such as *Metarhizium anisopliae*. The discovery of two potential subtilisin-like proteases in the haploid Lp19 endophyte was an unexpected result.

1.6 AIMS OF THIS PROJECT

At present, little is known about the factors required for the interaction between *Neotyphodium* endophytes and their hosts. Work by Lindstrom et al., 1993 showed that a subtilisin-like protease, At1, was expressed at very high levels in the interaction between *E. typhina* and its host, *Poa ampla*. The high levels of expression suggested this protease was important in the interaction between this endophyte and its host.

The aim of this thesis was to clone and characterise the two homologous subtilisin-like proteases from *Neotyphodium lolii* Lp19, which were first detected by G. Hotter (Section 1.5.2). This will allow the expression of both of these genes to be studied both in culture and within the plant. Expression studies will give information on conditions that regulate expression, and may also suggest a possible role for these proteases in the interaction between the endophyte and its host.

CHAPTER TWO

MATERIALS AND METHODS

2.1 FUNGAL AND BACTERIAL STRAINS, λ CLONES AND PLASMIDS

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

2.2 GROWTH OF CULTURES

Neotyphodium cultures were grown at 22°C on PD agar or in PD broth (Section 2.3.3), with shaking at 200 rpm. For cultures from which nucleic acid was to be extracted, 500 µL of PD broth containing homogenised mycelia was used to inoculate 50 ml of PD broth in a 125 mL flask. The cultures were incubated at 22°C with shaking at 200 rpm for 7-14 days. Some cultures were transferred to other media after this initial incubation. Mycelia were washed twice in MilliQ water, then used to inoculate the new media. After a further incubation mycelia were washed twice in MilliQ water and snap frozen in liquid nitrogen.

Penicillium paxilli cultures were grown at 22°C in PD broth (Section 2.3.3) with shaking at 200 rpm, or on PD agar plates (Section 2.3.3) supplemented where necessary with hygromycin. For cultures used for protoplast preparation, approximately 5 x 10^6 spores were used to inoculate 25 mL of PD broth (Section 2.3.3) in each 125 mL flask. The cultures were incubated at 22°C for two days with shaking at 200 rpm.

Escherichia coli cultures were grown at 37°C in LB broth or on LB agar (Section 2.3.2).

2.3 MEDIA

All media were prepared using MilliQ water, and sterilised by autoclaving at 121°C and 15 psi for 15 minutes before use. Liquid media were cooled to room temperature before inoculation or addition of supplements. Solid media were cooled to 50°C before antibiotics were added and plates poured. Uninoculated plates were stored at 4°C.

2.3.1 CD salts medium

CD salts medium contained (g/L): 1.0 g K_2HPO_4, 0.5 g $MgSO_4$, 0.5 g KCl, 0.01 g $FeSO_4$. The pH was adjusted to 6.8 before autoclaving. This media was supplemented with 2 g/L $NaNO_3$, 4.4 g/L sodium glutamate, 30 g/L sucrose, or 30 g/L mannitol where necessary.

Table 2.1 Fungal and bacterial strains, λ clones and plasmids

Strain or Plasmid	Relevant characteristics	Source or Reference
Fungal strains		
Penicillium paxilli	PN2013	(Itoh and Scott, 1994)
Endophyte strains		
AR510 (FaTG-3)	host *Festuca arundinacea*	MFFGC NZ[1]
E8 (*E. typhina*)	host *Lolium perenne*	(Schardl et al., 1991)
Fp5 (*N. uncinatum*)	host *Festuca pratensis*	(Christensen et al., 1993)
Lp1 (LpTG-2)	host *Lolium perenne*	(Christensen et al., 1993)
Lp19 (*N. lolii*)	host *Lolium perenne*	(Christensen et al., 1993)
Tf2 (*N. coenophialum*)	host *Festuca arundinacea*	(Christensen et al., 1993)
Tf13 (FaTG-2)	host *Festuca arundinacea*	(Christensen et al., 1993)
Bacterial strains		
E. coli		
KW251	F⁻ *supE44 galK galT22 metB1 hsdR2mcrB1 mcrA* [*argA81*:: Tn10]*recD1014*	Promega Corp.
XL-1	*supE44 hsdR17 recA1 endA1 gyrA46 thi relA1* Lac⁻/*proAB*⁺ *lacI*^{q} *lacZ* Δ M15 Tn10 (Tet^r)	(Bullock et al., 1987)
PN1670	XL-1 containing pMM1	This study
PN1671	XL-1 containing pMM2	This study
PN1672	XL-1 containing pMM3	This study
PN1673	XL-1 containing pMM4	This study
PN1674	XL-1 containing pMM7	This study

[1] Margot Forde Forage Germplasm Centre, AgResearch, Palmerston North, New Zealand

Strain or Plasmid	Relevant characteristics	Source or Reference
Phage lambda clones		
λMM3.1	λGEM-12 clone from Lp19	This study
λMM3.3	λGEM-12 clone from Lp19	This study
λMM3.5	λGEM-12 clone from Lp19	This study
λMM3.6	λGEM-12 clone from Lp19	This study
λMM30.1	λGEM-12 clone from Lp19	This study
λMM30.2	λGEM-12 clone from Lp19	This study
λMM30.3	λGEM-12 clone from Lp19	This study
λMM30.4	λGEM-12 clone from Lp19	This study
Plasmids		
pAN7-1	6.8 kb Hyg^R Amp^R	(Punt et al., 1987)
pCB1004	Hyg^R	(Carroll et al., 1994)
pFunGus	5 kb, Amp^R	(McGowan, 1996)
pGEM-T Easy	3 kb, Amp^R	Promega Corporation
pGH3	546 bp PCR product in pCR-Script, Amp^R	G. Hotter, unpublished
pGH30	478 bp PCR product in pCR-Script, Amp^R	G. Hotter, unpublished
pMM1	5 kb *Eco*RI fragment from λMM30.4 in pUC19, Amp^R	This study
pMM2	3.3 kb *Eco*RI fragment from λMM30.4 in pUC19, Amp^R	This study
pMM3	1.8 kb *Eco*RI fragment from λMM30.4 in pUC19, Amp^R	This study
pMM4	1.4 kb *Eco*RI fragment from λMM30.4 in pUC19, Amp^R	This study
pMM5	1.5 kb *Eco*RI/*Sal*I fragment from λMM30.4 in pUC19, Amp^R	This study

Strain or Plasmid	Relevant characteristics	Source or Reference
Plasmids cont.		
pMM7	2.0 kb *Bam*HI fragment from λMM3.3 in pUC118, Amp^R	This study
pMM8	0.6 kb *Eco*RI/*Nco*I PCR product in pFunGus	This study
pMM9	1.4 kb *Hin*dIII hygR fragment from pCB1004 in pMM8	This study
pMM10	0.7 kb 3' RACE product in pGEM-T Easy, Amp^R	This study
pUC19	2.7 kb, Amp^R	(Norrander et al., 1983)
pUC118	3.2 kb, Amp^R	(Vieira and Messing, 1987)

2.3.2 LB medium

LB medium contained 10 g tryptone (Difco), 5 g yeast extract and 5 g NaCl for each litre of MilliQ water. The pH was adjusted to 7.0 before autoclaving. For solid media, agar (Davis) was added to 15.0 g/L. Where needed, ampicillin was supplemented at a concentration of 100 µg/mL, isopropylthio-ß-D-galactoside (IPTG) and 5-bromo-4-chloro-3-indolyl-ß-D-galactoside (X-gal) in dimethylformamide were both supplemented at a concentration of 40 µg/mL.

2.3.3 Potato Dextrose Agar (PD agar) and Potato Dextrose Broth (PD broth)

PD broth contained 24.0 g of dehydrated potato dextrose broth (Difco) rehydrated in one litre of MilliQ water. The pH was adjusted to 6.5 prior to autoclaving.

PD agar was prepared by adding 15 g/L of bacteriological agar (Davis) to PD broth.

2.3.4 Regeneration medium

Regeneration medium contained 24 g dehydrated potato dextrose broth and 273.8 g of sucrose per litre of MilliQ water. The pH was adjusted to 6.5 prior to autoclaving, and agar (Davis) was added to a concentration of 15 g/L.

0.8% agar Regeneration medium was made as above, but agar (Davis) was added to a concentration of 8 g/L.

2.3.5 SOC medium

SOC medium contained 20 g tryptone, 5 g yeast extract, 0.6 g NaCl, 0.2 g KCl, 0.95 g $MgCl_2$, 2.5 g $MgSO_4.7H_2O$ and 3.6 g glucose per litre of MilliQ water.

2.3.6 Top Agarose

Top agarose contained 10 g tryptone (Difco), 5 g NaCl and 8 g agarose 15 (BDH) in 1 litre of MilliQ water. The media was cooled to 45-50°C after autoclaving, and supplemented with $MgSO_4.7H_2O$ to a concentration of 10 mM before use.

2.3.7 Media supplements

Supplement	Stock concentration	Final concentration
Ampicillin	100 mg/mL	100 µg/mL
Hygromycin	50 mg/mL	50 µg/mL, 25 µg/mL
Tetracycline	10 mg/mL	10-15 µg/mL

2.4 BUFFERS AND SOLUTIONS

2.4.1 Commonly used stock solutions

Stock	Concentration	pH
Ethanol	70%, 95% (v/v)	
Ethidium bromide	10 mg/mL	
Heparin	5 mM	
IPTG	24 mg/mL	
Maltose	20% (w/v)	
$MgSO_4.7H_2O$	1 M	
Na_2EDTA	250 mM	8.0
PEG solution	20% (w/v)	
Sodium acetate	3 M	7.0
Sodium dodecyl sulfate	10% (w/v)	
Spermidine	50 mM	
Tris-HCl	1 M	8.0
Triton X-100	0.01% (v/v)	
X-gal	20 mg/mL in dimethylformamide	

2.4.2 DNase (RNase free)

RNase free DNase contained deoxyribonuclease I at a concentration of 2 mg/min 0.15 M NaCl and 50% (v/v) glycerol.

2.4.3 DNase I buffer

DNase I buffer ((Bradshaw and Pillar, 1992)) contained 100 mM sodium acetate and 5 mM $MgSO_4$ at pH 5.0.

2.4.4 Electrophoresis solutions

2.4.4.1 SDS loading dye

SDS loading dye contained 20% (w/v) sucrose, 5 mM Na_2EDTA, 1% (w/v) sodium dodecyl sulfate and 0.02% (w/v) bromophenol blue.

2.4.4.2 1 x TAE electrophoresis buffer

1 x TAE buffer contained 40 mM Tris-HCl, 20 mM glacial acetic acid and 2 mM Na_2EDTA per litre.

2.4.4.3 1 x TBE electrophoresis buffer

1 x TBE buffer contained 89 mM Tris-HCl, 89 mM boric acid (pH 8.3) and 2.5 mM Na_2EDTA.

2.4.5 GUS analysis solutions

2.4.5.1 GUS extraction buffer

GUS extraction buffer contained 50 mM NaH_2PO_4 (pH 7.0), 10 mM ß-mercaptoethanol, 10 mM Na_2EDTA, 0.1% sodium lauryl sarcosine, and 0.1% Triton X-100.

2.4.5.2 GUS assay buffer

GUS assay buffer was prepared by adding 4-methylumbelliferyl ß-D-glucuronide (MUG, Sigma) to GUS extraction buffer to a final concentration of 1 mM.

2.4.6 Lysozyme

Lysozyme was prepared at 10 mg/mL in 10 mM Tris (pH 8.0).

2.4.7 Protoplasting solutions

2.4.7.1 OM buffer

OM buffer contained 1.5 M $MgSO_4.7H_2O$, and 10 mM Na_2HPO_4. The pH was adjusted to 5.8 by adding 100 mM $NaH_2PO_4.2H_2O$.

2.4.7.2 PEG buffer

PEG buffer contained 40% (w/v) PEG 4000, 50 mM $CaCl_2$, 1 M sorbitol and 50 mM Tris-HCl (pH 8.0).

2.4.7.3 ST buffer

ST buffer contained 0.6 M sorbitol and 100 mM Tris-HCl (pH 8.0).

2.4.7.4 STC buffer

STC buffer contained 1 M sorbitol, 50 mM $CaCl_2$, and 50 mM Tris-HCl (pH 8.0).

2.4.8 RNase (DNase free)

DNase-free RNase was prepared using RNaseA at 10 mg/mL in 10 mM Tris-HCl (pH 7.5) and 15 mM NaCl. The solution was heated to 100°C for 15 minutes then allowed to cool to room temperature, dispensed into aliquots and stored at –20°C.

2.4.9 SM buffer

SM buffer contained 0.1 M NaCl, 8 mM $MgSO_4.7H_2O$, 50 mM Tris-HCl (pH 7.5) and 0.01% gelatin.

2.4.10 Southern blot solutions

2.4.10.1 10 x Denhardt's hybridisation buffer

10 x Denhardt's buffer (Southern, 1975) contained (per litre): 50 mL 1 M HEPES (Sigma) pH 7.0, 150 mL 20 x SSC, 6 mL herring DNA (3 mg/mL, Sigma), 1.5 mL *E. coli* tRNA (10 mg/mL, Sigma), 5 mL 20% (w/v) sodium dodecyl sulfate, 2 g ficoll (Sigma 70), 2 g bovine serum albumin (Sigma) and 1 g polyvinylpyrrolidone (Sigma PVP-10). MilliQ water was added to one litre.

2.4.10.2 Solution 1

Solution 1 (Southern, 1975) contained 0.25 mM HCl.

2.4.10.3 Solution 2

Solution 2 (Southern, 1975) contained 0.5 M NaOH and 0.5 M NaCl.

2.4.10.4 Solution 3

Solution 3 (Southern, 1975) contained 0.5 M Tris-HCl and 2 M NaCl at pH 7.4.

2.4.10.5 20 x SSC

20 x SSC contained 3 M NaCl and 0.3 M sodium citrate.

2.4.10.6 2 x SSC

2 x SSC was prepared by 10 fold dilution of 20 x SSC.

2.4.11 STE (100/10/1) buffer

STE (100/10/1) buffer contained 100 mM NaCl, 10 mM Tris-HCl (pH 8.0) and 1 mM Na_2EDTA.

2.4.12 STET buffer

STET buffer contained 8% sucrose (w/v), 5% (v/v) Triton X-100, 50 mM Na_2EDTA (pH 8.0) and 50 mM Tris-HCl (pH 8.0).

2.4.13 TE (10/0.1) buffer

TE buffer (10/0.1 containing 10 mM Tris-HCl and 0.1 mM Na_2EDTA) was prepared by diluting 1 M Tris-HCl and 250 mM Na_2EDTA stock solutions described in Section 2.4.1.

2.4.14 10 x TNE

TNE (10 x) contained 0.1 M Tris base, 10 mM Na_2EDTA and 1 M NaCl.

2.4.15 Tris-Equilibrated Phenol

Tris-Equilibrated Phenol was purchased from Amersham, and 0.94 g 8-hydroxyquinoline was added per litre of phenol.

2.5 DNA ISOLATION

2.5.1 Isolation of DNA Fragments from Nusieve® and Seaplaque® Agarose Gels

DNA was recovered from 1.2-2% Nusieve® or 1% Seaplaque® agarose gels by the phenol-freeze extraction method (Thuring et al., 1975). The agarose was made using 1 x TAE (Section 2.4.4.2) and electrophoresis was also carried out in 1 x TAE. Gels were viewed under long wave UV light and fragments of the appropriate size were excised. Each fragment was placed into a 1.5 mL Eppendorf tube, and melted at 65°C in a heat block. An equal volume of Tris-equilibrated phenol was added before the tube was vortexed and left at 20°C for at least two hours. The tube was centrifuged for 10 minutes in a microcentrifuge and the aqueous phase was phenol/chloroform extracted (Section 2.6) and ethanol precipitated (Section 2.7).

2.5.2 Plasmid Isolation by the Rapid Boiling Method

E. coli cells were grown overnight at 37°C in LB broth (Section 2.3.2) supplemented with antibiotic. The cells were pelleted by centrifugation and resuspended in STET buffer (Section 2.4.12). Lysozyme (Section 2.4.6) was added and the solution was boiled for 40 seconds. The pellet was removed after centrifugation in a microcentrifuge for 10 min. An equal volume of isopropanol was added to the supernatant, mixed and left at –20°C for 10 to 20 min. The plasmid DNA was pelleted by centrifugation for 10 min, and washed once with 70 or 95% ethanol. The pellet was left to dry at 37°C for about 20 min and resuspended in 50 µL of MilliQ water. This method is based on that of (Holmes and Quigley, 1981).

2.5.3 Plasmid Isolation using the Quantum Plasmid Preparation Kit (BIO-RAD)

E. coli cells containing the plasmid of interest were grown overnight at 37°C in 3-5 mL of LB broth (Section 2.3.2) supplemented with antibiotic. One to five mL of culture was pelleted by centrifugation for 30 seconds, and the supernatant was discarded. The pelleted cells were resuspended in cell resuspension buffer. Cell lysis solution was added to the tube, followed by neutralisation solution. The tube was spun for five minutes, and the supernatant mixed with the Quantum preparation matrix. The supernatant was washed through a spin filter, and the matrix washed twice in wash solution. DNA was eluted from the matrix using MilliQ water and the DNA concentration checked.

2.5.4 Plasmid isolation using QIAGEN® Kit (Qiagen)

E. coli cells were grown as in Section 2.5.3, and cells harvested by centrifugation. The cells were resuspended in Buffer P1. Buffer P2 (lysis solution) was then added and mixed by inversion. Neutralisation solution (Buffer N3) was added and the tube was centrifuged for 10 minutes. The supernatant was poured onto a QIAprep column, which was washed with an optional wash buffer, then with a wash buffer. DNA was eluted using MilliQ water or TE (10/0.1).

2.5.5 Purification of phage λ DNA

Phage was plated as described in Section 2.13.1, with the phage diluted to give approximately 10^6 plaque forming units (PFU) per plate. Once confluent lysis had been obtained, 5 mL of SM buffer was added to each plate. The plates were left at 4°C overnight.

The lysate was collected and centrifuged at 3, 020 x g for 10 minutes at 4°C. DNase (Section 2.4.2) and RNase (Section 2.4.8) were added to the supernatant at a concentration of 1 µg/mL each. Reactions were incubated at 37°C for 30 minutes. 5 mL of 20% (w/v) PEG solution was added and the tubes left to stand on ice for one hour. The phage was pelleted by centrifugation at 7, 710 x g for 15 minutes and resuspended in 0.5 mL of SM buffer (Section 2.4.9), 5 µL 10% SDS and 10 µL of 250 mM Na_2EDTA. The tubes were incubated at 68°C for 15 minutes, then briefly vortexed. An equal volume of phenol was added, vortexed for 10 seconds, and left to stand at room temperature for five minutes. Samples were vortexed for another 10 seconds and the aqueous phase phenol/chloroform extracted (Section 2.6). The DNA was resuspended in 50 µL of MilliQ water and 10 µg of RNase (Section 2.4.8) was added. The DNA was quantitated as described in Section 2.8.

2.5.6 Purification of PCR products

PCR products were purified using the Concert™ Rapid PCR purification system (GibcoBRL). The sample was mixed with Binding solution (H1) and loaded onto a spin cartridge. The filtrate was discarded after centrifuging for one minute. Wash buffer (H2) was added to the spin filter and centrifuged for one minute. The filtrate was discarded, and tubes were centrifuged for one further minute. The DNA was eluted by adding 50 µL of warm TE buffer (Section 2.4.13) and incubated at room temperature for one minute, followed by centrifugation for two minutes. The purified PCR product was stored at –20°C.

2.6 PURIFICATION OF DNA BY PHENOL/CHLOROFORM EXTRACTION

Equal volumes of Tris-equilibrated phenol (Section 2.4.15) and chloroform were added to DNA samples and mixed before centrifugation for 5 minutes in a microcentrifuge. Two volumes of chloroform were added to the aqueous phase of the phenol/chloroform extraction. The tube was mixed before centrifugation for 3 min. The DNA in the aqueous phase from the chloroform extraction was precipitated with either ethanol or isopropanol (Section 2.7).

2.7 PRECIPITATION OF DNA WITH ETHANOL OR ISOPROPANOL

DNA samples were precipitated by adding $^{1}/_{10}$ volume 3 M sodium acetate (Section 2.4.1) and 2.5 volumes of 95% (v/v) ethanol or 0.6 volumes of isopropanol. Samples were mixed gently and left at –20°C for 30 minutes to 2 hours. The DNA was pelleted by centrifugation for 10 minutes and washed in 70% (v/v) or 95% (v/v) ethanol. The pellet was left to dry at 37°C for 15 to 30 minutes before resuspension in an appropriate volume of MilliQ water or TE (10/0.1) (Section 2.4.13).

2.8 DNA QUANTIFICATION

DNA was quantified by using a fluorometer, or by intensity of ethidium bromide fluorescence.

2.8.1 Fluorometric Quantitation of DNA

DNA samples were quantitated on a Hoefer Scientific TKO 100 fluorometer, in a solution containing 1 x TNE buffer (Section 2.4.14) and 0.1 µg/mL of Hoechst dye 33258 (Sigma). The fluorometer scale was set to 100 using 2 µL of 100 µg/mL calf thymus DNA added to 2 mL of dye solution. 2 µL of the sample DNA was added to 2 mL of dye solution and the resulting reading was recorded as the concentration of the solution in ng/µL.

2.8.2 Minigel Method for Determination of DNA concentration

A sample of the DNA solution was separated by agarose gel electrophoresis with a series of standard DNA solutions. This method was commonly used for plasmid and λ DNA. Smaller DNA fragments were quantitated against a low DNA Mass ladder (GibcoBRL) which consisted of a ladder of fragments of varying concentrations. Once the SDS dye front had migrated at least two-thirds of the way down the gel, the DNA was stained with ethidium bromide and photographed. The intensity of fluorescence from the unknown DNA sample was compared to that of the known standards.

2.9 RESTRICTION ENDONUCLEASE DIGESTION OF DNA

Restriction endonuclease digestions were carried out using manufacturer's buffer supplied with the enzyme. An excess of enzyme was used to cut the DNA. Digestions of plasmid and λ DNA were performed at 37°C for two hours and stored at -20°C while an aliquot was checked on an agarose gel. If the DNA had not digested to completion, more enzyme was added and the digests were reincubated. For DNA stocks that were difficult to digest, a further phenol/chloroform extraction (Section 2.6) was performed and the digest was repeated.

Digestions were performed in a similar manner for fungal genomic DNA, except incubation times were increased (three hours up to overnight if necessary). When endophyte genomic DNA was digested with *Eco*RI and *Sst*I restriction endonucleases, 50 nmol of spermidine was added to each digest to relax the DNA and facilitate digestion. No spermidine was added to *Hin*dIII digests.

2.10 AGAROSE GEL ELECTROPHORESIS OF DNA

Agarose gel electrophoresis was performed either in a mini-gel apparatus for 1-2 hours, or in a Bio-Rad DNA Sub-Cell overnight. Agarose concentrations varied from 0.7% to 2% (w/v) dissolved in TAE (Section 2.4.4.2) or TBE (Section 2.4.4.3) electrophoresis buffer. The agarose was melted in either a pressure cooker or microwave, and allowed to equilibrate to 50°C before the gel was poured. DNA samples were mixed with SDS dye (Section 2.4.3.1) before loading. Electrophoresis was carried out at 80 to 120 V for minigels, or at 30 V for overnight gels (Bio-Rad Sub-Cell).

2.10.1 Staining and photographing gels

Once the loading dye had moved at least half way down the gel, it was stained in an ethidium bromide bath (1 μg/mL ethidium bromide in MilliQ water) before destaining in MilliQ water. Bands on the gel were visualised using an UV transilluminator and photographed using an Alpha Innotech gel documentation system.

2.10.2 Sizing DNA fragments

DNA fragment sizes were determined after electrophoresis by using Alpha Innotech gel documentation system software. This program compares the distance travelled by molecular markers of known size, such as the λ/*Hin*dIII ladder, with the fragments of unknown size.

2.11 SUBCLONING

Subcloning involved ligation of DNA fragments into a suitable vector, and electroporation of ligation mixtures into *E. coli* cells. Recombinants were screened by two methods: either examining restriction digests of rapid boil plasmid DNA (Section 2.5.2) by gel electrophoresis (Section 2.10) or by analysing the products generated by colony PCR (Section 2.14.2).

2.11.1 DNA ligations

2.11.1.1 CAP treatment of Vector DNA

Approximately 5 µg of vector DNA was digested with an appropriate amount of restriction enzyme. Where possible, the restriction enzyme was inactivated by heat. Half a unit of calf intestine alkaline phosphatase (Boehringer Mannheim) was added and the reaction was incubated at 37°C for 30 minutes. Na_2EDTA and SDS were added to concentrations of 5 mM and 0.5% (w/v) respectively. Proteinase K was added to a final concentration of 50 µg/mL and the mixture was incubated at 56°C for 30 minutes. The reaction was cooled to room temperature then phenol/chloroform extracted (Section 2.6) and ethanol precipitated (Section 2.7). Control ligations were carried out using before and after ligation samples of vector from before CAP, after CAP and CAP'd vector + insert samples.

2.11.1.2 SAP Treatment of Vectors

DNA was digested as described for CAP treatment (Section 2.11.1.1). The vector was purified using phenol/chloroform purification (Section 2.6), and ethanol precipitated (Section 2.7). A sample was set aside as a before SAP control, before a reaction was set up containing the purified vector, 1 x SAP buffer (50 mM Tris-HCl, 0.1 mM Na_2EDTA, pH 8.5) and 1-2 µL of shrimp alkaline phosphatase (SAP, Boehringer Mannheim). The reaction

was incubated at 37°C for 30-50 minutes and the enzyme inactivated at 65°C for 15 minutes. Any excess SAP and salts were removed by phenol/chloroform purification (Section 2.6). The SAP treated vector underwent the same control ligations described for CAP treated vectors (Section 2.11.1.1).

2.11.1.3 Ligations

Ligation mixtures contained 2 µL 10 x ligation buffer (NE Biolabs), a three times molar excess of insert to vector, and 2 µL of a five-fold dilution of T4 DNA ligase (NE Biolabs). For non-directional cloning, vectors were either pre-treated with CAP or SAP (Sections 2.11.1.1 and 2.11.1.2). For each ligation, two different ratios of insert to vector were performed. A vector only control was included. Pre-ligation samples were removed and placed in 5 µL of SDS loading dye and the remaining ligation mix was incubated at 4°C overnight.

2.11.1.4 Shotgun cloning of λ DNA fragments

λ clones were digested with an appropriate restriction enzyme (Section 2.9), and then phenol/chloroform purified (Section 2.6) and ethanol precipitated (Section 2.7). The digestion was then included in a ligation reaction with 20 ng of vector as described in Section 2.11.1.3.

2.11.2 Transformation of *E. coli* by electroporation

2.11.2.1 Preparation of Electro-Competent E. coli cells

One litre of LB broth was inoculated with XL-1 *E. coli* cells and grown at 37 °C with shaking until the optical density (A_{600}) of the culture is between 0.5 and 1.0. The cells were chilled on ice for 20 minutes then harvested by centrifugation at 4, 000 g for 10 minutes (all centrifugations were performed at 4 °C). The cells were washed in ice cold water (1 L, then 500 mL), then 20 mL of ice cold 10% glycerol. The pelleted cells were resuspended in 4 mL of ice cold 10% glycerol and stored in 40 µL aliquots at –80°C.

2.11.2.2 Electroporation

Electro-competent XL-1 *E. coli* cells were gently thawed, then left on ice. The Gene Pulser (Bio-Rad) was set at 25μF, 2.5 kV and pulse controller to 200 Ω resistance. One to 2 μL of DNA was mixed with the electrocompetent cells, which were left on ice for one minute. The mixture was transferred to an ice cold 0.2 cm cuvette. The mixture was shaken to the bottom of the cuvette and pulsed at the settings described above. If the time constant was between 4 and 5 milliseconds, the electroporated cells were resuspended in 500 μL to 1 mL of SOC medium. A cells only sample was included as a negative control and a sample transformed with 20 ng of pUC118 was included as a positive control. The transformed cells were incubated at 37°C for one hour, then plated with suitable dilutions onto selective LB plates.

2.11.3 Screening of transformants

Several methods were used to screen for transformants that contained DNA of interest. Blue-white screening was used when transforming a single fragment in a vector containing the lacZ gene (e.g. pUC118). Sometimes blue-white selection could not be used because of the prior insertion of another fragment. When blue-white screening was available, the transformed cells were plated on selective LB agar plates with IPTG and X-gal. Transformants were always screened by restriction mapping of rapid boil DNA (Section 2.5.3).

PCR screening was used to identify colonies that contained the plasmid with the desired fragment after transformation (Section 2.11.2.2). A 'cocktail' was set up as described in Section 2.14.1, except the volume of each reaction was made up to 25 μL with MilliQ water. Instead of applying 5 μL of template, a 10 μL Gilson pipette tip was used to introduce a small amount of a colony into the PCR reaction. The PCR machine was preheated to 95°C, and after 3 min at 95°C, the reactions underwent 30 cycles at 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for x minutes (x = 1 min/kb DNA amplified). Transformants screened by PCR were also analysed by restriction mapping of rapid boil DNA.

2.12 SOUTHERN BLOTTING AND HYBRIDISATION

2.12.1 Southern Blotting

Overnight agarose gel electrophoresis was used to separate the DNA to be transferred to the membrane, and the gel was stained, visualised and photographed as described in Section 2.10. The gel was placed in a tray containing Solution 1 (Section 2.4.10.2) and gently agitated for 15 minutes. The solution was poured off and replaced with Solution 2 (Section 2.4.10.3), and gently agitated for 30 minutes. This solution was replaced with Solution 3 (Section 2.4.10.4) and agitated for 15 minutes, before a final wash for two minutes in 2 x SSC (Section 2.4.10.6).

During the gel treatment, a plastic trough with wells at both ends was prepared so two sheets of Whatman 3MM chromatography paper soaked in 20 x SSC (Section 2.4.10.5) protruded into the wells, which were then filled with 20 x SSC. Plastic wrap was placed over the trough and pressed flat. A grid 2 mm smaller than the gel size was removed from the plastic wrap and gel was placed so it overlapped the edges of the grid. A piece of positively-charged nylon membrane (Boehringer Mannheim) was cut to 2 mm greater than the gel size, pre-soaked in 2 x SSC, and laid on the gel. Four sheets of Whatman 3MM chromatography paper were cut to 2 mm less than the gel size. Two of these were pre-soaked in 2 x SSC and then laid on the membrane, followed by the other two dry sheets. A pile of paper towels was placed at the top of the blotting apparatus and weighed down with a media bottle. The apparatus was left to blot overnight.

The next day the blot apparatus was disassembled, and the membrane was washed in 2 x SSC for 5 minutes. After drying between blotting paper, the Southern blot was baked in a vacuum oven at 80°C for 2 hours.

2.12.2 Preparation of Labelled Probe with High Prime DNA Labelling Kit

DNA to be labelled (25-50 ng) was placed in 11 µL volume, denatured by boiling for 3 min, then placed immediately on ice. 4 µL of High Prime solution (Boehringer Mannheim) and 5 µL of [α^{32}P]dCTP (3000 Ci/mmol, Amersham) was added. Reactions were incubated at 37°C for 1 hour before the reaction was stopped by adding 30 µL of STE buffer (Section 2.4.11). Unincorporated nucleotides were removed from the mixture using a Sephadex G-50 column (ProbeQuant). The vortexed column was prespun to remove the void volume from

the column. The probe sample was applied and the column was spun for two minutes at 735 x g.

2.12.3 DNA Hybridisation

A Southern blot (Section 2.12.1) membrane was prehybridised for at least 2 hours in 10 x Denhardt's hybridisation buffer (Section 2.4.10.1) at 65°C. The boiled [α^{32}P]dCTP-labelled probe (Section 2.12.2) was added and left to hybridise at 65°C overnight. Following removal from the tube, the blot was washed three times in 2 x SSC for 15 minutes at room temperature.

2.12.4 Autoradiography

The blot was wrapped in plastic, and exposed in the presence of a Cronex intensifying screen to a sheet of Fuji Medical X-ray film in an X-ray cassette at –70°C for an appropriate period of time. The film was developed in Kodak D19 developer for two minutes, rinsed in water, and fixed in Kodak Rapid fixer for two minutes. After rinsing in water, the film was left to dry.

2.12.5 Stripping filters and blots for re-use

A solution of boiling 0.1% (w/v) sodium dodecyl sulfate (SDS) was poured over the filter or blot and gently agitated overnight while the solution cooled to room temperature. This process was repeated three times to ensure all probe DNA was removed from the filter. The filter was checked by autoradiography (Section 2.12.3) to ensure that no signal from the probe remained. If stripping was incomplete, the process was repeated.

2.13 LIBRARY SCREENING BY PLAQUE HYBRIDISATION

A Lp19 genomic DNA library in λGEM-12 was screened by plating phage, taking filter lifts off the resulting plaques, and hybridising an appropriate probe to the filters. Positive plaques were cut out of the plate and stored in SM buffer with a drop of chloroform. Positive plaques were screened twice more to purify the phage before DNA was extracted from positive plaques (Section 2.5.5).

2.13.1 Plating phage λ

E. coli strain KW251 was used as a host for phage λ. The λ population to be screened was diluted to 1.8 x 10^5 PFU/mL and 50 µL was combined with 50 µL of SM buffer (Section 2.4.9). The diluted phage was combined with 100 µL of *E. coli* KW251 cells grown overnight in LB (Section 2.3.2) supplemented with 10mM $MgSO_4.7H_2O$ and 0.2% (w/v) maltose. The phage-*E. coli* mixture was incubated at 37°C for 30 minutes. The mixture was combined with 3 ml of supplemented Top agarose (Section 2.3.6) and overlaid on a LB agar or agarose plate. The plates were incubated at 37°C for about six hours, then stored at 4°C overnight.

2.13.2 Filter lifts

Filters (NEN Life Science) were placed on the bacteria/phage lawn and marked asymmetrically so the filter position relative to the plate could be determined later. For the first round of screening, three lifts were taken. The first lift was left for 45 seconds after the filter was completely wet; the second was left for 90 seconds and the third lift for 2 minutes. For the second and third rounds of screening, the first lift was left for one minute and the second lift for two minutes. Filters removed from the plate were placed DNA side up on three layers of 3MM paper soaked in Solution 2 (Section 2.4.10.3) for two minutes, Solution 3 (2.4.10.4) for five minutes, and in 2 x SSC (Section 2.4.10.6) for two minutes. The filters were left to air dry and baked in a vacuum oven at 80°C for 2 hours.

2.13.3 Hybridisation of phage λ DNA to [α-^{32}P]dCTP-labelled DNA probes

Probe labelling, filter hybridisation, and autoradiography were performed as described in Sections 2.12.2- 2.12.4, except hybridisations were performed in plastic pots with up to 15 filters per pot. Positive plaques were identified as autoradiograph signals in identical positions on duplicate filters. The plates were aligned with the autoradiograph, and plaques corresponding to signals were removed from the plate using a cut-off 1 mL Gilson pipette tip.

2.14 DNA AMPLIFICATION BY THE POLYMERASE CHAIN REACTION (PCR)

2.14.1 Primers

Oligonucleotide primers were synthesised by Life Technologies or Sigma Genosys. Each primer was resuspended to a final concentration of 100 µM. For PCR reactions, primer stocks were diluted to 10 µM. Sequencing stocks were prepared from primer stocks at a concentration of 3.2 µM. Primers were stored at –20°C until needed. Primers used in this study are shown in Table 2.2.

2.14.2 General PCR reaction

A 'cocktail' which contained the ingredients necessary for (n+1) reactions was prepared for all PCR reactions. A normal PCR reaction contained: 2.5 µL of 10 x Taq polymerase buffer (containing $MgCl_2$), 1.25 mM each of dNTPs, 5 pmol of "forward" primer, 5 pmol of "reverse" primer, 0.05 U of Taq polymerase (Roche), and MilliQ water to give a 20 µL volume. 5 µL of template was added was added to each PCR reaction. A negative control containing water as a template was used for each 'cocktail' and a positive control was included where possible. All PCR reactions were placed in a thermal cycler (Corbett) preheated to 95°C. Commonly used PCR programs are shown in the table below.

Reaction type	Primers used	Program
RT-PCR	MM2/MM5	94°C 2 min
	hmg29/hmg30	94°C 30 sec ┐
		60°C 30 sec ├ x 35
		72 °C 1 min/kb ┘
		72°C 5 min
		4°C/25°C 1 min
PCR direct	pUChph3/	95°C 3 min
from colony	pUChph4	95°C 30 sec ┐
		60°C 30 sec ├ x 30
		72°C 1 min/kb ┘
		72 °C 5 min
		4°C/25°C 1 min

Table 2.2 Primers used in this study

Primer	Sequence	Anneals to
MM1	CAA CGA CAT CGT CCG AAA G	*prt*1 (Lp19)
MM2	GTG ATCCAG TCG AGA GTC	*prt*1 (Lp19)
MM3	AAG TCT CGC CAT GAC CAC	*prt*1 (Lp19)
MM4	CAG GTC GAG GTT GTT GAG	*prt*1 (Lp19)
MM5	TGA TGC CTG GAC ATG TTG	*prt*1 (Lp19)
MM6	TCG TTC AGC GAC TGC GAG	*prt*2 (Lp19)
MM7	GCC TCC CAG TTA GCA TTC	*prt*1 (Lp19)
MM8	CGG ACG CGT GTG ACT GAC	*prt*1 (Lp19)
MM9	CCA AAC CAA CAT GTC CAG	*prt*1 (Lp19)
MM10	TGG GCA ACG ACA AGG ATG	*prt*1 (Lp19)
MM11B	CTC GAT GTA GTC AAC CTG	*prt*2 (Lp19)
MM12	AGC AGC GCG ATG CTC CTC	*prt*2 (Lp19)
MM13	ATC AAG GTA CTC AGC GAC	*prt*2 (Lp19)
MM14	GAC TTC TTT GAG CCC GAG	*prt*1 (Lp19)
MM15	GTG ACA TTG GTG GCT ACG	*prt*2 (Lp19)
MM16	GAT CGA ACA TCA CCT CTG	*prt*1 (Lp19)
prot1Nco	GTT CAC CAT GGT GGT TCT CAA CTG C	*prt*1 translation start
hmg29[1]	GAC GTG GTC AAA TCC GTG TTG	hmg (Lp19)
hmg30[1]	GTT GGA TGT GAT CCT CGC AC	hmg (Lp19)
pUChph3	CTG CAT CAT CGA AAT TGC	hph (pCB1004)
pUChph4	AAA CCG AAC TGC CCG CTG TTC	hph (pCB1004)
for[2]	GCC AGG GTT TTC CCA GTC ACG A	M13 lacZ
rev[2]	GAG CGG ATA ACA ATT TCA CAC AGG	M13 lacZ
T7	TAA TAC GAC TCA CTA TAG GG	T7 promoter

[1] (Dobson, 1997)

[2] These primers are the forward and reverse primers from the Amplicycle sequencing kit (PE Biosystems)

2.15 DNA SEQUENCING

2.15.1 Automated DNA sequencing

DNA prepared using plasmid isolation kits (Section 2.5.4 and 2.5.5) was sequenced by the Sanger method using BigDye (ABI) dye terminator chemistry, and samples separated using an ABI 377 automatic sequencer.

2.15.2 Sequence analysis

DNA sequences were assembled and analysed using the Wisconsin Package version 9.1, Genetics Computer Group (GCG), Madison, Wisconsin, and Sequencher™ 3.1.1. For the Wisconsin Package (GCG) the programs used were SeqEd, GetSeq, fragment assembly packages, Map and FastA.

Protein sequences were analysed for motifs using the Findpattern program. Protein sequences were aligned using ClustalW and PIMA sequence alignment programs. Blast-X searches were performed using the Basic BLAST option of BLAST 2.0 (Altschul et al., 1997) on the Internet at the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nih.gov). Codon frequency tables for genes were constructed using the CODONFREQUENCY program from the Wisconsion Package (GCG).

2.16 TECHNIQUE FOR WORKING WITH RNA

In preparation for RNA work, all glassware was washed in chromic acid overnight and oven baked at 180°C for 2 hours. Solutions were made with reagents used for RNA work only and DEPC-treated water, and placed in acid-washed oven-baked glassware. Mortars, pestles and metal spatulas were soaked overnight in 0.3% hydrogen peroxide before use. DNase and RNase-free barrier tips were used to further reduce RNase contamination. Gloves were worn throughout all RNA work.

2.17 ISOLATION OF RNA FROM FUNGAL CULTURES

Fungal mycelia were harvested from Lp19 cultures grown in liquid media, and stored at –80°C or in liquid nitrogen. During RNA isolation from fungal material, samples were kept in liquid nitrogen to prevent RNase digestion until Trizol reagent was added. Mycelia were

ground to a fine powder in liquid nitrogen, and 10 mL of Trizol was added for each gram of mycelia. The sample was centrifuged at 11, 366 x g for 10 minutes at 4°C, and the aqueous phase was extracted in chloroform by centrifugation at 11, 366 x g for 15 minutes at 4°C.

The RNA from the aqueous phase of the chloroform extraction was precipitated with isopropanol, and pelleted by centrifugation at 11, 366 x g for 10 minutes at 4°C. The pellet was washed in 70% ethanol, then left to air dry. The sample was resuspended in DEPC-treated water, and RNA was quantitated by the absorbance of the sample at 260 and 280 nm. RNA samples were stored at –80°C in preparation for DNaseI digestion.

2.17.1 DNase I treatment of RNA

RNA prepared in section 2.17 was treated with DNaseI in the presence of RNase inhibitor (Boehringer Mannheim), DTT, and a DNaseI buffer (Bradshaw and Pillar, 1992). After incubation with DNaseI, the RNA was phenol/chloroform extracted and centrifuged at 15, 900 x g at 4°C for 10 min. The aqueous phase was extracted in 2 volumes of chloroform, then centrifuged at 15, 900 x g for 10 minutes at 4 °C. The RNA was ethanol precipitated (Section 2.7), pelleted and washed in 75% ethanol. The pellet was resuspended in DEPC-treated water and stored at –80°C.

2.18 RT-PCR ANALYSIS OF RNA

2.18.1 Synthesis of cDNA using random primers

To make cDNA using random primers, 1 µg of DNased RNA (Section 2.17) was incubated with random hexamer primers at 90°C for 5 minutes. No RT and no RNA controls were also set up for each cDNA reaction. All of the techniques described in Section 2.16 were used to prevent RNase digestion. After the initial annealing step, a 'cocktail' was prepared, and added to the cDNA, no RT and no RNA reactions. Each reaction contained 1 x Expand RT buffer (Boehringer Mannheim), 200 nmol DTT, 1 nmol dNTPs and 8 U RNase inhibitor (Roche). Expand RT (Boehringer Mannheim) was added to the cDNA and no RNA reactions.

Reactions were incubated at room temperature for 10 minutes to allow primer annealing, then incubated at 42°C for 45 minutes. Ten-fold, hundred-fold and thousand-fold dilutions

were prepared for each cDNA sample synthesised. Diluted stocks of cDNA and the control reactions were stored at –20°C until use.

2.18.2 RT-PCR using cDNA

cDNA prepared in Section 2.18.1 was used as a template for RT-PCR, using PCR conditions described in Section 2.14.2. Primers used for RT-PCR were generally designed to anneal to sequences flanking an intron, which should give a size difference between the genomic and cDNA PCR products.

2.19 3' RACE ANALYSIS

2.19.1 Preparation of cDNA using oligo dT primer

In order to make cDNA that represented Lp19 mRNA, cDNA was prepared in a similar manner to Section 2.18.1. An oligo-dT primer, which anneals to the polyA tail of mRNA, was used instead of random primers. Instead of the Expand RT reverse transcriptase used in Section 2.18.1, *C. therm* reverse transcriptase (Roche) was used. Each reaction contained 1 x RT buffer, 100 nmol DTT, 16 nmol dNTPs and 8% DMSO. The reactions were amplified at 60°C for 30 minutes, followed by heat inactivation of RT at 95°C for 2 minutes.

2.19.2 Amplification using gene specific primers

The cDNA pool created in Section 2.19.1 was amplified using a gene-specific primer (GSP1) and a universal adapter primer (UAP), which is complementary to sequence in the oligo dT primer. The products of this reaction were amplified using a nested gene specific primer (GSP2) and UAP. The product of the GSP2-UAP reaction was subcloned into a pGEM-T Easy vector (Appendix A1.4) in preparation for sequencing.

2.20 PREPARATION OF FUNGAL PROTOPLASTS

Fungal cultures inoculated with 5 x 10^6 spores of *P. paxilli* PN2013 were grown in 25 mL of PD broth at 22°C with shaking at 200 rpm for 48 hours. The mycelia were washed three times in sterile MilliQ water. Mycelia were washed in OM buffer (Section 2.4.7.1), before being mixed in 30 mL of 10 mg/mL Glucanex (ChemColour Industries N. Z. Ltd.) in sterile OM buffer (Section 2.4.7.1). The mycelia were left to shake at 80 rpm overnight at 30°C.

The mycelial solution was filtered using Miracloth, and the filtrate was overlaid with 2 mL of ST buffer (Section 2.4.7.3). The protoplasts formed a white interface between the two liquid phases following centrifugation at 3, 020 x g for 5 min. The protoplasts were removed, and washed three times in STC buffer (Section 2.4.7.4) by centrifugation at 7, 710 x g. Protoplasts were resuspended in STC buffer, at a concentration of 1.25 x 10^8 protoplasts/mL, in preparation for transformation.

2.21 TRANSFORMATION OF *PENICILLIUM PAXILLI*

2.21.1 Transformation of *P. paxilli*

In order to transform protoplasts, 80 µL of the protoplasts prepared in Section 2.20 were combined with 20 µL of PEG buffer (Section 2.4.7.2), 2 µL spermidine, 5 µL heparin and 5 µg of DNA. This PEG-mediated transformation method is based on that used by Yelton et al.. The transformation mixtures were left on ice for 30 minutes, before a further 900 µL of PEG buffer was added. Transformations were incubated on ice for 15-20 minutes, before 100 µL aliquots of the transformation mixture were added to 5 mL of 0.8% RG agar (Section 2.3.4) and overlaid on RG plates. The plates were incubated at 22°C overnight. On the next day, transformation plates requiring selection were overlaid with hygromycin in a 0.8% RG overlay to give a final hygromycin concentration of 50 µg/mL.

2.21.2 Single spore purification of transformants

Transformants were single spore purified by suspending the spores in 0.01% Triton, and streaking for single colonies on PD agar plates containing 50 µg/mL hygromycin. Plates were incubated at 22°C for 5 days. Spores from the first round of single spore purification were restreaked on to new PD agar plates with 25 µg/mL hygromycin. The plates were incubated at 22°C for 5 days.

2.22 ß-GLUCURONIDASE ASSAY

A loopful of spores from transformants were incubated with 100 µL of GUS assay buffer (Section 2.4.5.2) for one hour at 22°C in the dark. GUS assay buffer only was used as a negative control, and duplicate assays were carried out for spores from each transformant.

The fluorescence was visualised using UV light and the Alpha Innotech gel documentation system.

CHAPTER THREE

ANALYSIS OF THE LP19*PRT*1 GENE

3.1 SOUTHERN BLOTTING OF ENDOPHYTE DNA

A 550 bp *Eco*RI/*Sst*I fragment from pGH30 (Appendix A2.2) was [α^{32}P]dCTP-labelled and used to probe a Southern blot containing Lp1 (LpTG-2), Lp19 (*N. lolii*) and E8 (*E. typhina*) genomic DNA. All of these endophytes contained DNA that hybridised to the probe (Figure 3.1). The sizes of hybridising fragments are shown in Table 3.1. Lp1 had at least two hybridising bands in each of the *Eco*RI and *Hin*dIII digests (Figure 3.1B, lanes 2-3). For Lp19 single hybridising bands were observed for *Eco*RI, *Hin*dIII and *Sst*I digests. E8 *Eco*RI and *Sst*I digests also contained single hybridising bands. It was difficult to determine which *Hin*dIII fragment hybridised for E8 because digestion appeared to be incomplete (lane 9).

Southern blots had confirmed that sequences homologous to the pGH30 insert were in three different endophytes. To look for the presence of similar sequences in other endophytes, genomic DNA from representatives of each asexual endophyte taxonomic group was digested with *Hin*dIII and Southern blotted using the same probe as in Figure 3.1 (Figure 3.2). The sizes of hybridising fragments are shown in Table 3.2. Lanes 2, 3 and 4 contain DNA from the same endophytes as in the initial blot. Lp19 (lane 2) and Lp1 (lane 3) had hybridising fragments similar in size to those in the first blot while E8 (lane 4) contained an 11.5 kb hybridising fragment which was unclear in the first blot.

The other genomic DNA on the second blot came from AR510 (FaTG-3), Fp5 (*N. uncinatum*), Tf2 (*N. coenophialum*) and Tf13 (FaTG-2). All of the endophytes contained at least one band that hybridised to the pGH30 insert. Lp1 (lane 3), E8 (lane 4), AR510 (lane 6) and Tf2 (lane 8) all shared a common band of 11.5 kb. Tf2 (lane 8) and Tf13 (lane 9) also share a common band of approximately 10 kb with Lp19 (lane2). The *N. uncinatum* band (lane7) is larger in size than any of the other hybridising bands.

3.2 LIBRARY SCREENING AND MAPPING

3.2.1 Library screening

Filters containing 48,000 plaque-forming units from a λGEM-12 (Appendix A1.6) Lp19 library (prepared by J. M. Dobson) were screened by plaque hybridisation using a [α^{32}P]dCTP-labelled *Eco*RI/*Sst*I insert from pGH30. For the first round of screening

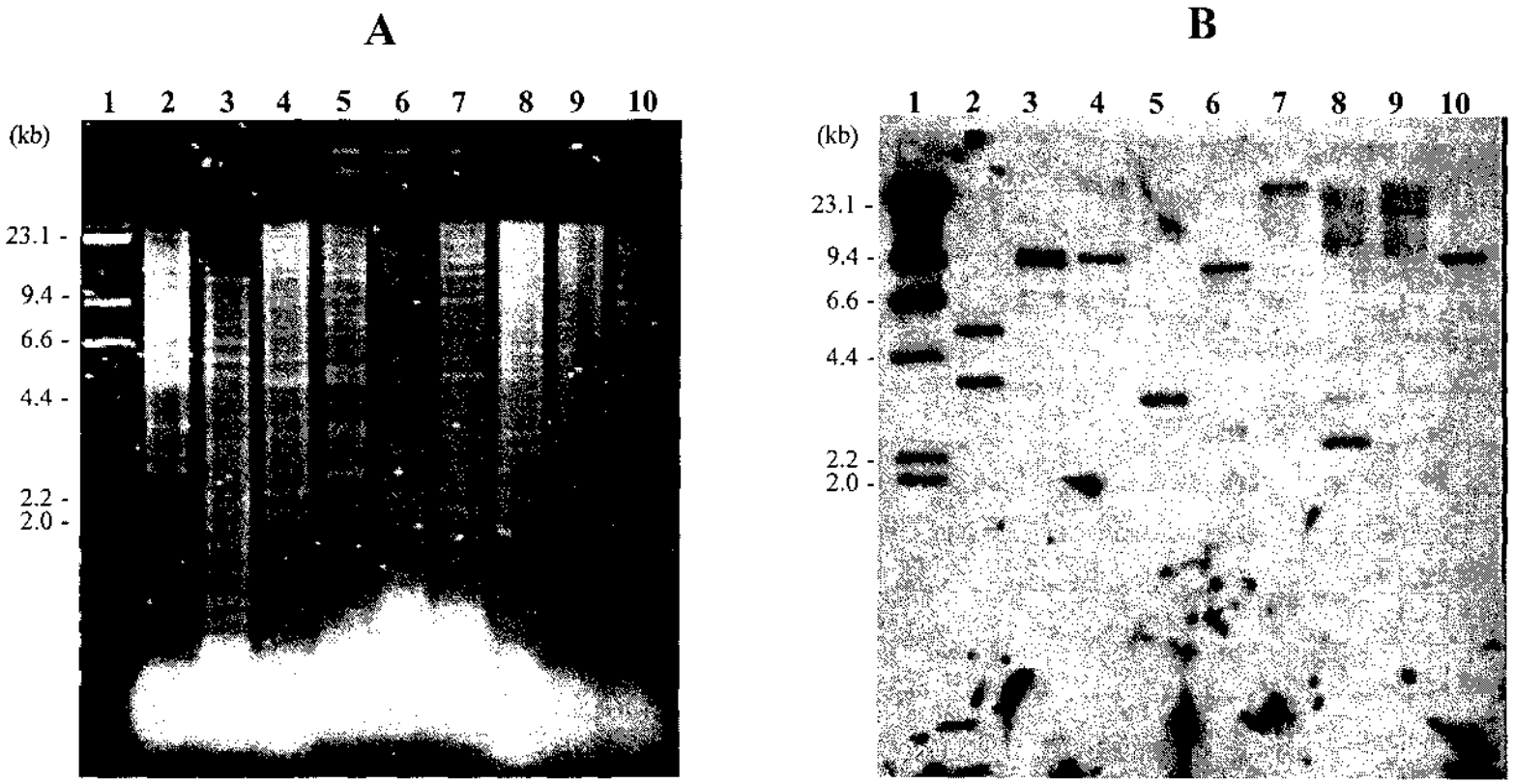


Figure 3.1 Southern blot analysis of the *prt*1 gene.

A. Southern blot of 2 µg of Lp1 (lanes 2-4), Lp19 (lanes 5-7) and E8 (lanes 8-10) genomic DNA digested with *Eco*RI (lanes 2, 5 and 8), *Hin*dIII (lanes 3, 6 and 9) and *Sst*I (lanes 4, 7 and 10). Lane 1 contains λ DNA digested with *Hin*dIII.

B. Autoradiograph of this blot probed with [α^{32}P]dCTP-labelled pGH30 insert.

Table 3.1 Genomic DNA fragments homologous to the pGH30 insert.

	*Eco*RI	*Hin*dIII	*Sst*I
Lp1	6.2 3.8	11.5 10.7 8.3	20.7 11.3
Lp19	3.3	10.7 8.5	20.7
E8	3.3 2.6	17.3 12.5	12.1

Fragment sizes are shown in kb.

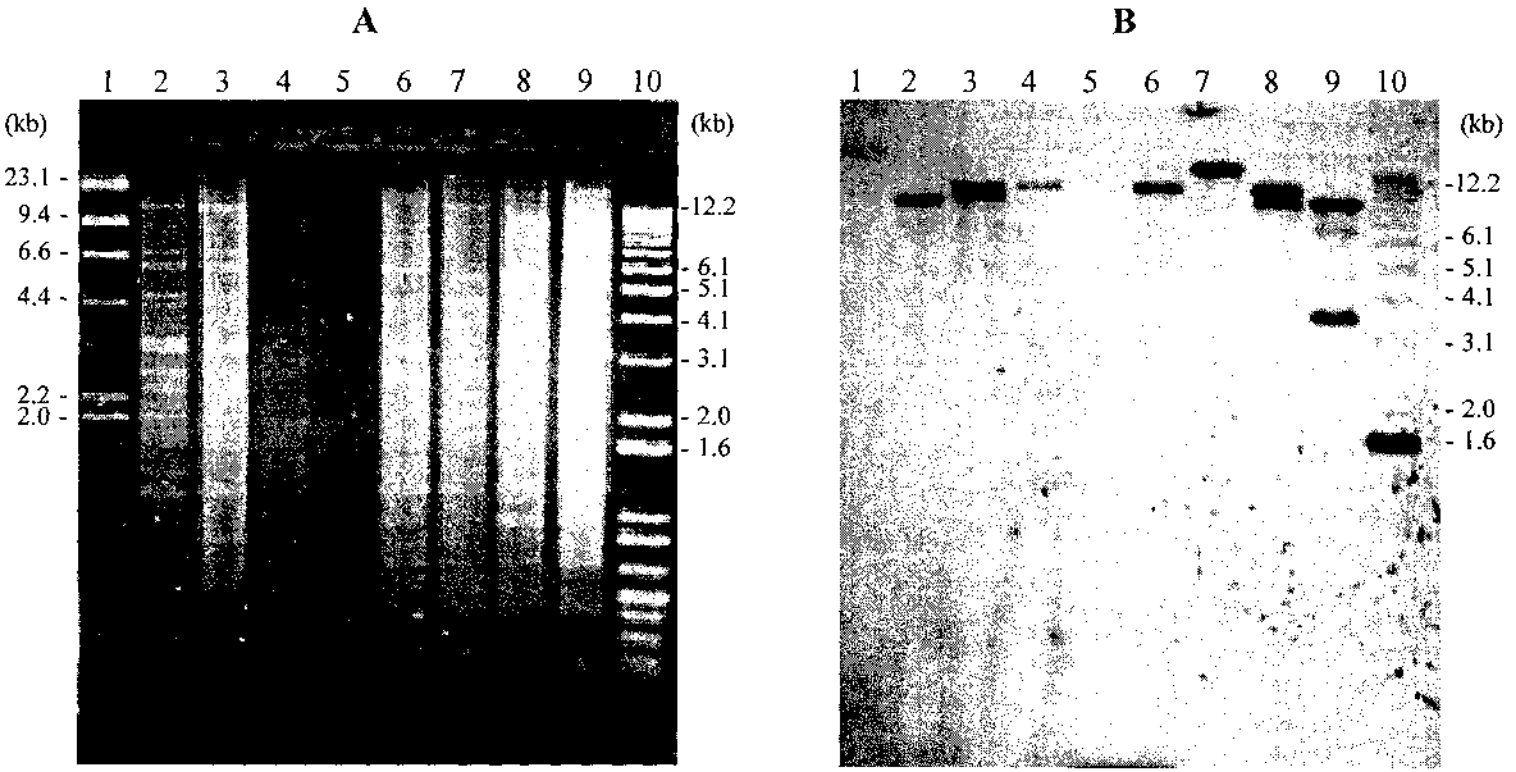


Figure 3.2 Southern analysis endophyte genomic DNA homologous to the pGH30 insert

A. Southern blot of 1 µg of Lp19 (lane 2), Lp1 (lane 3), E8 (lane 4), AR510 (lane 6), Fp5 (lane 7), Tf2 (lane 8) and Tf13 (lane 9) genomic DNA digested with *Hin*dIII. Lane 1 contains λ DNA digested with *Hin*dIII, and lane 10 contains 1kb + ladder (GibcoBRL). No DNA was loaded in lane 5.

B. Autoradiograph of this blot probed with [α^{32}P]dCTP-labelled pGH30 insert.

Table 3.2 Sizes of *Hin*dIII fragments homologous to the pGH30 insert.

	***Hin*dIII digest**
Lp19 (*N. lolii*)	10.0 kb
Lp1 (LpTG-2)	11.5 kb, 10.6 kb
E8 (*E. typhina*)	11.5 kb
AR510 (FaTG-3)	11.5 kb
Fp5 (*N. uncinatum*)	19.9 kb
Tf2 (*N. coenophialum*)	11.5 kb, 10.0 kb
Tf13 (FaTG-2)	10.0 kb, 6.6 kb, 3.5 kb

Fragment sizes are shown in kb.

triplicate lifts were taken from each plate. One set of lifts was probed with the pGH30 insert, the second set with the pGH3 insert (Section 4.2), and the third set was screened with both probes. Some of the λ clones that hybridised to the pGH30 insert on duplicate lifts were taken through two more rounds of library screening. Four positive λ clones, λMM30.1, λMM30.2, λMM30.3 andλMM30.4, were selected for further analysis.

3.2.2 Mapping of *prt*1 positive clones

3.2.2.1 Restriction digestion of positive λ clones

λ DNA was digested with *Eco*RI, *Hin*dIII, or with both *Eco*RI and *Hin*dIII (Figure 3.3A). These enzymes were selected because *Eco*RI cuts within the λGEM-12 multicloning cassette, while *Hin*dIII cuts within the λ arms. Fragment sizes are shown in Table 3.3. All four of the positive λ clones appeared to share a 1.8 kb *Eco*RI fragment (lanes 2, 5, 8 and 11) that is also present in the *Eco*RI/*Hin*dIII digests (lanes 3, 6, 9 and 12). This suggests that all of the clones overlap. Both λMM30.1 and λMM30.2 contained an insert of approximately 15.8 kb, and had identically sized bands for each restriction digest. The sum of the insert fragments from λMM30.3 and λMM30.4 revealed an insert size of approximately 12 kb.

When a Southern blot of the λ clones was probed with the probe used in Section 3.1 λMM30.1 and λMM30.2 both contained a hybridising band of 2 kb (Figure 3.3B, lanes 2 and 5). The same size bands also hybridised in the *Eco*RI/*Hin*dIII digest (lanes 3 and 6). For the *Hin*dIII digest a band of at least 23 kb hybridised to the probe (lanes 4 and 7). The presence of a 2.0 kb *Eco*RI fragment in λMM30.1 (lane 2) and λMM30.2 (lane 5) compared to the 3.3 kb *Eco*RI band from the Lp19 genomic blot (Figure 3.1B) suggests these λ clones contain a truncated hybridising fragment. The same sized fragment also hybridised in *Eco*RI/*Hin*dIII digests in lanes 3 and 6.

The Southern blot of λMM30.3 and λMM30.4 showed they both contained 3.3 kb *Eco*RI hybridising bands (Figure 3.3B, lanes 8 and 11) identical in size to the genomic band (Figure 3.1B). The same sized bands hybridised in the *Eco*RI/*Hin*dIII digests (Figure 3.3B, lanes 9 and 12). The *Hin*dIII fragment hybridising band was at least 23 kb in size.

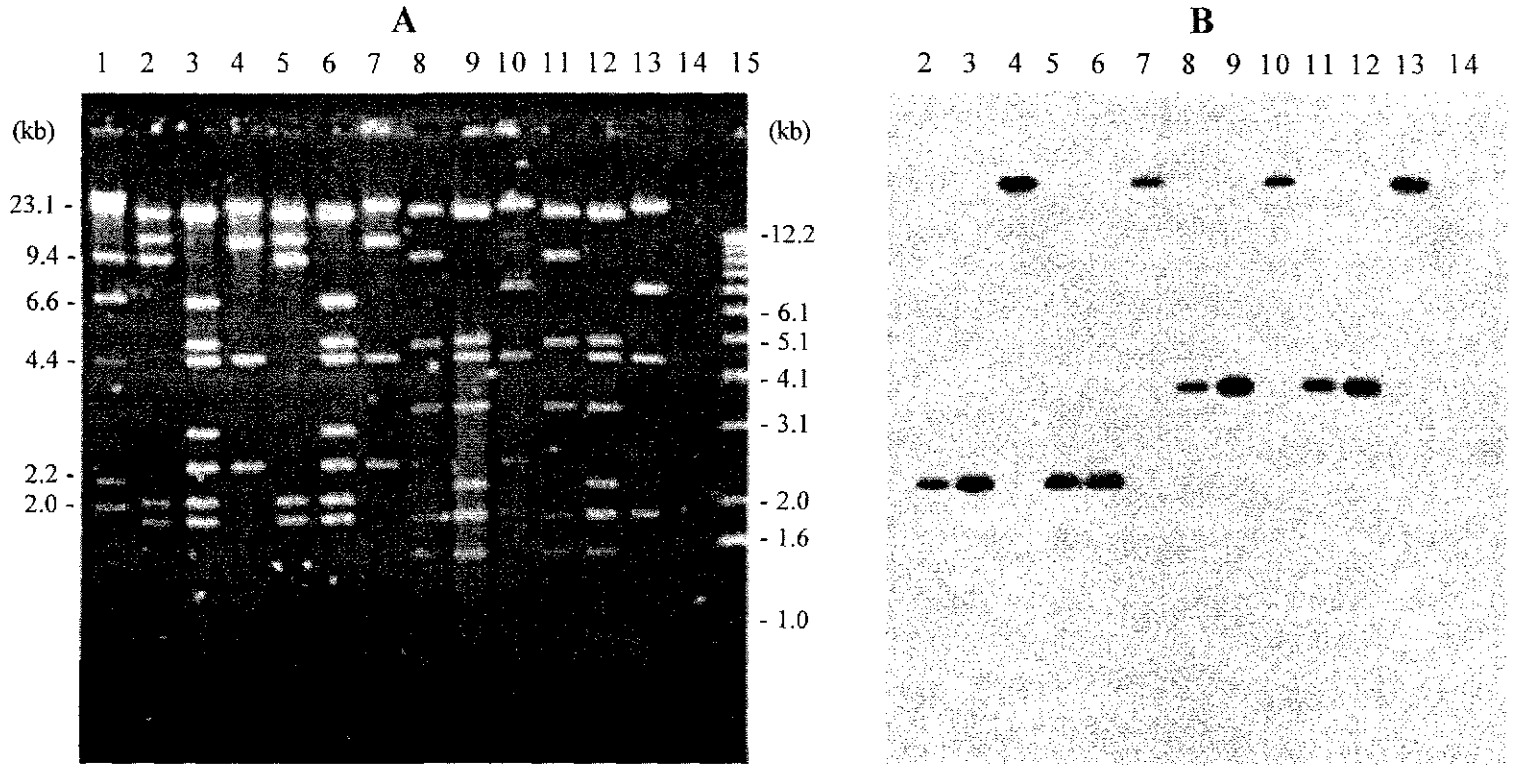


Figure 3.3 Southern analysis of *prt*1 positive λ clones and fragments homologous to the pGH30 insert.

A: Southern blot of *Eco*RI, *Eco*RI/*Hin*dIII and *Hin*dIII digested λ clones. Clones λMM30.1 (lanes 2-4), λMM30.2 (lanes 5-7), λMM30.3 (lanes 8-10), λMM30.4 (lanes 11-13). Lane 1 contains λ DNA digested with *Hin*dIII; lane 15 contains 1 kb ladder (GibcoBRL).

B: Autoradiograph of the Southern blot shown in Figure 3.3A probed with [α^{32}P]dCTP-labelled pGH30 insert.

Table 3.3 **Restriction fragments generated by digestion of *prt*1 positive λ clones.**

	EcoRI	*EcoRI/Hin*dIII	*Hin*dIII
λMM30.1/ λMM30.2	20.0 11.9 9.2	20.0	20.0* 11.1
		6.5	
		4.8 4.4	4.4
		2.9 2.5	2.5
	2.0* 1.8	2.0* 1.8	
λMM30.3/ λMM30.4	20.0	20.0	23.1*
	9.2		7.3
	5.0	5.0 4.6	4.7
	3.3*	3.3*	
		2.3	2.0
	1.8 1.4	1.8 1.4	

Fragment sizes are shown in kb. Hybridising fragments are indicated with an asterisk.

3.2.2.2 Restriction mapping of λMM30.2 and λMM30.4

The selected λ clones λMM30.2 (Figure 3.4) and λMM30.4 (Figure 3.5) were digested by the restriction enzymes *Bam*HI, *Eco*RI, *Hin*dIII and *Sst*I or in combination with *Eco*RI. From Southern analysis and restriction mapping, it was known these λ clones share 1.8 kb *Eco*RI fragments as well as 2 kb of the 3.3 kb *Eco*RI fragments. The data from the restriction digests was used to develop a restriction map for λMM30.2 and λMM30.4. The band sizes are summarised in Tables 3.4 and 3.5.

Southern analysis with the probe used in Section 3.1 showed that λMM30.2 had an 8.4 kb *Bam*HI hybridising fragment (Figure 3.4B, lane 2) . A common 2 kb fragment was seen in the *Bam*HI/*Eco*RI, *Eco*RI, *Eco*RI/*Hin*dIII and *Eco*RI/*Sst*I digests (lanes 3, 4, 5 and 7 respectively). The *Hin*dIII hybridising fragment was at least 23 kb in size. Unfortunately it was hard to analyse the *Sst*I digest (Figure 3.4A, lane 8) because of a lack of DNA loaded on the gel, so other gels were examined to determine the sizes of *Sst*I fragments (lane 11). Southern analysis of λMM30.4 showed a single 13 kb *Bam*HI fragment hybridised to the pGH30 insert (Figure 3.5, lane2). A common 3.3 kb hybridising band was observed in the *Bam*HI/*Eco*RI, *Eco*RI, *Eco*RI/*Hin*dIII and *Eco*RI/*Sst*I digests (Figure 3.5B, lanes3, 4, 5 and 7 respectively). For the *Hin*dIII digest (lane 6) a large fragment hybridised. In the *Sst*I digest (lane 8) a 9 kb fragment hybridised to the probe.

The data generated by Southern analysis and restriction mapping was used to determine the overall organisation of the λ clones (Figure 3.6). Due to a lack of detailed restriction mapping the organisation of λMM30.2 is still ambiguous, but it appears that λMM30.2 and λMM30.4 contain an overlapping region of about 4.8 kb.

3.2.3 Shot gun cloning of fragments from λMM30.4

Products from an *Eco*RI digest of λMM30.4 were subcloned into pUC19 (Appendix A1.5, Figure 3.7B). The resulting plasmids were analysed by restriction mapping to determine what sites were in each of the *Eco*RI fragments aiding in the restriction mapping of λMM30.4 (restriction maps are shown in Figure 3.7B). pMM1 contained a 5 kb *Eco*RI fragment (lanes 6-8, 12), pMM3 contained a 1.8 kb *Eco*RI band (lane 11) and pMM4 contained a 1 kb *Eco*RI fragment (lanes 10 and 11). pMM2 (lanes 5, 13 and 14), which contained the 3.3 kb *Eco*RI hybridising fragment, was initially chosen for sequencing.

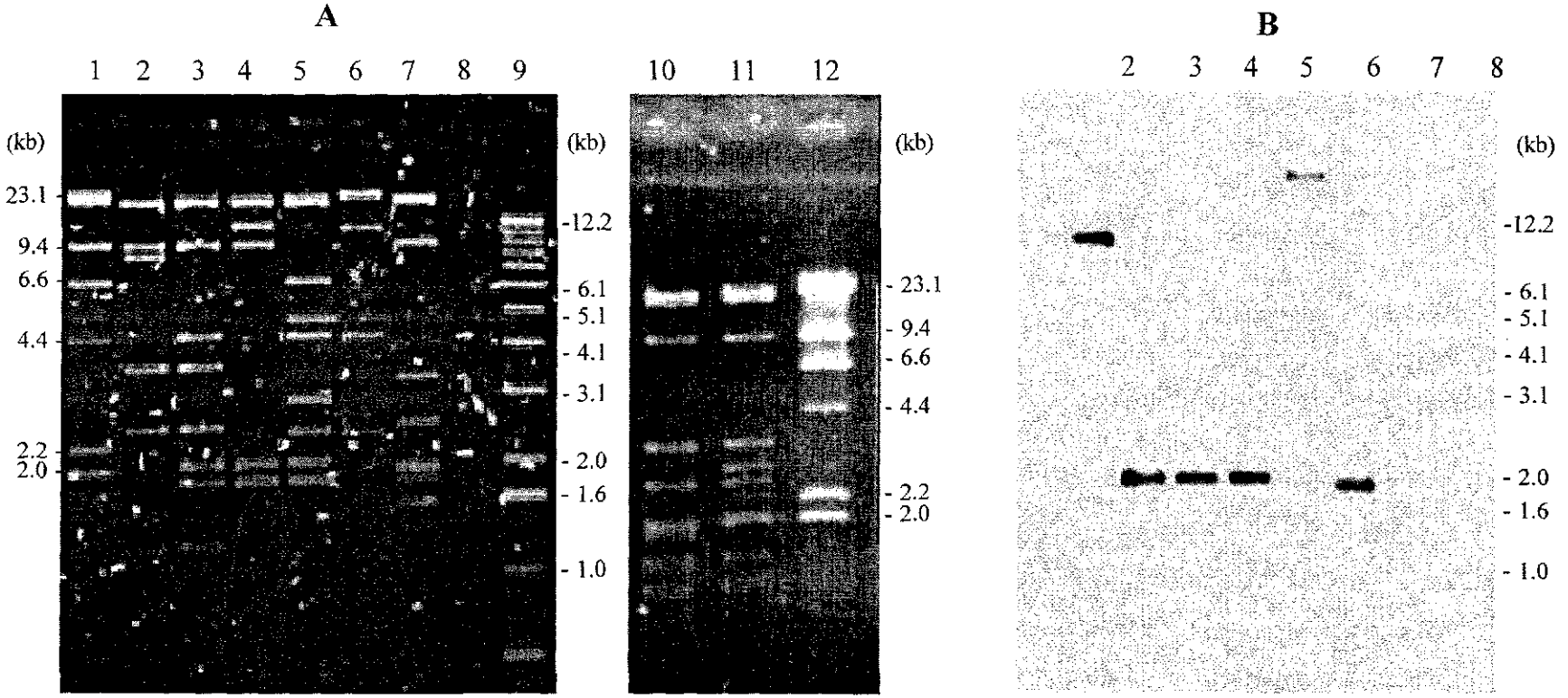


Figure 3.4 Restriction mapping of λMM30.2 and fragments homologous to pGH30.

A: Southern blot of λMM30.2 digested with *Bam*HI (lane 2), *Bam*HI/*Eco*RI (lane 3), *Eco*RI (lane 4), *Eco*RI/*Hin*dIII (lane 5), *Hin*dIII (lane 6), *Eco*RI/*Sst*I (lane 7) and *Sst*I (lane 8). Lane 10 contains the *Eco*RI/*Sst*I digest seen in lane 7, lane 11 contains the *Sst*I digest not visible in lane 8. Lanes 1 and 12 contain λ DNA digested with *Hin*dIII; lane 9 contains 1 kb ladder (GibcoBRL).

B: Autoradiograph of the Southern blot shown in Figure 3.4A probed with [α^{32}P]dCTP-labelled pGH30 insert.

Table 3.4 Fragment sizes of λMM30.2 restriction digests

*Bam*HI	*Bam*HI/ *Eco*RI	*Eco*RI	*Eco*RI/ *Hin*dIII	*Hin*dIII	*Eco*RI/ *Sst*I	*Sst*I
20.7	20.7	20.7	20.7	>23.0	20.7	20.7
		12.0		11.0		
9.4	9.4	9.4			9.4	9.4
8.4*			6.7			
			4.9			
	4.2		4.4	4.4		
3.6	3.6				3.4	3.4
			2.9			2.8
2.5	2.6		2.5	2.5	2.6	2.6
	2.0*	2.0*	2.0*		2.0*	2.0
	1.8	1.8	1.8		1.8	
					1.6	1.6
1.1	1.1				1.2	1.2

Fragment sizes are shown in kb. Bands that hybridise to the pGH30 insert are indicated with an asterisk.

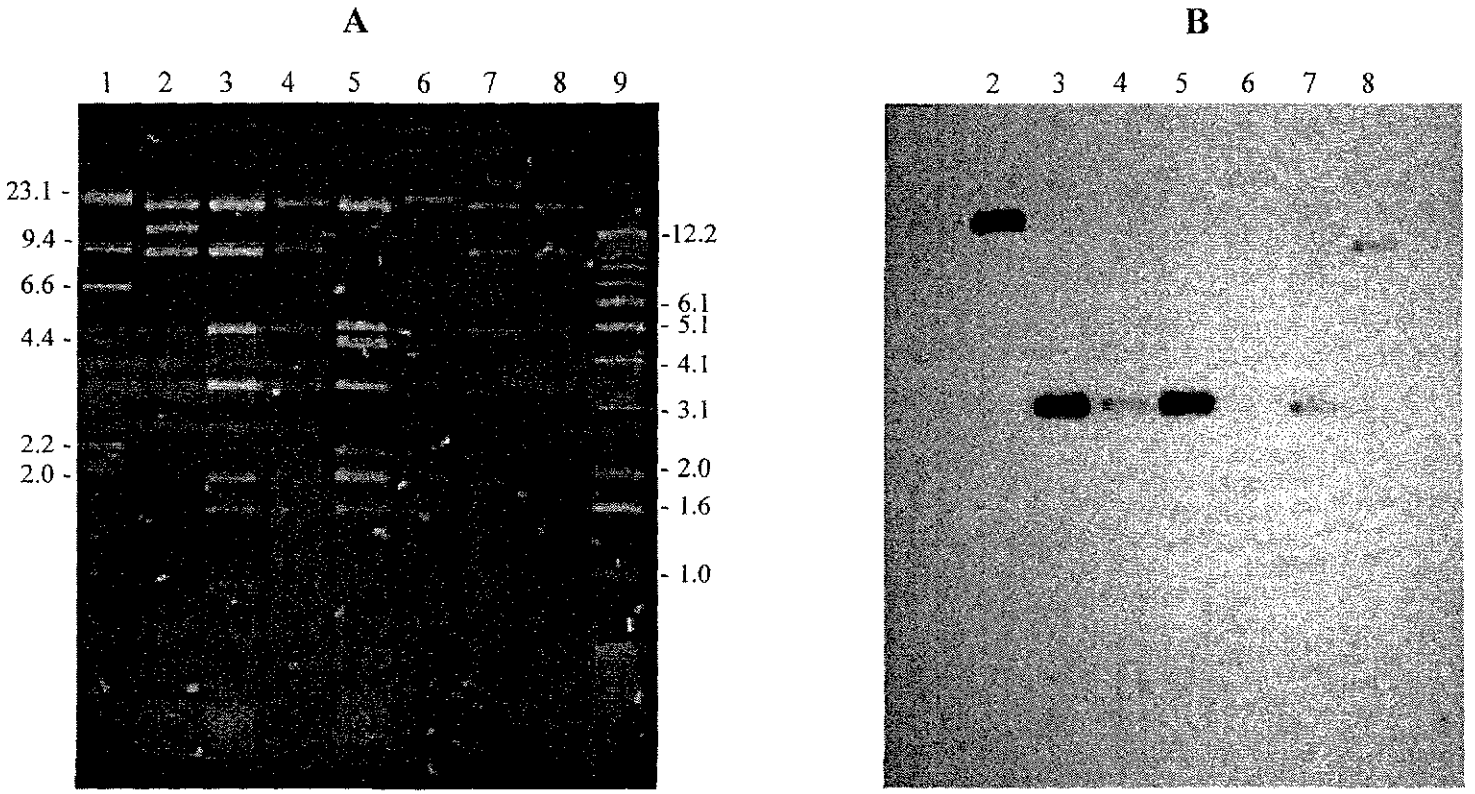


Figure 3.5 Restriction mapping of λMM30.4 and fragments homologous to the pGH30 insert.

A: Southern blot of λMM30.4 digested with *Bam*HI (lane 2), *Bam*HI/*Eco*RI (lane 3), *Eco*RI (lane 4), *Eco*RI/*Hin*dIII (lane 5), *Hin*dIII (lane 6), *Eco*RI/*Sst*I (lane 7) and *Sst*I (lane 8). Lane 1 contains λ DNA digested with *Hin*dIII; lane 9 contains 1 kb ladder (GibcoBRL).

B: Autoradiograph of the Southern blot in Figure 3.5A probed with [α^{32}P]dCTP-labelled pGH30 insert.

Table 3.5 Fragment sizes of λMM30.4 restriction digests.

*Bam*HI	*Bam*HI/ *Eco*RI	*Eco*RI	*Eco*RI/ *Hin*dIII	*Hin*dIII	*Eco*RI/ *Sst*I	*Sst*I
20.0	20.0	20.0	20.0	>23.0*	20.0	20.0
13.0*						
9.0	9.0	9.0			9.0	9.0*
				7.2	5.5	
	5.0	5.0	5.2	4.7		
			4.4			
	3.3*	3.3*	3.3*		3.3*	2.8
			2.3			
	2.0	2.0	2.0		2.0	
	1.8	1.8	1.8			
	1.0	1.0				

Fragment sizes are shown in kb. Bands that hybridise to the pGH30 insert are indicated with an asterisk.

Figure 3.6 Overall organisation of λ clones and sequencing strategy for the fragment homologous to the pGH30 insert.

A. Restriction maps for the overlapping λ clones λMM30.2 and λMM30.4 are shown in blue. The position of the pGH30 insert sequence relative to the λ clones is shown in orange.

B. The lower part of this figure shows the strategy used to sequence this gene. Sites where primers annealed within the 3.3 kb *Eco*RI fragment in pMM2 are shown with red arrows, with the primer names in black text. Restriction sites are also indicated in purple text. The open reading frames for the *prt*1 gene are shown with the green arrows.

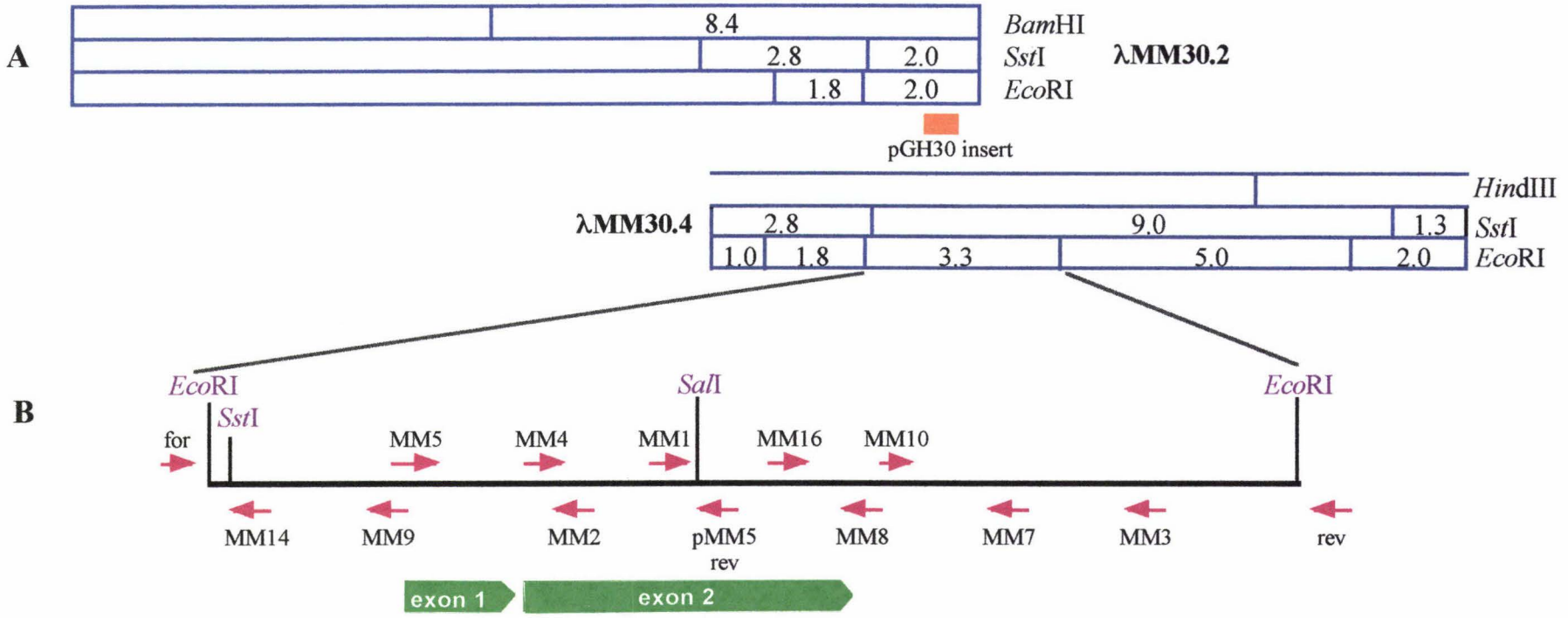


Figure 3.7 Subcloning of fragments from λMM30.4

A. Shot gun cloning of *Eco*RI fragments from λMM30.4

This gel contains: *Eco*RI-digested λMM30.4 (lane 2), *Eco*RI-digested pUC19 (lane 3) and rapid boil DNA from λMM30.4 subclones digested with *Eco*RI (lanes 4-14). Lane 1 contains λ DNA digested with *Hin*dIII, while lane 15 contains 1 kb ladder (Gibco BRL).

Lanes 6, 7, 8 and 12 contain pMM1. Lane 5, 13 and 14 contain pMM2; lane 11 contains pMM3, and lanes 9 and 10 contain pMM4.

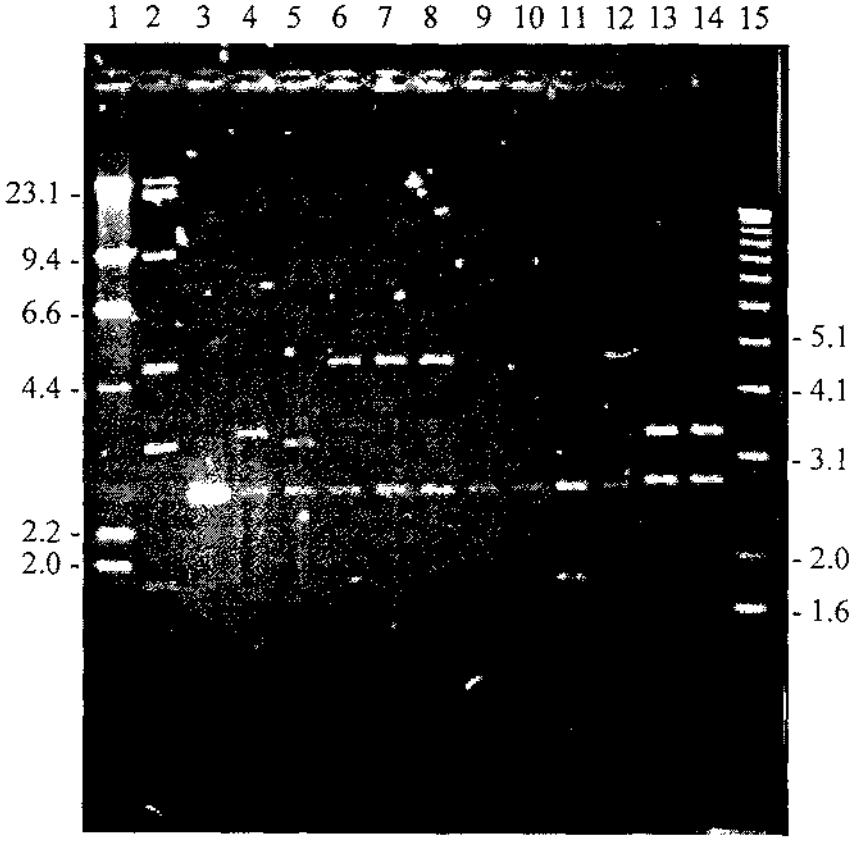
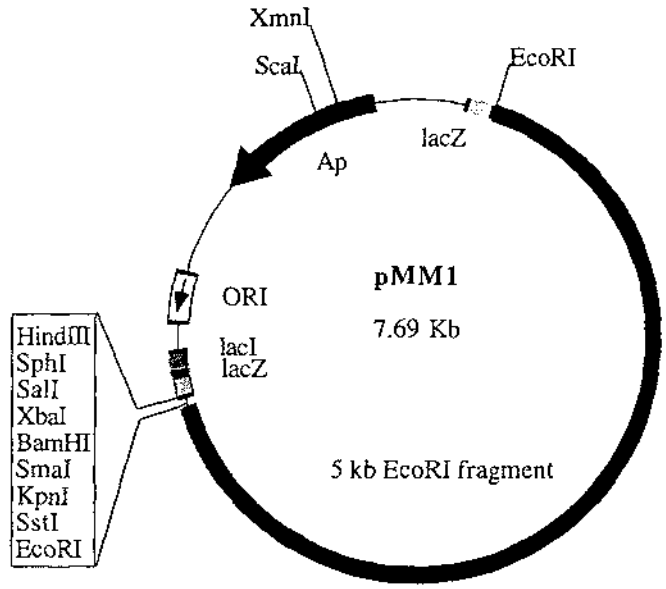


Figure 3.7 Subcloning of fragments from λMM30.4

B. Restriction maps of plasmids from shotgun cloning of λMM30.4

These plasmids were generated by shotgun cloning of *Eco*RI fragments from λMM30.4 into pUC19 (Appendix 1.5).



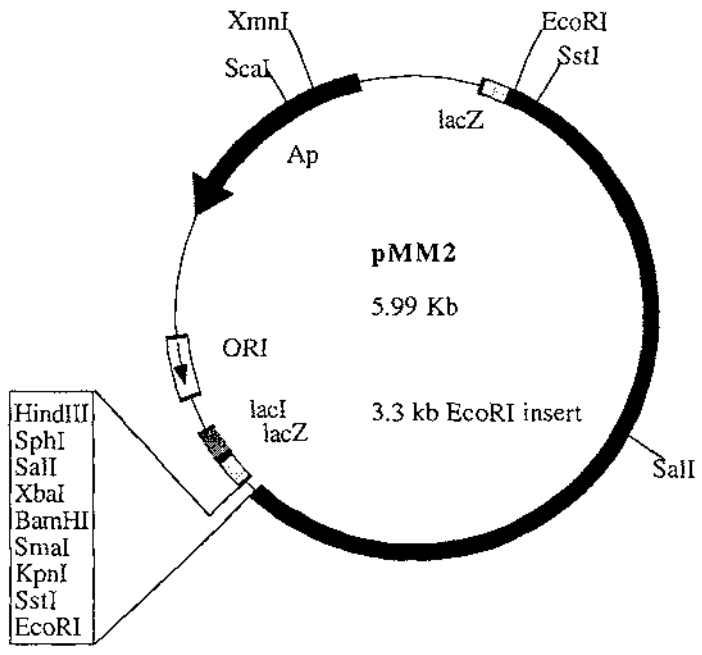
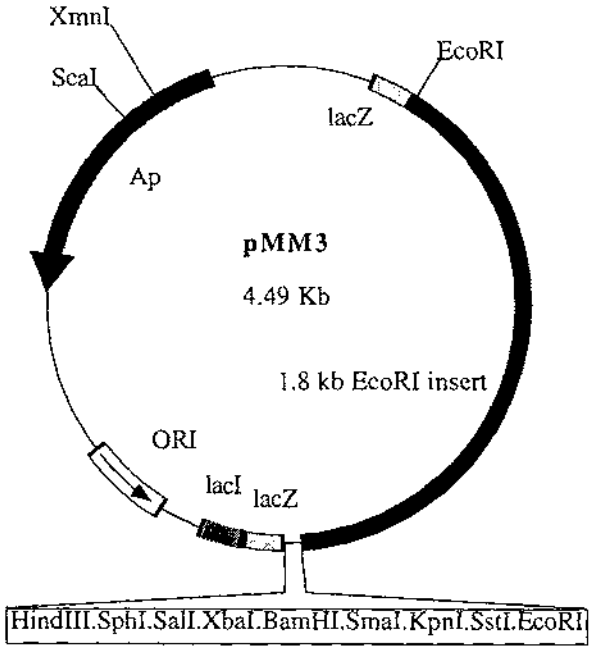


Figure 3.7 Subcloning of fragments from λMM30.4

B. Restriction maps of plasmids from shotgun cloning of λMM30.4 cont.

These plasmids were generated by shotgun cloning of *Eco*RI fragments from λMM30.4 into pUC19 (Appendix 1.5).



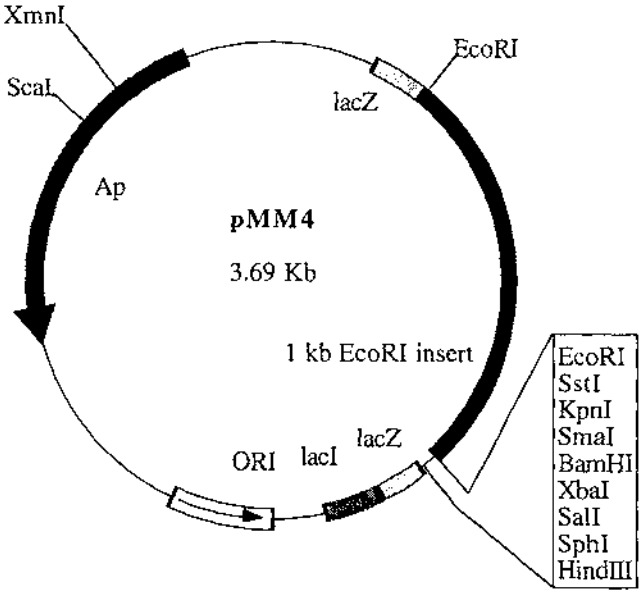
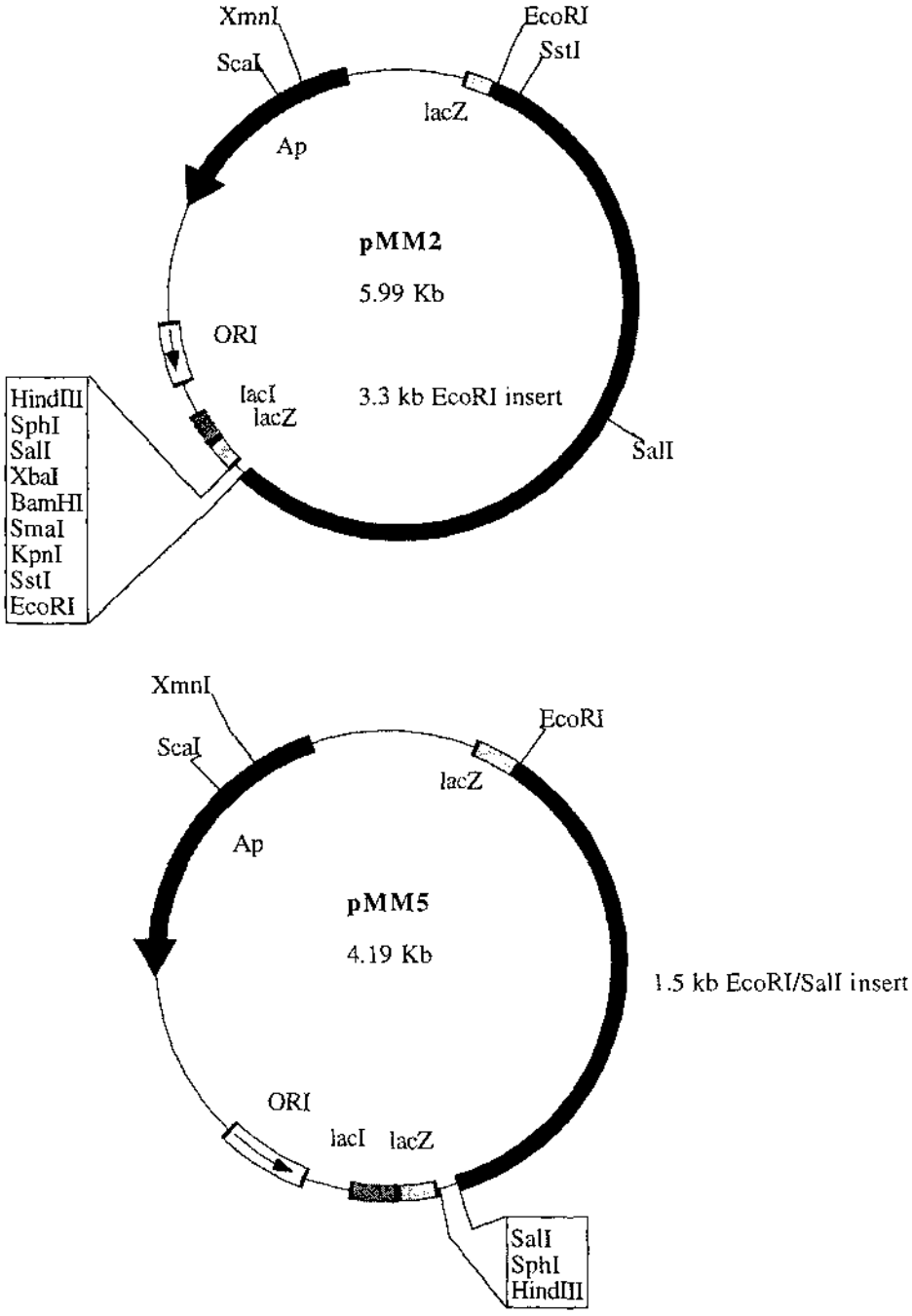


Figure 3.7 Subcloning of fragments from λMM30.4

C. Restriction map of the pMM2/*Sal*I dropout vector pMM5

The pMM5 dropout vector was constructed by cutting out a 1.8 kb *Sal*I restriction fragment from the pMM2 vector and religating the vector to itself.



3.3 SEQUENCE ANALYSIS

3.3.1 Construction of a drop out vector from pMM2

Restriction mapping of pMM2 revealed a *Sal*I site in middle of the 3.3 kb *Eco*RI insert (Section 3.2.3). A *Sal*I site was also found in the middle of the pGH30 insert so it was probable that this *Sal*I site was within the sequence homologous to the probe. Digestion with *Sal*I resulted in a 1.8 kb *Sal*I fragment being lost from pMM2. After religation the new vector contained a 1.5 kb *Eco*RI/*Sal*I fragment (Figure 3.7C). The drop out vector, known as pMM5, allowed sequencing from a region of pMM2 that hybridises to the pGH30 probe outwards into sequences downstream of the gene.

3.3.2 Sequencing

The pMM2 vector was initially sequenced with M13 forward and reverse primers using Big Dye chemistry (Section 2.15). Blast-X searches of these sequences revealed no homology to protease genes. When pMM5 (Section 3.3.1) was sequenced with the reverse primer, homology to other proteases was detected. The sequence was extended using new primers until double stranded sequence was obtained for the whole *prt*1 gene (Figures 3.6B and 3.8). The Wisconsisn Package 9.1 (GCG) and Sequencher™ packages were used to assemble the sequences in preparation for further analysis (Section 2.15.2). Sequence data is shown in Appendix A3.1.

3.3.3 Analysis of the *prt*1 promoter region

The *prt*1 promoter region (Figure 3.8, bases –623 to –100) contained several potential transcription factor-binding sites. Of particular interest are the GATA sites, which are often involved in nitrogen regulation in filamentous fungi. Two GATA sites, at –375 and -351, and four GATT sites, at –547, -537, -287 and -180, are found in the 5' region (Figure 3.8). There is also one binding site consensus sequence (GTGGAG) at -587 for CreA, which is involved in carbon catabolite repression. A PacC binding consensus sequence, GCCAAG, is present at –530. PacC is a global regulator of pH response in filamentous fungi (Tilburn et al., 1995).

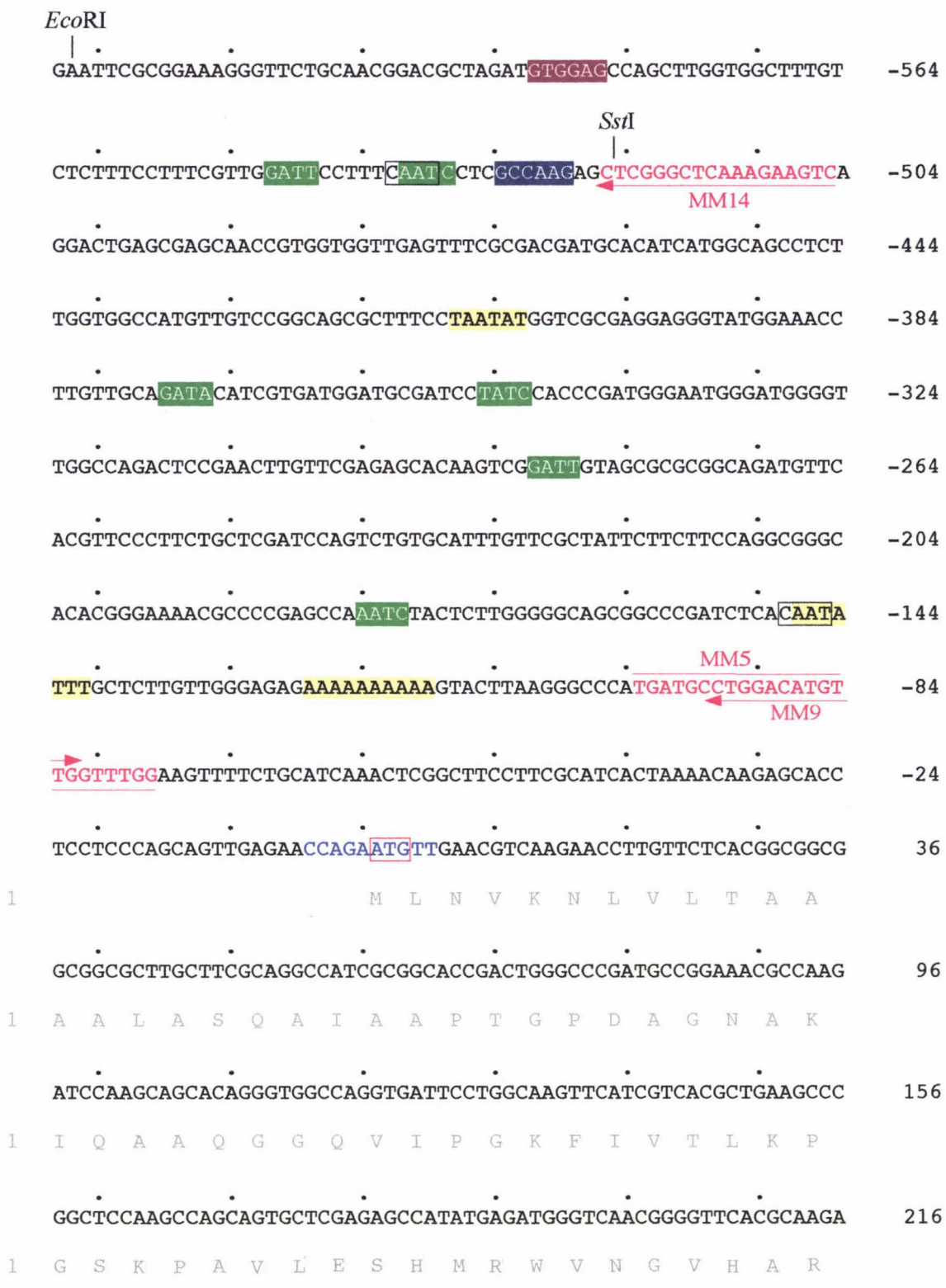
There is no strong match for a TATA box in the *prt*1 promoter region, but there are several AT-rich regions that may have a similar function. The sequence AATATTT is found at -141

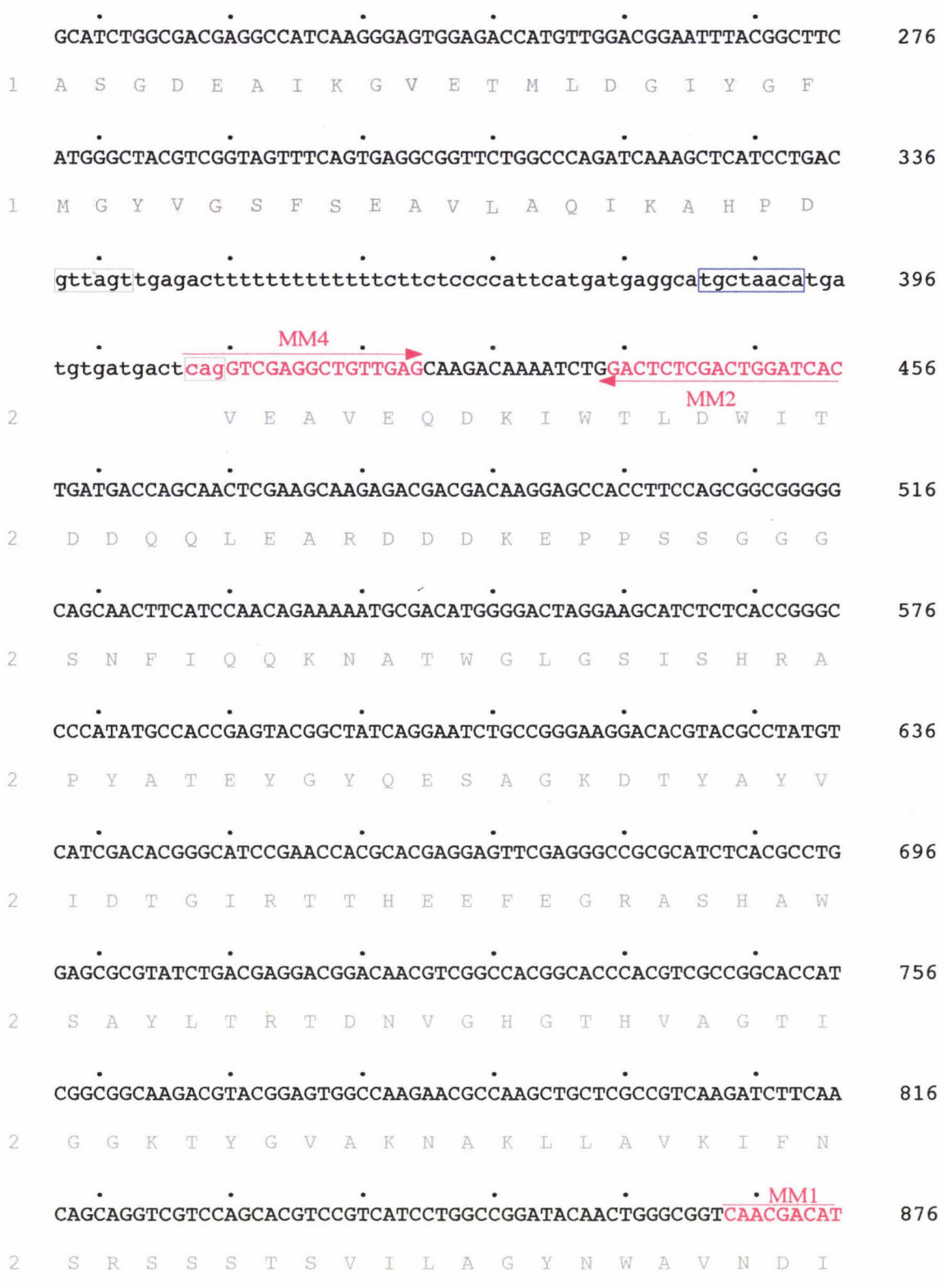
Figure 3.8 Nucleotide sequence of the *prt*1 gene.

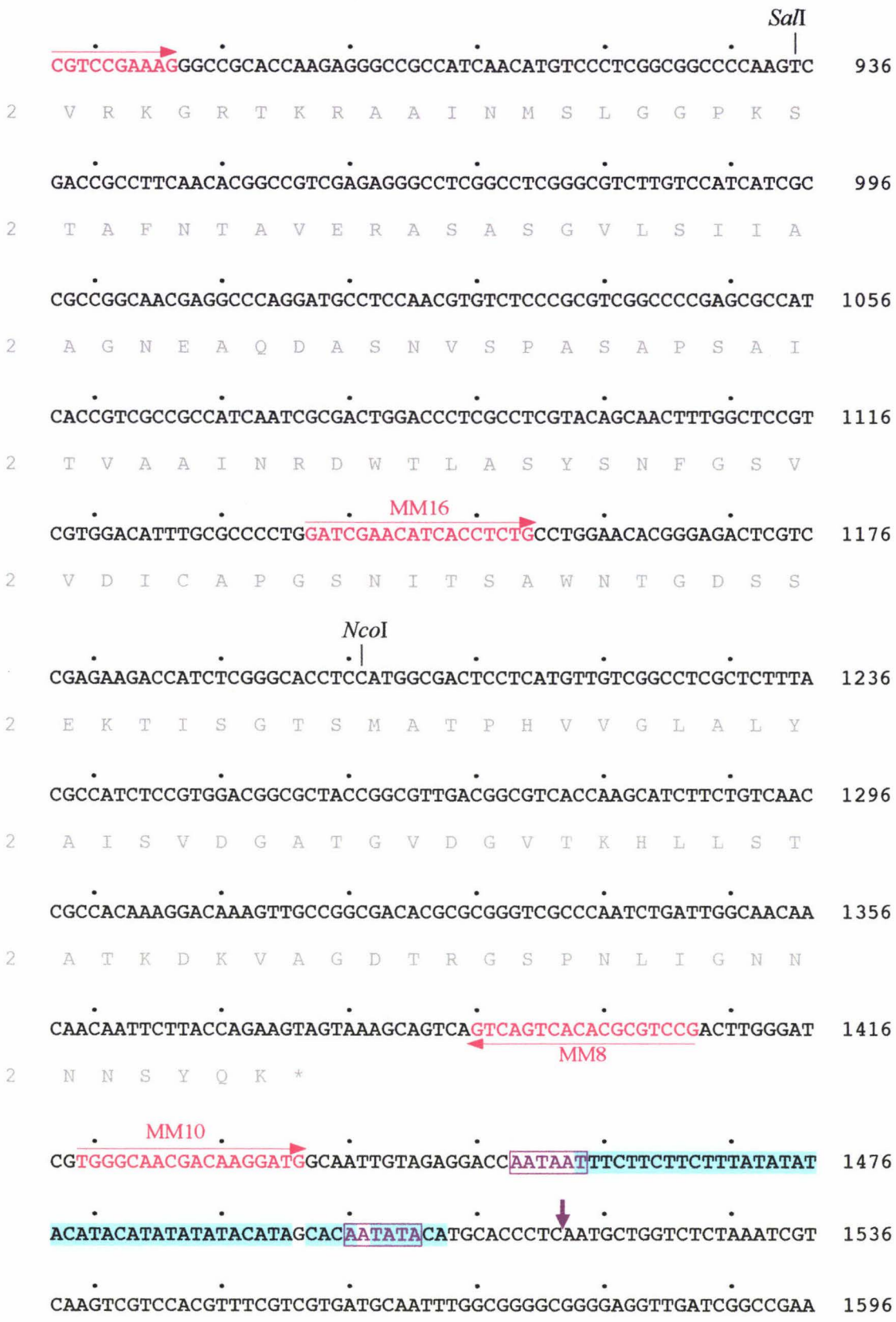
The 3300 base pair *Eco*RI fragment from pMM2 contains two open reading frames. Within the promoter region, AreA binding sites (GATA and GATT) are highlighted with green shading, CreA binding sites (SYGGRG) are highlighted in dark red shading and PacC binding sites (GCCARG) are shown in navy shading. Putative CAAT boxes are outlined in black, and AT-rich regions which could act like TATA boxes are highlighted with yellow shading.

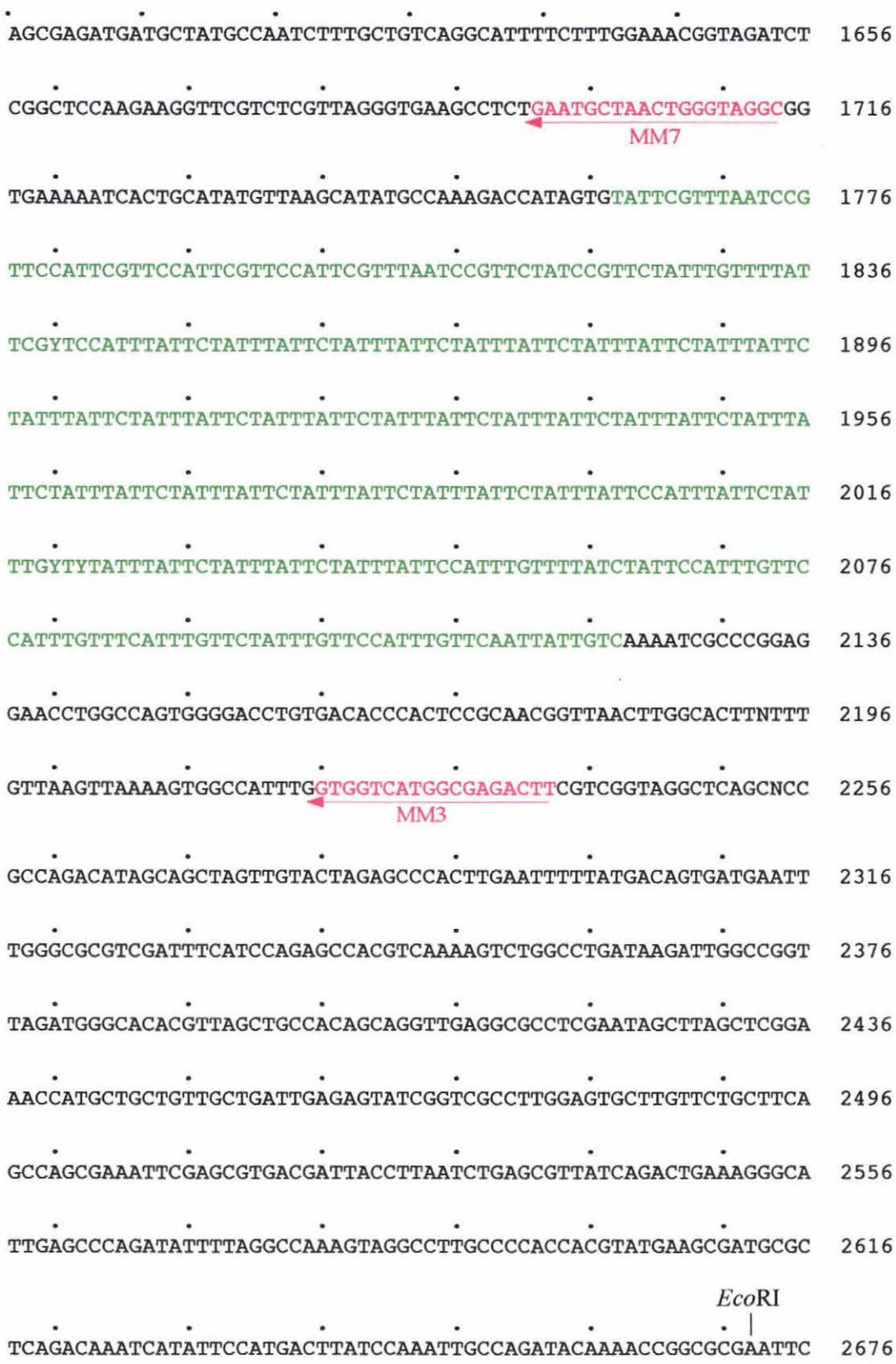
The sequence is numbered relative to the translation start, where A of the start codon ATG (outlined in a red box) is equal to position 1. A good match for the Kozak consensus sequence, which surrounds the translation start sequence, is shown in dark blue text. The translation, which was deduced by homology with peptide sequences from other fungal subtilisin-like protease, is shown in grey text directly below the nucleotide sequence. The reading frame number is indicated at the beginning of each line of the translation. The intron, at 337 to 409, is shown in lower case letters. The potential 5' splice and 3' splice sites are outlined in grey boxes. The lariat sequence is outlined in a dark blue box.

Primers used for sequencing are indicated in red text, with the accompanying red arrow showing the direction of extension from the primer. The two potential polyadenylation recognition sequences at 1453-1458 and 1500-1505 are shown in purple text and outlined by a purple box. The polyadenylation site determined by 3' RACE is indicated by a purple arrow. The complex microsatellite in the 3' UTR is highlighted with light blue shading. The minisatellite downstream of the *prt*1 gene is shown in green text.


EcoRI
GAATTCGCGGAAAGGGTTCTGCAACGGACGCTAGATGTGGAGCCAGCTTGGTGGCTTTGT -564
SstI
CTCTTTCCTTTCGTTGGATTCCTTTCAATCCTCGCCAAGAGCTCGGGCTCAAAGAAGTCA -504
MM14
GGACTGAGCGAGCAACCGTGGTGGTTGAGTTTCGCGACGATGCACATCATGGCAGCCTCT -444
TGGTGGCCATGTTGTCCGGCAGCGCTTTCCTAATATGGTCGCGAGGAGGGTATGGAAACC -384
TTGTTGCAGATACATCGTGATGGATGCGATCCTATCCACCCGATGGGAATGGGATGGGGT -324
TGGCCAGACTCCGAACTTGTTCGAGAGCACAAGTCGGATTGTAGCGCGCGGCAGATGTTC -264
ACGTTCCCTTCTGCTCGATCCAGTCTGTGCATTTGTTCGCTATTCTTCTTCCAGGCGGGC -204
ACACGGGAAAACGCCCCGAGCCAAATCTACTCTTGGGGGCAGCGGCCCGATCTCACAATA -144
MM5
TTTGCTCTTGTTGGGAGAGAAAAAAAAAAGTACTTAAGGGCCCATGATGCCTGGACATGT -84
MM9
TGGTTTGGAAGTTTTCTGCATCAAACTCGGCTTCCTTCGCATCACTAAAACAAGAGCACC -24
TCCTCCCAGCAGTTGAGAACCAGAATGTTGAACGTCAAGAACCTTGTTCTCACGGCGGCG 36
1 M L N V K N L V L T A A
GCGGCGCTTGCTTCGCAGGCCATCGCGGCACCGACTGGGCCCGATGCCGGAAACGCCAAG 96
1 A A L A S Q A I A A P T G P D A G N A K
ATCCAAGCAGCACAGGGTGGCCAGGTGATTCCTGGCAAGTTCATCGTCACGCTGAAGCCC 156
1 I Q A A Q G G Q V I P G K F I V T L K P
GGCTCCAAGCCAGCAGTGCTCGAGAGCCATATGAGATGGGTCAACGGGGTTCACGCAAGA 216
1 G S K P A V L E S H M R W V N G V H A R


GCATCTGGCGACGAGGCCATCAAGGGAGTGGAGACCATGTTGGACGGAATTTACGGCTTC 276
1 A S G D E A I K G V E T M L D G I Y G F
ATGGGCTACGTCGGTAGTTTCAGTGAGGCGGTTCTGGCCCAGATCAAAGCTCATCCTGAC 336
1 M G Y V G S F S E A V L A Q I K A H P D
gttagttgagacttttttttttttcttctccccattcatgatgaggcatgctaacatga 396
MM4
tgtgatgactcagGTCGAGGCTGTTGAGCAAGACAAAATCTGGACTCTCGACTGGATCAC 456
MM2
2 V E A V E Q D K I W T L D W I T
TGATGACCAGCAACTCGAAGCAAGAGACGACGACAAGGAGCCACCTTCCAGCGGCGGGGG 516
2 D D Q Q L E A R D D D K E P P S S G G G
CAGCAACTTCATCCAACAGAAAAATGCGACATGGGGACTAGGAAGCATCTCTCACCGGGC 576
2 S N F I Q Q K N A T W G L G S I S H R A
CCCATATGCCACCGAGTACGGCTATCAGGAATCTGCCGGGAAGGACACGTACGCCTATGT 636
2 P Y A T E Y G Y Q E S A G K D T Y A Y V
CATCGACACGGGCATCCGAACCACGCACGAGGAGTTCGAGGGCCGCGCATCTCACGCCTG 696
2 I D T G I R T T H E E F E G R A S H A W
GAGCGCGTATCTGACGAGGACGGACAACGTCGGCCACGGCACCCACGTCGCCGGCACCAT 756
2 S A Y L T R T D N V G H G T H V A G T I
CGGCGGCAAGACGTACGGAGTGGCCAAGAACGCCAAGCTGCTCGCCGTCAAGATCTTCAA 816
2 G G K T Y G V A K N A K L L A V K I F N
MM1
CAGCAGGTCGTCCAGCACGTCCGTCATCCTGGCCGGATACAACTGGGCGGTCAACGACAT 876
2 S R S S S T S V I L A G Y N W A V N D I


SalI
CGTCCGAAAGGGCCGCACCAAGAGGGCCGCCATCAACATGTCCCTCGGCGGCCCCAAGTC 936
2 V R K G R T K R A A I N M S L G G P K S
GACCGCCTTCAACACGGCCGTCGAGAGGGCCTCGGCCTCGGGCGTCTTGTCCATCATCGC 996
2 T A F N T A V E R A S A S G V L S I I A
CGCCGGCAACGAGGCCCAGGATGCCTCCAACGTGTCTCCCGCGTCGGCCCCGAGCGCCAT 1056
2 A G N E A Q D A S N V S P A S A P S A I
CACCGTCGCCGCCATCAATCGCGACTGGACCCTCGCCTCGTACAGCAACTTTGGCTCCGT 1116
2 T V A A I N R D W T L A S Y S N F G S V
MM16
CGTGGACATTTGCGCCCCTGGATCGAACATCACCTCTGCCTGGAACACGGGAGACTCGTC 1176
2 V D I C A P G S N I T S A W N T G D S S
NcoI
CGAGAAGACCATCTCGGGCACCTCCATGGCGACTCCTCATGTTGTCGGCCTCGCTCTTTA 1236
2 E K T I S G T S M A T P H V V G L A L Y
CGCCATCTCCGTGGACGGCGCTACCGGCGTTGACGGCGTCACCAAGCATCTTCTGTCAAC 1296
2 A I S V D G A T G V D G V T K H L L S T
CGCCACAAAGGACAAAGTTGCCGGCGACACGCGCGGGTCGCCCAATCTGATTGGCAACAA 1356
2 A T K D K V A G D T R G S P N L I G N N
CAACAATTCTTACCAGAAGTAGTAAAGCAGTCAGTCAGTCACACGCGTCCGACTTGGGAT 1416
MM8
2 N N S Y Q K *
MM10
CGTGGGCAACGACAAGGATGGCAATTGTAGAGGACCAATAATTTCTTCTTCTTTATATAT 1476
ACATACATATATATACATAGCACAATATACATGCACCCTCAATGCTGGTCTCTAAATCGT 1536
CAAGTCGTCCACGTTTCGTCGTGATGCAATTTGGCGGGGCGGGGAGGTTGATCGGCCGAA 1596


AGCGAGATGATGCTATGCCAATCTTTGCTGTCAGGCATTTTCTTTGGAAACGGTAGATCT 1656
CGGCTCCAAGAAGGTTCGTCTCGTTAGGGTGAAGCCTCTGAATGCTAACTGGGTAGGCGG 1716
MM7
TGAAAAATCACTGCATATGTTAAGCATATGCCAAAGACCATAGTGTATTCGTTTAATCCG 1776
TTCCATTCGTTCCATTCGTTCCATTCGTTTAATCCGTTCTATCCGTTCTATTTGTTTTAT 1836
TCGYTCCATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTC 1896
TATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTA 1956
TTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCCATTTATTCTAT 2016
TTGYTYTATTTATTCTATTTATTCTATTTATTCCATTTGTTTTATCTATTCCATTTGTTC 2076
CATTTGTTTCATTTGTTCTATTTGTTCCATTTGTTCAATTATTGTCAAAATCGCCCGGAG 2136
GAACCTGGCCAGTGGGGACCTGTGACACCCACTCCGCAACGGTTAACTTGGCACTTNTTT 2196
GTTAAGTTAAAAGTGGCCATTTGGTGGTCATGGCGAGACTTCGTCGGTAGGCTCAGCNCC 2256
MM3
GCCAGACATAGCAGCTAGTTGTACTAGAGCCCACTTGAATTTTTATGACAGTGATGAATT 2316
TGGGCGCGTCGATTTCATCCAGAGCCACGTCAAAAGTCTGGCCTGATAAGATTGGCCGGT 2376
TAGATGGGCACACGTTAGCTGCCACAGCAGGTTGAGGCGCCTCGAATAGCTTAGCTCGGA 2436
AACCATGCTGCTGTTGCTGATTGAGAGTATCGGTCGCCTTGGAGTGCTTGTTCTGCTTCA 2496
GCCAGCGAAATTCGAGCGTGACGATTACCTTAATCTGAGCGTTATCAGACTGAAAGGGCA 2556
TTGAGCCCAGATATTTTAGGCCAAAGTAGGCCTTGCCCCACCACGTATGAAGCGATGCGC 2616
*Eco*RI
TCAGACAAATCATATTCCATGACTTATCCAAATTGCCAGATACAAAACCGGCGCGAATTC 2676

followed by a run of ten adenines at -124, as well as an AT rich region TAATAT found at -413. There also CAAT sequences at -538 and -148 (Figure 3.8).

3.3.4 Proposed sites for transcription and translation initiation for *prt*1

The transcription initiation site of the *prt*1 gene is so far undefined. Based on RT-PCR results using the MM2 and MM5 primers (Figure 3.8), it appears to be upstream of the MM5 primer (at -100). Based on homology with the amino acid sequence of other fungal subtilisin-like proteases, the translation start site appears to be at position 1. The sequence around the start codon also shows close similarity to the Kozak consensus sequence (Gurr et al., 1987). Instead of the fungal Kozak consensus of CCACCATGGC, the *prt*1 gene has CCAGAATGTT.

3.3.5 Analysis of the *prt*1 coding region

The *prt*1 gene appears appears to contain one intron of 73 base pairs. Supporting evidence comes from RT-PCR studies (Section 3.6) and from homology with the peptide sequence of products of homologous genes. This intron is yet to be confirmed by sequencing of *prt*1 cDNA. There are four stop codons within the intron sequence and a frame change between the proposed coding sequences on either side of the intron. The proposed 5' splice site for the intron, GTTAGT (Figure 3.8, 337 to 342), is a good match for the consensus sequence GTANGT (Gurr et al., 1987).

An interesting feature of this intron is the 13 thymines in a row at 349-361, which is part of a pyrimidine-rich region. The *prt*1 intron contains the lariat sequence TGCTAACA (386 to 393) necessary for RNA splicing (Gurr et al., 1987). The proposed 3' splice site CAG (at 407 to 409) matches the fungal consensus of PyAG. The intron position is highly conserved with subtilisin like protease genes from other fungi.

The codon frequency of *prt*1 was analysed using the CODONFREQUENCY program from the Wisconsin Package 9.1, GCG (Table 3.6). There was a strong preference for cytosine at the third position of each codon. Where a purine was found at the third position, it is usually a guanine. This is a characteristic of highly expressed genes (Gurr et al., 1987). The codons GTA (Val), ATA (Ile), TGA (End), TGT (Cys), TAA (End), TTA (Leu), and CGT (Arg) are not used.

Table 3.6 Codon bias table for the *prt*1 gene

Amino acid	Codon	Number	Fraction	Amino acid	Codon	Number	Fraction
Gly	GGG	5.00	0.11	Trp	TGG	8.00	1.00
Gly	GGA	9.00	0.20	End	TGA	0.00	0.00
Gly	GGT	2.00	0.04	Cys	TGT	0.00	0.00
Gly	GGC	29.00	0.64	Cys	TGC	1.00	1.00
Glu	GAG	14.00	0.88	End	TAG	1.00	1.00
Glu	GAA	2.00	0.12	End	TAA	0.00	0.00
Asp	GAT	3.00	0.13	Tyr	TAT	4.00	0.31
Asp	GAC	20.00	0.87	Tyr	TAC	9.00	0.69
Val	GTG	7.00	0.22	Leu	TTG	3.00	0.13
Val	GTA	0.00	0.00	Leu	TTA	0.00	0.00
Val	GTT	7.00	0.22	Phe	TTT	1.00	0.12
Val	GTC	18.00	0.56	Phe	TTC	7.00	0.88
Ala	GCG	11.00	0.19	Ser	TCG	11.00	0.27
Ala	GCA	8.00	0.14	Ser	TCA	1.00	0.02
Ala	GCT	5.00	0.08	Ser	TCT	7.00	0.17
Ala	GCC	35.00	0.59	Ser	TCC	11.00	0.27
Arg	AGG	4.00	0.29	Arg	CGG	1.00	0.07
Arg	AGA	3.00	0.21	Arg	CGA	2.00	0.14
Ser	AGT	2.00	0.05	Arg	CGT	0.00	0.00
Ser	AGC	9.00	0.22	Arg	CGC	4.00	0.29
Lys	AAG	19.00	0.83	Gln	CAG	9.00	0.69
Lys	AAA	4.00	0.17	Gln	CAA	4.00	0.31
Asn	AAT	4.00	0.17	His	CAT	4.00	0.40
Asn	AAC	20.00	0.83	His	CAC	6.00	0.60
Met	ATG	6.00	1.00	Leu	CTG	7.00	0.30
Ile	ATA	0.00	0.00	Leu	CTA	1.00	0.04
Ile	ATT	4.00	0.15	Leu	CTT	4.00	0.17
Ile	ATC	23.00	0.85	Leu	CTC	8.00	0.35
Thr	ACG	12.00	0.36	Pro	CCG	2.00	0.13
Thr	ACA	2.00	0.06	Pro	CCA	3.00	0.20
Thr	ACT	4.00	0.12	Pro	CCT	5.00	0.33
Thr	ACC	15.00	0.45	Pro	CCC	5.00	0.33

The codon frequency for the *prt*1 gene coding region was calculated using the CODONFREQUENCY program from Wisconsin Package 9.1, GCG.

3.3.6 Analysis of the 3' UTR of *prt*1

3.3.6.1 Determination of the polyadenylation site by 3' RACE

The polyadenylation site of the prt1 gene was confirmed by 3' RACE using RNA from Lp19 cultures grown in PD broth for 14 days. Initially, a pool of cDNA was amplified using an oligo dT primer, which contains an adapter sequence and anneals to poly-A^+ mRNA. Using a primer complementary to the oligo dT adapter (UAP) and a gene-specific primer (MM4), PCR products were amplified. The products of this reaction were amplified using MM4 and MM8 as a control (Figure 3.8). This reaction showed at least four possible products (Figure 3.9A, lane 5).

The UAP-MM4 reaction was amplified using UAP and a nested gene specific primer, MM1. No product was visible in the PCR reaction (Figure 3.9B lane 7) but reamplification using the same primers gave a single product of approximately 700 bases in size (Figure 3.9C, lanes 2 and 3). Amplification of the UAP-MM4 reaction with the nested primers MM1 and MM8 also gave a single product (Figure 3.9B, lane 5). The reamplified UAP-MM1 product was ligated into a pGEM-T vector to generate pMM10 (Figure 3.10), which was sequenced using the MM16 primer. The sequence from the 3' RACE product is shown aligned with the *prt*1 gene sequence in Figure 3.10.

The polyadenylation site is 141 bases downstream of the translation stop site. In eukaryotes, the recognition sequence AAUAAA is required for both RNA cleavage and polyadenylation at the 3' end of many mRNAs (Zhao et al., 1999). The closest matches in the 3'UTR of *prt*1 are AATAAT (Figure 3.8, 1453 to 1458) and AATATA (1500 to 1505). In higher eukaryotes, this sequence is 11 to 30 base pairs away from the polyadenylation site. In *prt*1, the AATAAT sequence at 1453-1458 is 58 bp upstream of the polyadenylation site, while the AATATA sequence at 1500-1505 is 11 bp upstream.

3.3.6.2 Composition of the 3' UTR of prt1

A complex microsatellite was found within the 3' untranslated region of *prt*1 (Figure 3.8). The first part of the microsatellite covers twelve base pairs, consisting of a CTT repeat, with an imperfect TTT repeat at the beginning of the microsatellite. The second part of the

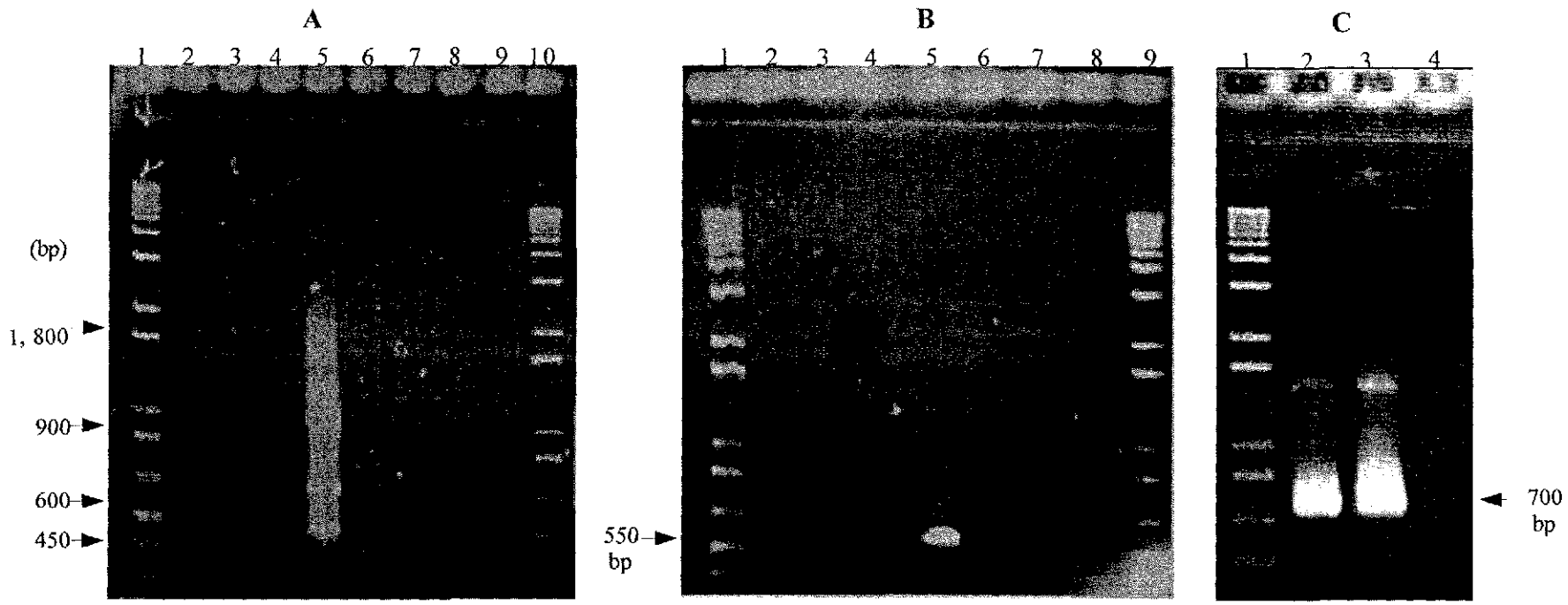


Figure 3.9 Amplification of fragments by 3' RACE

A. Amplification with MM4 primer. Lanes 2-6 were amplified with MM4 and MM8 (internal control); Lp19 genomic DNA (lane 2), cDNA before (lane 3) and after PCR clean up (lane 4), $^{1}/_{10}$ dilution of MM4-UAP reaction (lane 5), and water only control. Lanes 7-9 were amplified with MM4 and UAP: cDNA before clean up (lane 7), after clean up (lane 8) and water only control (lane 9). Lanes 1 and 10 contain 1kb + ladder (GibcoBRL).

B. Amplification with nested MM1 primer. Lanes 2-6 are the same as for A, except they were amplified with MM1 and MM8. Lanes 7 and 8 were amplified with MM1 and UAP; $^{1}/_{10}$ MM4-UAP reaction (lane 7), water only (lane 8). Lanes 1 and 9 contain 1kb + ladder (Gibco BRL).

C. Reamplification of MM1-UAP reaction: $^{1}/_{10}$ dilution MM1-UAP (lane 2), $^{1}/_{100}$ dilutions MM1-UAP (lane 3), water only control (lane 4). Lane 1 contains 1kb + ladder (GibcoBRL).

Figure 3.10 Sequencing of 3'RACE product

Sequence from the pMM10 vector containing the 3'RACE product for *prt*1 are shown aligned with the *prt*1 sequence below. The numbering shown in this figure is relative to the translation start, as in Figure 3.8. The sequence of the RACE product is shown in black text, while the *prt*1 sequence is shown in blue text. The probable polyadenylation site is indicated by a red arrow. Sequences shown in red boxes are putative polyadenylation recognition sequences, at 1453-1458 and 1500-1505, although the AT rich microsatellite sequence upstream of the second polyadenylation sequence could also contribute to the efficiency of polyadenylation. The distance between the second recognition sequence at 1500-1505 and the polyadenylation site is similar to the distance seen in higher eukaryotes.

microsatellite consists of a dinucleotide TA repeat, and is 38 base pairs in length. For some bases, thymines have changed to cytosines, and there are some insertions towards the 3' end of the microsatellite. This microsatellite was found between the proposed polyadenylation recognition sequence (Figure 3.8, 1453-1458) and the polyadenylation site (Figure 3.8, at 1516-1517) which was confirmed by 3' RACE (Section 3.3.6).

3.3.7 Repetitive DNA downstream of the *prt*1 gene

A minisatellite was found 388 bases downstream of the translation stop site and 247 bases downstream of the polyadenylation site. This minisatellite consisted of a nine base pair motif repeated 40 times to give a total length of 360 base pairs (Figure 3.8, Figure 3.11). In the centre of the minisatellite, the sequence TATTTATTC is repeated 17 times, with other variations towards the edges of the minisatellite. The most common substitutions take place at positions 1, 4, 5, 6 and 9 of the repeating unit (Figure 3.11B). At positions 1 and 5 of this unit adenine or cytosine substitutes thymine, while at position 4 thymine is only substituted by cytosine. At position 6, guanine or thymine sometimes substitutes adenine, and at position 9 thymine substitutes cytosine .

3.3.8 Proposed peptide sequence of Prt1

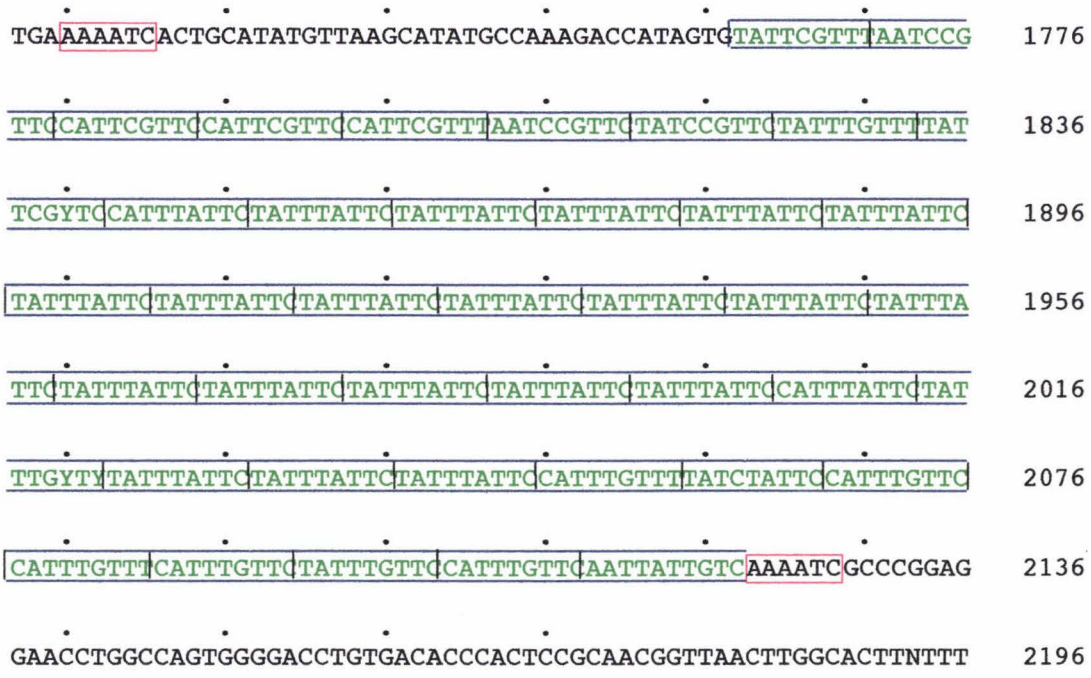
The proposed peptide sequence of Prt1, which is 434 amino acid residues in length, shares significant similarity with other fungal subtilisin-like proteases, particularly towards the C terminus of the protein. The signal peptide and propeptide were determined based on homology with other fungal subtilisin-like proteases. The proposed signal peptide of Prt1 consists of 24 amino acid residues (Figure 3.12, residues 1-24), with a large number of alanine residues. The Prt1 propeptide consisted of 101 amino acid residues (residues 25-125). Two motifs conserved with other subtilisin-like protease are found in the propeptide region of Prt1: the H1 box and the H3 box (Siezen et al., 1995).

The mature protein after signal peptide and propeptide cleavage consist of 309 amino acid residues (126-434). Active site residues were in highly conserved regions; aspartate was at 190, histidine was at 220, and serine was at 376. The proposed protein sequence contained four possible N-glycosylation sites, at 280, 318, 357 and 429.

Figure 3.11 Minisatellite DNA downstream of *prt*1

A. Minisatellite sequence

Minisatellite sequence is shown in green text, with flanking sequences in black text. The flanking repeats are shown in red boxes. The repeating units are indicated within navy blue boxes.



B. Frequency of bases at each position in the minisatellite repeat

This table shows the frequency of each nucleotide at each position in the minisatellite repeat.

Position in repeat	T	C	A	G	Y
1	0.675	0.250	0.075	0.000	-
2	0.000	0.000	1.000	0.000	-
3	1.000	0.000	0.000	0.000	-
4	0.900	0.100	0.000	0.000	-
5	0.775	0.200	0.025	0.000	-
6	0.025	0.000	0.575	0.400	-
7	0.950	0.000	0.000	0.000	0.050
8	1.000	0.000	0.000	0.000	-
9	0.125	0.850	0.000	0.000	0.025

HATYHDTTY

Consenus sequence where H = T, C or A, Y = T or C , and D = T, A or G.

Figure 3.12 Proposed peptide sequence of Prt1, a subtilisin-like protease

The signal peptide is shown in purple text, and the propeptide in green. The H1 and H3 regions of homology in the propeptide sequence are highlighted with yellow shading. The mature protease is shown in black text, with active site residues outlined with red boxes. N-glycosylation sites (NXS/T) are shown in blue text. The black arrow points to a possible autocatalytic site.

```
                                                H1
MLNVKNLVLTAAAALASQAIAAPTGPDAGNAKIQAAQGGQVIPGKFIVTLKPGSKPAVLE   60

                                                    H3
SHMRWVNGVHARASGDEAIKGVETMLDGIYGFMGYVGSFSEAVLAQIKAHPDVEAVEQDK  120

IWTLDWITDDQQLEARDDDKEPPSSGGGSNFIQQKNATWGLGSISHRAPYATEPGPQESA  180

GKDTYAYVIDTGIRTTHEEFEGRASHAWSAYLTRTDNVGHGTHVAGTIGGKTYGVAKNAK  240

LLAVKIFNSRSSSTSVILAGYNWAVNDIVRKGRTKRAAINMSLGGPKSTAFNTAVERASA  300

                          ↓
SGVLSIIAAGNEAQDASNVSPASAPSAITVAAINRDWTLASYSNFGSVVDICAPGSNITS  360

AWNTGDSSEKTISGTSMATPHVVGLALYAISVDGATGVDGVTKHLLSTATKDKVAGDTRG  420

SPNLIGNNNNSYQK                                                 434
```

Blast-X searches using the *prt*1 nucleotide sequence showed the translation of the *prt*1 gene had highest similarity to subtilisin-like proteases from *Aspergillus oryzae* (swissprot locus ORYZ_ASPOR, accession P12547) and *Aspergillus fumigatus* (swissprot locus ORYZ_ASPFU, accession P28296). FastA searches using Wisconsin Package version 9.1 (GCG) showed *that A. oryzae* and *A. fumigatus* subtilisin-like proteases shared 49% amino acid identity with Prt1. All of the other fungal subtilisin-like proteases analysed in a similar manner shared at least 35% identity with Prt1.

An alignment was prepared to show similarities between the Prt1 and other fungal subtilisin-like protease amino acid sequences. The alignment was based on ClustalW searches and PIMA alignments using maximal linkage clustering and sequential branching clustering. The mature part of Prt1 shared a large degree of sequence identity, especially around active site residues (Figure 3.13). The alignment shows that Prt1 contains an insertion at the beginning of the propeptide sequence and a large insertion at the amino terminus of the mature protein from residues 130 to 153 compared to other fungal subtilisin-like proteases.

3.4 DEVELOPMENT OF A *PRT*1-GUS EXPRESSION VECTOR

3.4.1 Amplification of *prt*1 promoter fragment

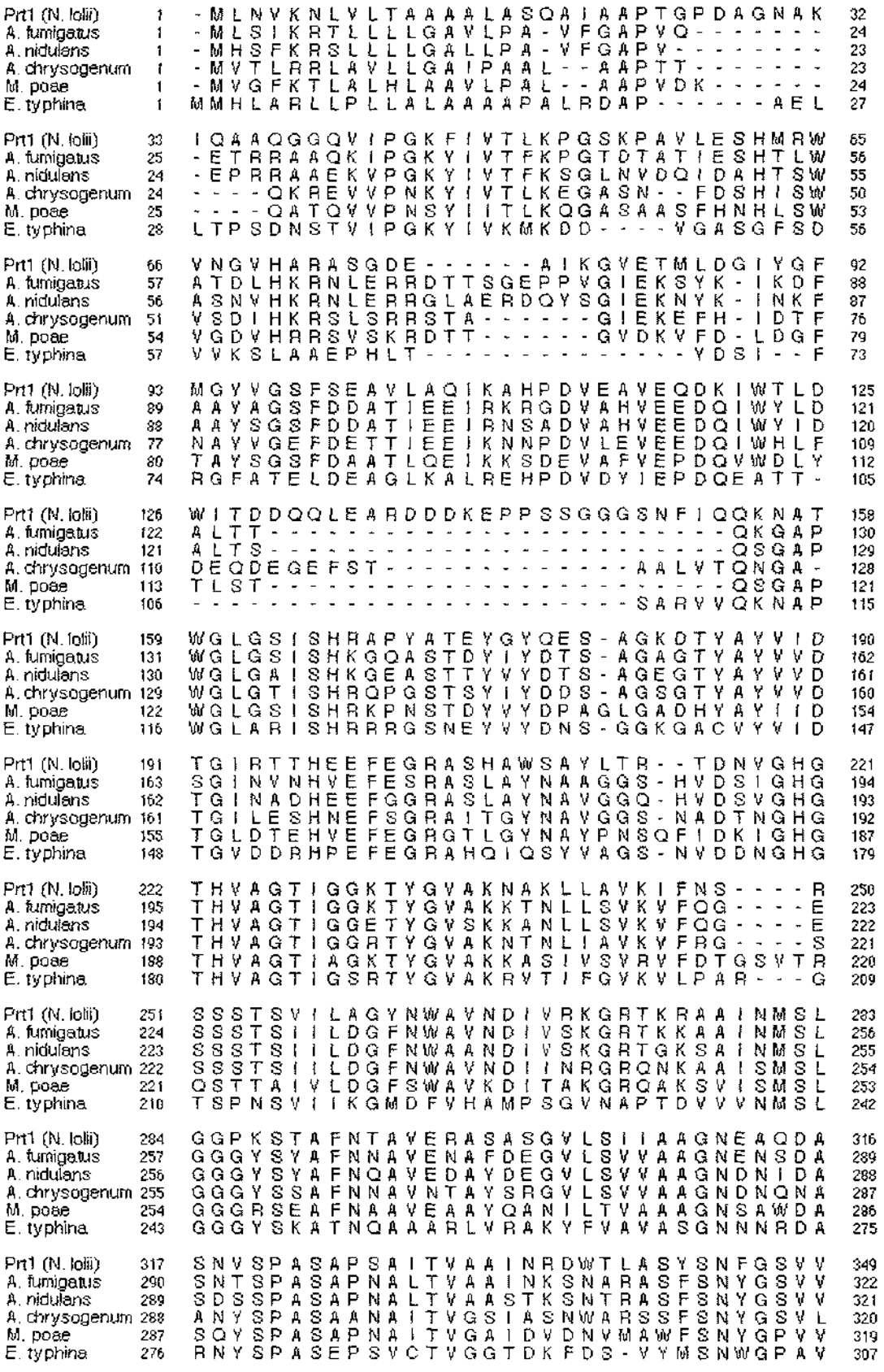
The promoter fragment was amplified for use in an expression vector by designing a primer, prot1Nco, which was complementary to the translation initiation site of *prt*1. Several base substitutions were made to introduce a *Nco*I site into the primer and therefore the amplified fragment. The M13 forward and prot1Nco primers were used to amplify the promoter region from pMM2 (Section 3.2.3, Figure 3.14). The product of this reaction was digested to give an *Eco*RI/*Nco*I fragment containing the promoter region.

3.4.2 Subcloning of *prt*1 promoter fragment into pFunGus

The promoter fragment was ligated into a pFunGus vector (Appendix A1.3) that was also digested with *Eco*RI and *Nco*I. The transformants were screened by digestion with *Eco*RI and *Nco*I. This new vector, pMM8 (Figure 3.14), contained a translational fusion of the *prt*1 promoter with the GUS gene of the pFunGus vector. However, pMM8 required a selectable marker for easier detection in fungal transformants.

Figure 3.13 Alignment of Prt1 with other fungal subtilisin-like protease sequences.

The Prt1 peptide sequence was aligned with peptide sequences from other fungi which showed high degrees of similarity in Blast-P searches. The accession numbers for the fungal protease sequences were as follows: *Aspergillus fumigatus* (swissprot: locus ORYZ_ASPFU, accession P28296), *Aspergillus nidulans* (GenBank locus ASNPRTA accession L31778.1), *Acremonium chrysogenum* (swissprot: locus ALP_CEPAC, accession P29118), *Magnaporthe poae* (GenBank locus AF118126 accession AF118126.1) and *Epichloë typhina* (GenBank locus APEPROT accession L76740.1).


Prt1 (N. lolii) 1 - M L N V K N L V L T A A A A L A S Q A I A A P T G P D A G N A K 32
A. fumigatus 1 - M L S I K R T L L L L G A V L P A - V F G A P V Q - - - - - - - 24
A. nidulans 1 - M H S F K R S L L L L G A L L P A - V F G A P V - - - - - - - - 23
A. chrysogenum 1 - M V T L R R L A V L L G A I P A A L - - A A P T T - - - - - - - 23
M. poae 1 - M V G F K T L A L H L A A V L P A L - - A A P V D K - - - - - - - 24
E. typhina 1 M M H L A R L L P L L A L A A A A P A L R D A P - - - - - - A E L 27
Prt1 (N. lolii) 33 I Q A A Q G G Q V I P G K F I V T L K P G S K P A V L E S H M R W 65
A. fumigatus 25 - E T R R A A Q K I P G K Y I V T F K P G T D T A T I E S H T L W 56
A. nidulans 24 - E P R R A A E K V P G K Y I V T F K S G L N V D Q I D A H T S W 55
A. chrysogenum 24 - - - - Q K R E V V P N K Y I V T L K E G A S N - - F D S H I S W 50
M. poae 25 - - - - Q A T Q V V P N S Y I I T L K Q G A S A A S F H N H L S W 53
E. typhina 28 L T P S D N S T V I P G K Y I V K M K D D - - - - V G A S G F S D 56
Prt1 (N. lolii) 66 V N G V H A R A S G D E - - - - - - - A I K G V E T M L D G I Y G F 92
A. fumigatus 57 A T D L H K R N L E R R D T T S G E P P V G I E K S Y K - I K D F 88
A. nidulans 56 A S N V H K R N L E R R G L A E R D Q Y S G I E K N Y K - I N K F 87
A. chrysogenum 51 V S D I H K R R S L S R R S T A - - - - - - - G I E K E F H - I D T F 76
M. poae 54 V G D V H R R S V S K R D T T - - - - - - - G V D K V F D - L D G F 79
E. typhina 57 V V K S L A A E P H L T - - - - - - - - - - - - - - Y D S I - - F 73
Prt1 (N. lolii) 93 M G Y V G S F S E A V L A Q I K A H P D V E A V E Q D K I W T L D 125
A. fumigatus 89 A A Y A G S F D D A T I E E I R K R G D V A H V E E D Q I W Y L D 121
A. nidulans 88 A A Y S G S F D D A T I E E I R N S A D V A H V E E D Q I W Y I D 120
A. chrysogenum 77 N A Y V G E F D E T T I E E I K N N P D V L E V E E D Q I W H L F 109
M. poae 80 T A Y S G S F D A A T L Q E I K K S D E V A F V E P D Q V W D L Y 112
E. typhina 74 R G F A T E L D E A G L K A L R E H P D V D Y I E P D Q E A T T - 105
Prt1 (N. lolii) 126 W I T D D Q Q L E A R D D D K E P P S S G G G S N F I Q Q K N A T 158
A. fumigatus 122 A L T T - Q K G A P 130
A. nidulans 121 A L T S - Q S G A P 129
A. chrysogenum 110 D E Q D E G E F S T - - - - - - - - - - - - A A L V T Q N G A - 128
M. poae 113 T L S T - Q S G A P 121
E. typhina 106 - S A R V V Q K N A P 115
Prt1 (N. lolii) 159 W G L G S I S H R A P Y A T E Y G Y Q E S - A G K D T Y A Y V I D 190
A. fumigatus 131 W G L G S I S H K G Q A S T D Y I Y D T S - A G A G T Y A Y V V D 162
A. nidulans 130 W G L G A I S H K G E A S T T Y W Y D T S - A G E G T Y A Y V V D 161
A. chrysogenum 129 W G L G T I S H R Q P G S T S Y I Y D D S - A G S G T Y A Y V V D 160
M. poae 122 W G L G S I S H R K P N S T D Y V Y D P A G L G A D H Y A Y I I D 154
E. typhina 116 W G L A R I S H R R R G S N E Y V Y D N S - G G K G A C V Y V I D 147
Prt1 (N. lolii) 191 T G I R T T H E E F E G R A S H A W S A Y L T R - - T D N V G H G 221
A. fumigatus 163 S G I N V N H V E F E S R A S L A Y N A A G G S - H V D S I G H G 194
A. nidulans 162 T G I N A D H E E F G G R A S L A Y N A V G G Q - H V D S V G H G 193
A. chrysogenum 161 T G I L E S H N E F S G R A I T G Y N A V G G S - N A D T N G H G 192
M. poae 155 T G L D T E H V E F E G R G T L G Y N A Y P N S Q F I D K I G H G 187
E. typhina 148 T G V D D D R H P E F E G R A H Q I Q S Y V A G S - N V D D N G H G 179
Prt1 (N. lolii) 222 T H V A G T I G G K T Y G V A K N A K L L A V K I F N S - - - - R 250
A. fumigatus 195 T H V A G T I G G K T Y G V A K K T N L L S V K V F Q G - - - - E 223
A. nidulans 194 T H V A G T I G G E T Y G V S K K A N L L S V K V F Q G - - - - E 222
A. chrysogenum 193 T H V A G T I G G R T Y G V A K N T N L I A V K V F R G - - - - S 221
M. poae 188 T H V A G T I A G K T Y G V A K K A S I V S V R V F D T G S V T R 220
E. typhina 180 T H V A G T I G S R T Y G V A K R V T I F G V K V L P A R - - - G 209
Prt1 (N. lolii) 251 S S S T S V I L A G Y N W A V N D I V R K G R T K R A A I N M S L 283
A. fumigatus 224 S S S T S I I L D G F N W A V N D I V S K G R T K K A A I N M S L 256
A. nidulans 223 S S S T S I I L D G F N W A A N D I V S K G R T G K S A I N M S L 255
A. chrysogenum 222 S S S T S I I L D G F N W A V N D I I N R G R Q N K A A I S M S L 254
M. poae 221 Q S T T A I V L D G F S W A V K D I T A K G R Q A K S V I S M S L 253
E. typhina 210 T S P N S V I I K G M D F V H A M P S G V N A P T D V V V N M S L 242
Prt1 (N. lolii) 284 G G P K S T A F N T A V E R A S A S G V L S I I A A G N E A Q D A 316
A. fumigatus 257 G G G Y S Y A F N N A V E N A F D E G V L S V V A A G N E N S D A 289
A. nidulans 256 G G G Y S Y A F N Q A V E D A Y D E G V L S V V A A G N D N I D A 288
A. chrysogenum 255 G G G Y S S A F N N A V N T A Y S R G V L S V V A A G N D N Q N A 287
M. poae 254 G G G R S E A F N A A V E A A Y Q A N I L T V A A A G N S A W D A 286
E. typhina 243 G G G Y S K A T N Q A A A R L V R A K Y F V A V A S G N N N R D A 275
Prt1 (N. lolii) 317 S N V S P A S A P S A I T V A A I N R D W T L A S Y S N F G S V V 349
A. fumigatus 290 S N T S P A S A P N A L T V A A I N K S N A R A S F S N Y G S V V 322
A. nidulans 289 S D S S P A S A P N A L T V A A S T K S N T R A S F S N Y G S V V 321
A. chrysogenum 288 A N Y S P A S A A N A I T V G S I A S N W A R S S F S N Y G S V L 320
M. poae 287 S Q Y S P A S A P N A I T V G A I D V D N V M A W F S N Y G P V V 319
E. typhina 276 R N Y S P A S E P S V C T V G G T D K F D S - V Y M S N W G P A V 307

```
Prt1 (N. lolii)     350  D I C A P G S N I T S A W N T G D S S E K T I - S G T S M A T P H  381
A. fumigatus        323  D I F A P G Q D I L S A W I G S T T A T N T I - S G T S M A T P H  354
A. nidulans         322  D I F A P G Q D I L S A W I G S T T A T N T I - S G T S M A T P H  353
A. chrysogenum      321  D I F A P G T S I L S A W I G G N S A T N T I - S G T S M A T P H  352
M. poae             320  D V F A P G V A V E S A W I G S S H A E H D V L D G T S M A T P H  352
E. typhina          308  D I N G P G V D V L S T L P N R R T G R L - - - T G T S M A T P H  337

Prt1 (N. lolii)     382  V V G L A L Y A I S V D G A T G V D G V T K H L L S T A T K D K V  414
A. fumigatus        355  I V G L S V Y L M G L E N L S G P A A V T A R I K E L A T N G V V  387
A. nidulans         354  V V G L S L Y L I A L E G L S S A S A V V S R I K E L A T Q G V L  386
A. chrysogenum      353  V T G V V L Y L Q A L E G L T T S G A A A R L N A L A T T G R V S  385
M. poae             353  V S G L V L Y L K S L E G F A S A A A V T D R I K A L G T N D V V  385
E. typhina          338  I A G L G A Y L A A K N G R R A G P G L C R T I K D M A T K N V I  370

Prt1 (N. lolii)     415  A G D T R - G S P N L I G N N N N S Y Q K                          434
A. fumigatus        388  T N V K - - G S P N K L A Y N G N A                               403
A. nidulans         387  S N V Q - - G S P N L L A Y N G A D E                             403
A. chrysogenum      386  N P G S - - G S P N R I L Y N G N G A                             402
M. poae             386  T G L E G T D S P N L I A F N G V T A                             404
E. typhina          371  T N Q V A - G T V N L L A F N G E K                               387
```

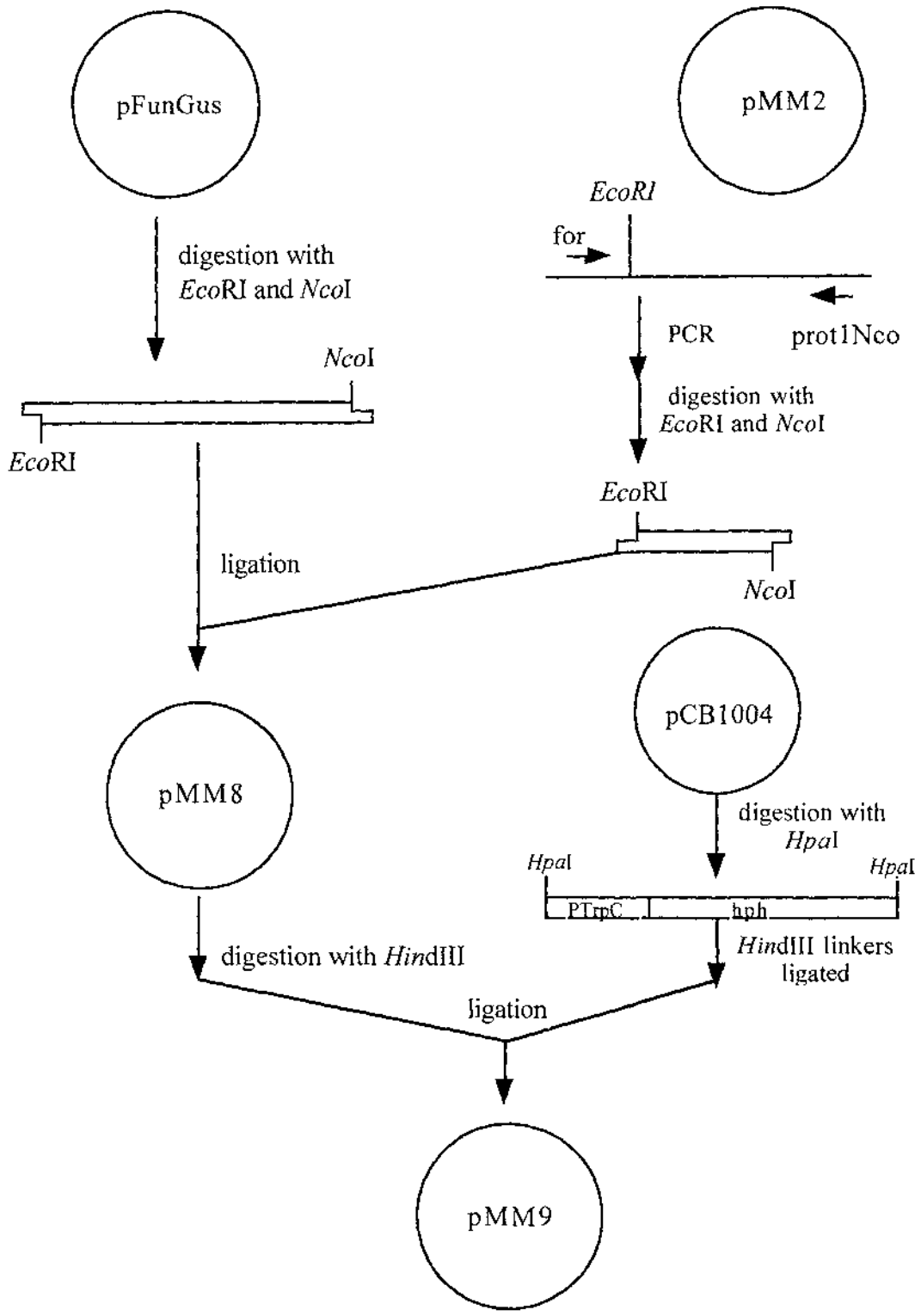


Figure 3.14 Development of the *prt*1 expression vector pMM9

3.4.3 Subcloning of a *hph* selectable marker into pMM8

A *hph* selectable marker from pCB1004 (Appendix A1.2) was introduced into the *prt*1 expression vector pMM8. Initially a 1.4 kb fragment containing the marker was removed by digestion with *Hpa*I, which creates a blunt-ended restriction site. *Hin*dIII linkers were ligated onto a 1.4 kb *Hpa*I fragment, which was then ligated into *Hin*dIII-digested pMM8. The resulting vector pMM9 contained a translational fusion between the *prt*1 promoter and the GUS gene and a hygromycin resistance selectable marker (Figure 3.14).

3.4.4 Sequencing of promoter fragment in pMM9

Before transformation of pMM9 it was necessary to check that the *prt*1 promoter fragment had no base substitutions as a result of PCR amplification. If any alteration to the sequence had taken place it could change the expression of GUS from this promoter. Sequencing using a primer annealing to the T7 promoter region of pFunGus showed that the *prt*1 promoter fragment was identical to that of pMM2, which contained the *prt*1 fragment used for sequencing (sequence data shown in Appendix A3.3). The pMM9 vector was now ready for transformation into *Penicillium paxilli*.

3.5 TRANSFORMATION OF pMM9 INTO *PENICILLIUM PAXILLI*

3.5.1 Preparation and transformation of *P. paxilli* protoplasts

Protoplasts were prepared from 48 hour cultures of *P. paxilli* (Section 2.2). The concentration of protoplasts was checked using a haemocytometer. This revealed the concentration of protoplasts was 6.8 x 10^8 protoplasts/mL (Table 3.7).

P. paxilli protoplasts were transformed using a PEG-mediated DNA uptake method. Circular pMM9 DNA was transformed and pAN7-1 (Appendix 1.1) linearised with *Hin*dIII was used as a positive control. Protoplasts were also used both as a negative control for hygromycin resistance and to show protoplast viability after transformation (Table 3.7). Transformation results showed low numbers of hygromycin-resistant transformants, where the frequency of transformants were 5.7x10^{-7} transformants/5 ug DNA/mL of viable protoplasts for pAN7-1, and 8.5x10^{-7} transformants/5 ug DNA/mL of viable protoplasts for pMM9. Despite the low transformation frequencies pMM9 hygromycin-resistant transformants were single-spore purified three times to ensure stable hygromycin resistance in preparation for GUS analysis.

Table 3.7 Transformation of *Penicillium paxilli*

Plasmid [a]	Viable protoplasts [b]	Viable protoplasts after transformtion [b]	Hyg^R transformants [c]	Hyg^R Gus^+ transformants [3]
Cells only	$6.8x10^8$	$1.4x10^7$		
Linear pAN7-I			8 ($5.7x10^{-7}$)[d]	
Circular pMM9			12 ($8.5x10^{-7}$)[d]	7

a pAN7-1 was linearised with *Hin*dIII

b Viable protoplasts per mL

c Colonies per 5 μg of DNA

d Frequency of transformants/ 5 μg DNA/ mL of viable protoplasts

3.5.2 GUS qualitative assays of transformants

GUS activity in pMM9 hygromycin resistant transformants was assayed using a qualitative GUS microtitre assay (Figure 3.15) with MUG as a substrate. As a negative control GUS assay buffer was incubated alone (lane A). For spores from each transformant, duplicates were assayed (lanes B and C). The GUS qualitative assay revealed that seven of the 12 hygromycin resistant transformants were also producing GUS (rows 1, 2, 4-6, 8 and 12). These positive results indicate that the 0.5-kb *prt*1 promoter fragment in pMM9 is capable of promoting GUS expression in a heterologous system.

3.6 EXPRESSION OF THE *PRT*1 GENE IN RESPONSE TO NUTRIENTS

RNA was isolated from *N. lolii* strain Lp19 cultures grown in PD broth for nine days then transferred to a supplemented CD salts media for two days. The cultures were supplemented with sucrose, mannitol, glutamate or nitrate, either singly or in combination. As well as examining *prt*1 the expression of the Lp19 *hmg* gene, which encodes hmg CoA reductase (Dobson, 1997), was analysed in these cultures as a positive control (Figures 3.16-3.21, Part A).

When the cultures were grown in unsupplemented CD salts media the *prt*1 gene was expressed (Figure 3.16). The CD salts medium (Section 2.3.1) did not contain any carbon or nitrogen sources. If the media was supplemented with glutamate the *prt*1 gene was also expressed (Figure 3.17). Similar results were also seen for cultures supplemented with sucrose only (Figure 3.18). *prt*1 expression was strong in cultures supplemented with sucrose and nitrate (Figure 3.19), but weak in cultures supplemented with sucrose and glutamate (Figure 3.20). In contrast, *prt*1 expression in cultures supplemented with mannitol and glutamate (Figure 3.21) was comparable to *prt*1 expression seen in cultures containing no supplements (Figure 3.16).

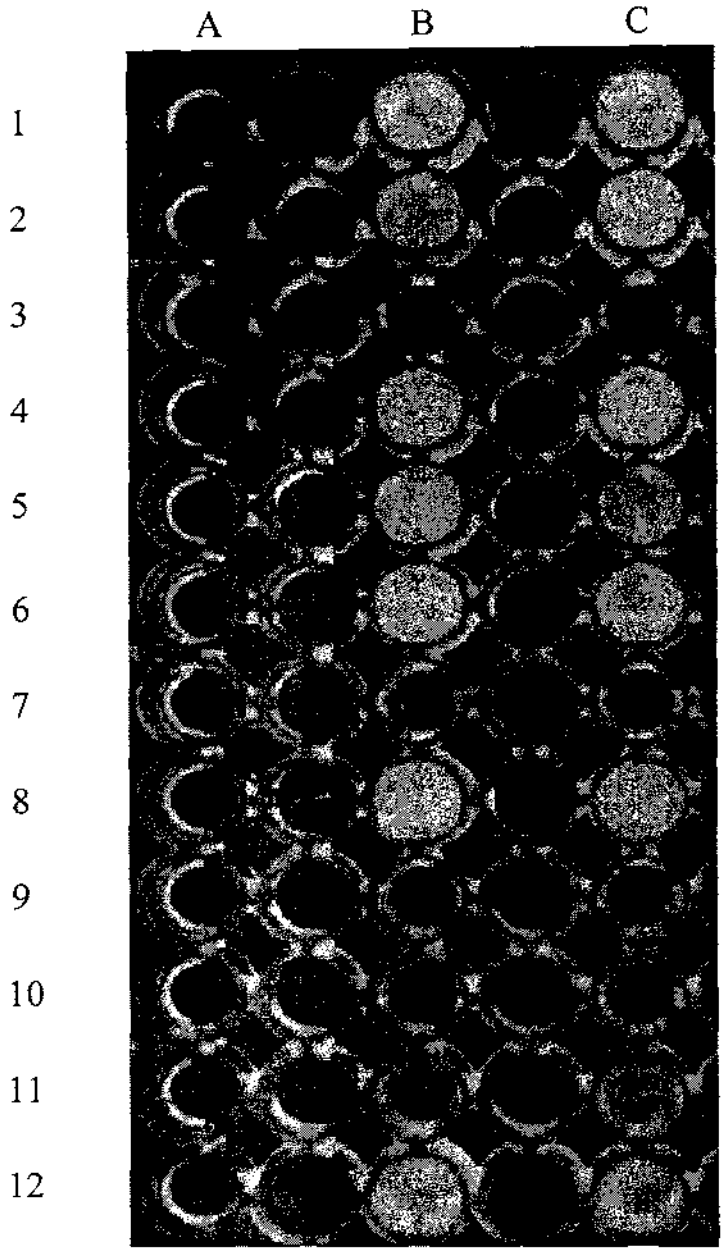


Figure 3.15 **GUS Qualitative Assay for *Penicillium paxilli* transformed with pMM9**

Qualitative assays for GUS expression from *Penicillium paxilli* transformed with the *prt*1 expression vector pMM9. Lane A contains GUS assay buffer only, and lanes B and C are duplicates containing GUS assay buffer and transformant spores. The numbers 1-12 refer to different transformant colonies. A light gray colour in both lanes B and C indicates GUS expression.

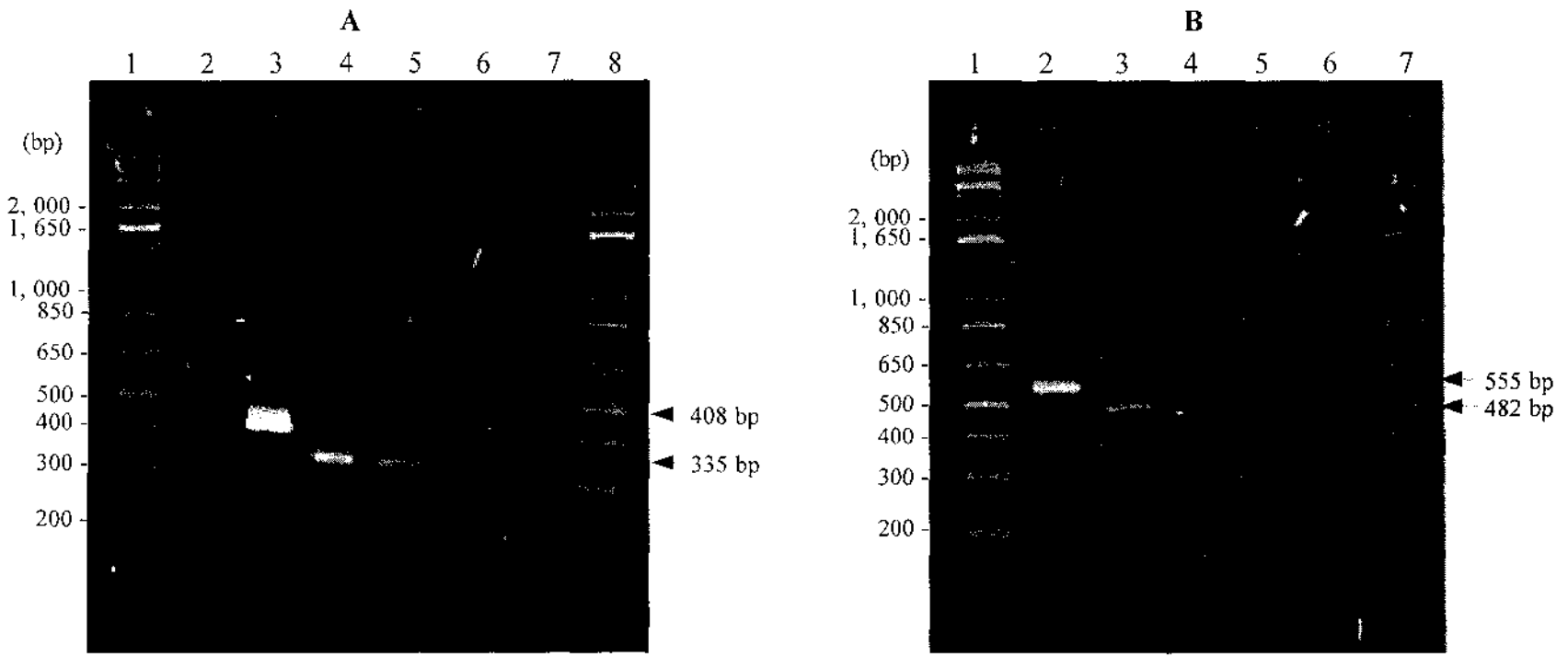


Figure 3.16 Gene expression in Lp19 cultures grown in CD salts only.

A: Expression of hmg CoA reductase gene using primers hmg29 and hmg30: no RT control (lane 2), Lp19 genomic DNA (lane 3), $^{1}/_{10}$ cDNA dilution (lane 4), $^{1}/_{100}$ cDNA dilution (lane 5), $^{1}/_{1000}$ cDNA dilution (lane 6) and water only control (lane 7). Lanes 1 and 8 contain 1 kb + ladder (GibcoBRL).

B. Expression of the *prt*1 gene using primers MM2 and MM5: Lp19 genomic DNA (lane 2), $^{1}/_{10}$ cDNA (lane 3), $^{1}/_{100}$ cDNA (lane 4), $^{1}/_{1000}$ cDNA (lane 5) and water only control (lane 6). Lanes 1 and 7 contain 1 kb + ladder (Gibco BRL).

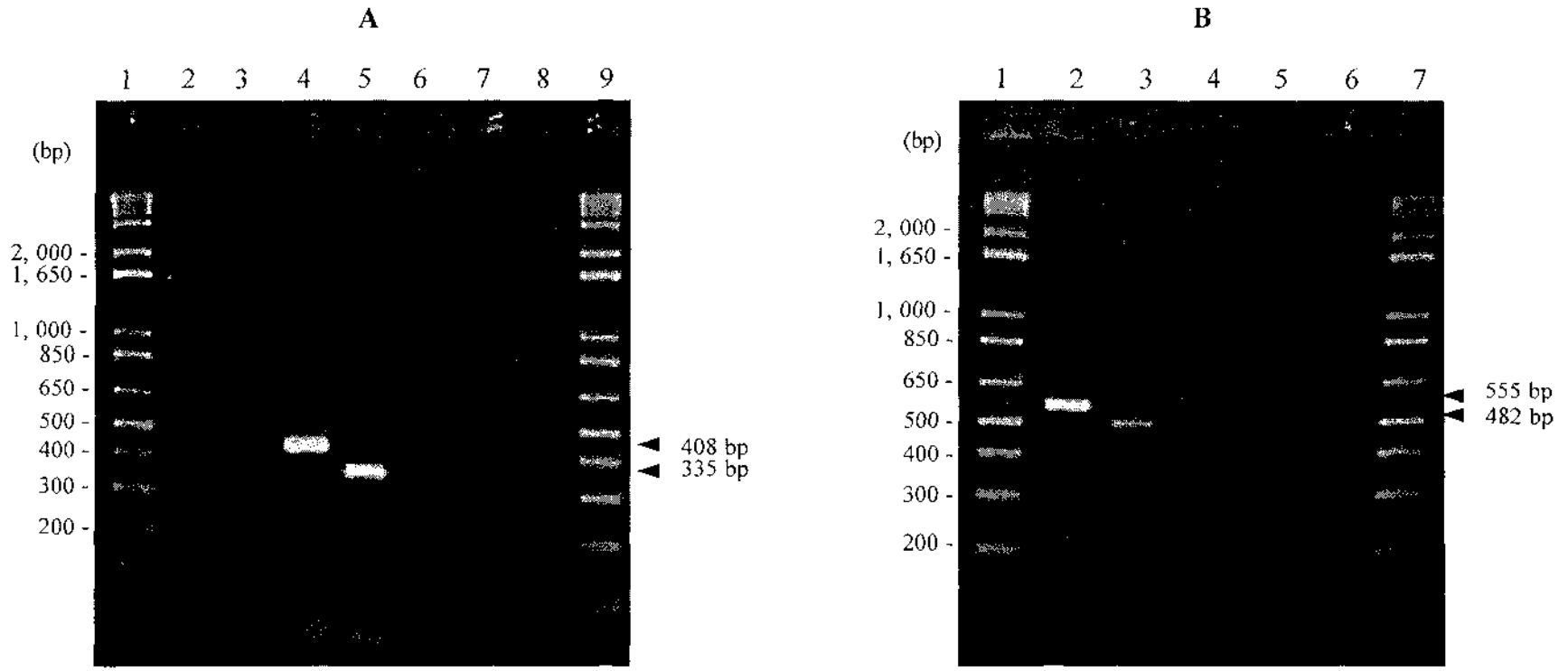


Figure 3.17 Gene expression in Lp19 cultures grown in CD salts supplemented with glutamate.

A: Expression of hmg CoA reductase gene using primers hmg29 and hmg30: no RT control (lane 2), no RNA control (lane 3), Lp19 genomic DNA (lane 4), $^1/_{10}$ cDNA dilution (lane 5), $^1/_{100}$ cDNA dilution (lane 6), $^1/_{1000}$ cDNA dilution (lane 7) and water only control (lane 8). Lanes 1 and 9 contain 1 kb + ladder (GibcoBRL).

B. Expression of the *prt*1 gene using primers MM2 and MM5: Lp19 genomic DNA (lane 2), $^1/_{10}$ cDNA (lane 3), $^1/_{100}$ cDNA (lane 4), $^1/_{1000}$ cDNA (lane 5) and water only control (lane 6). Lanes 1 and 7 contain 1 kb + ladder (Gibco BRL).

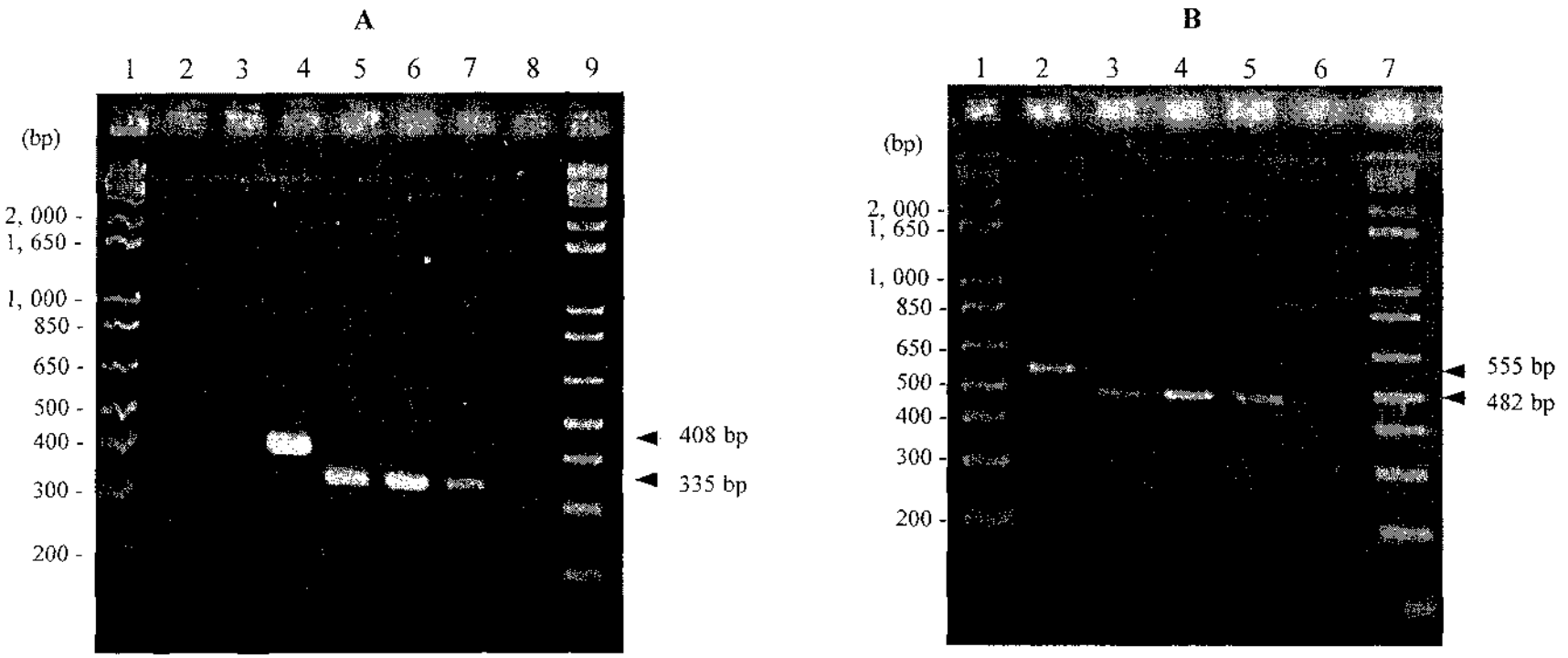


Figure 3.18 Gene expression in Lp19 cultures grown in CD salts supplemented with sucrose.

A: Expression of hmg CoA reductase gene using primers hmg29 and hmg30: no RT control (lane 2), no RNA control (lane 3), Lp19 genomic DNA (lane 4), $^1/_{10}$ cDNA dilution (lane 5), $^1/_{100}$ cDNA dilution (lane 6), $^1/_{1000}$ cDNA dilution (lane 7) and water only control (lane 8). Lanes 1 and 9 contain 1 kb + ladder (GibcoBRL).

B. Expression of the *prt*1 gene using primers MM2 and MM5: Lp19 genomic DNA (lane 2), $^1/_{10}$ cDNA (lane 3), $^1/_{100}$ cDNA (lane 4), $^1/_{1000}$ cDNA (lane 5) and water only control (lane 6). Lanes 1 and 7 contain 1 kb + ladder (Gibco BRL).

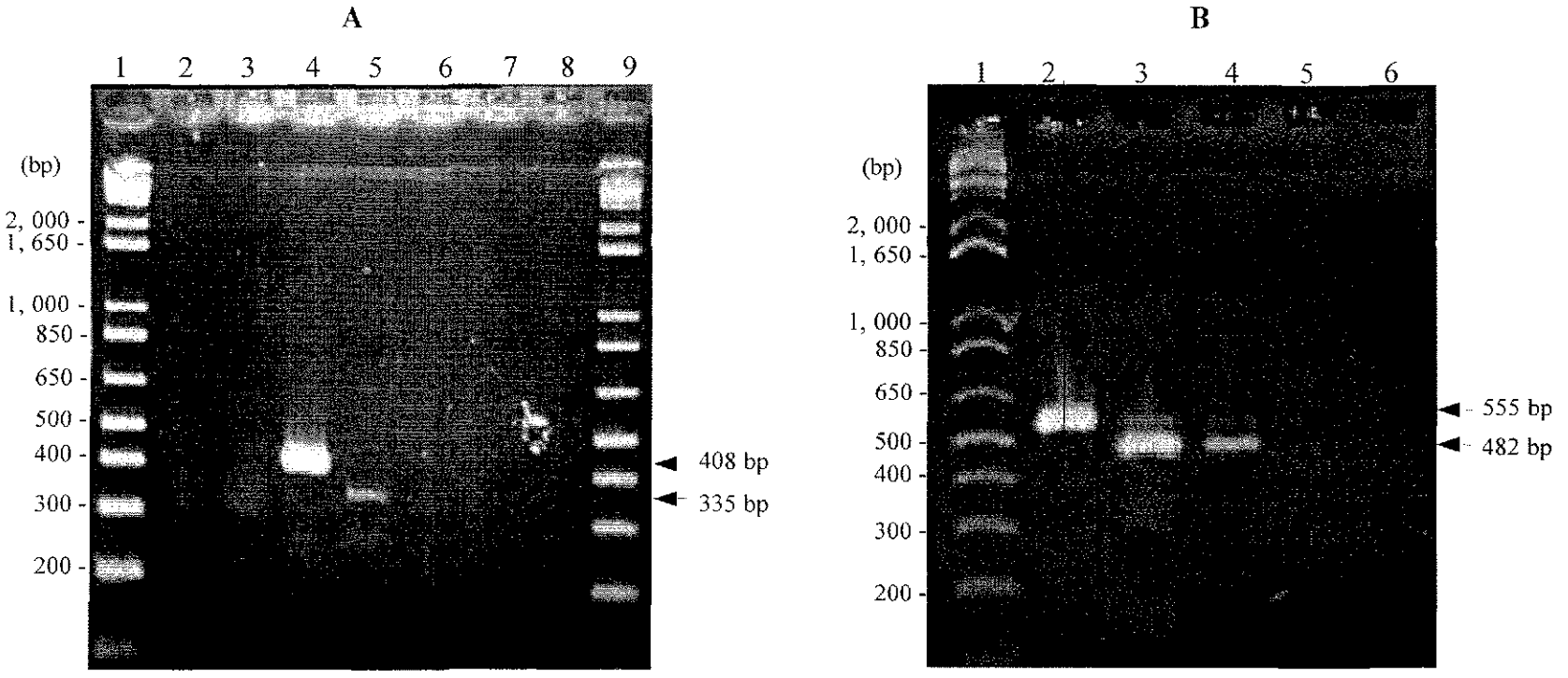


Figure 3.19 Gene expression in Lp19 cultures grown in CD salts supplemented with sucrose and nitrate.

A: Expression of hmg CoA reductase gene using primers hmg29 and hmg30: no RT control (lane 2), no RNA control (lane 3), Lp19 genomic DNA (lane 4), $^{1}/_{10}$ cDNA dilution (lane 5), $^{1}/_{100}$ cDNA dilution (lane 6), $^{1}/_{1000}$ cDNA dilution (lane 7) and water only control (lane 8). Lanes 1 and 9 contain 1 kb + ladder (GibcoBRL).

B. Expression of the *prt*1 gene using primers MM2 and MM5: Lp19 genomic DNA (lane 2), $^{1}/_{10}$ cDNA (lane 3), $^{1}/_{100}$ cDNA (lane 4), $^{1}/_{1000}$ cDNA (lane 5) and water only control (lane 6). Lane 1 contains 1 kb + ladder (Gibco BRL).

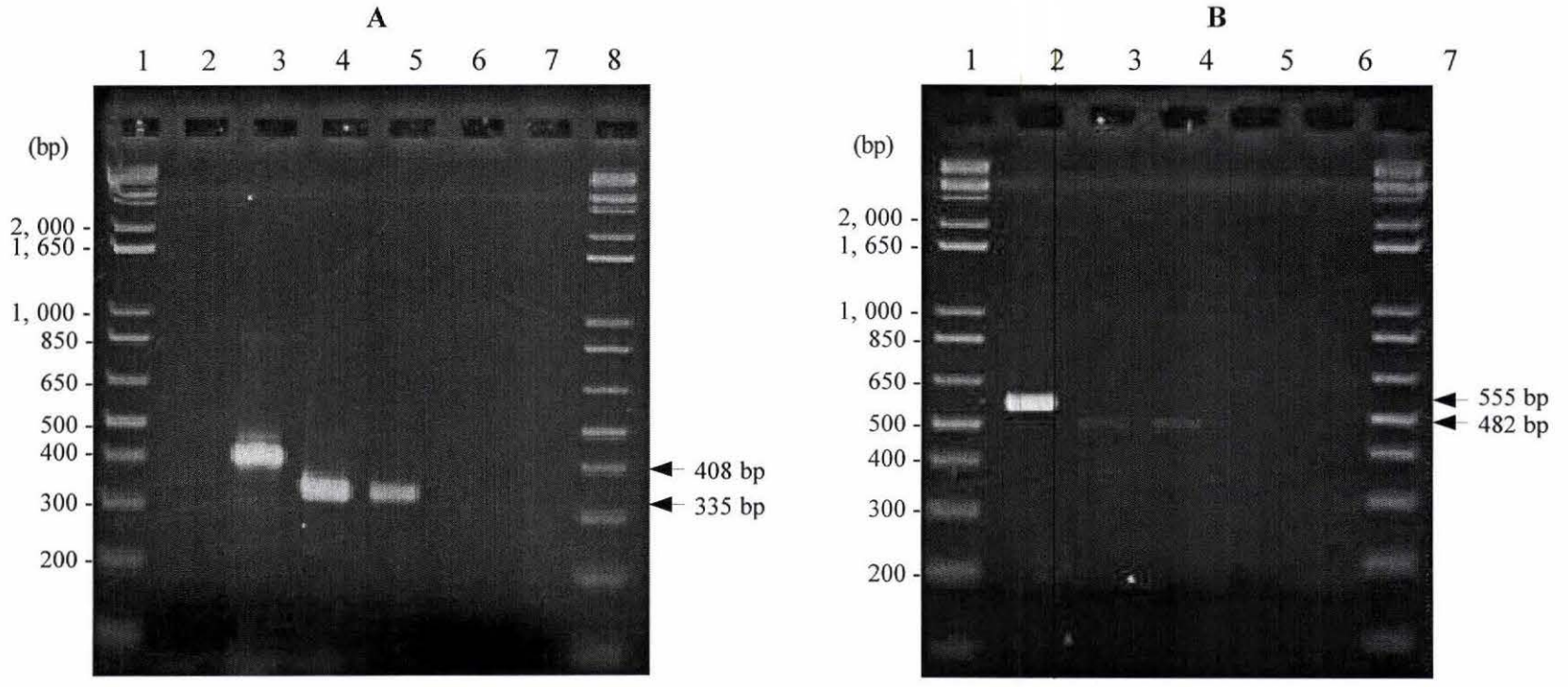


Figure 3.20 Gene expression in Lp19 cultures grown in CD salts supplemented with sucrose and glutamate.

A: Expression of hmg CoA reductase gene using primers hmg29 and hmg30: no RT control (lane 2), Lp19 genomic DNA (lane 3), $^1/_{10}$ cDNA dilution (lane 4), $^1/_{100}$ cDNA dilution (lane 5), $^1/_{1000}$ cDNA dilution (lane 6) and water only control (lane 7). Lanes 1 and 8 contain 1 kb + ladder (GibcoBRL).

B. Expression of the *prt*1 gene using primers MM2 and MM5: Lp19 genomic DNA (lane 2), $^1/_{10}$ cDNA (lane 3), $^1/_{100}$ cDNA (lane 4), $^1/_{1000}$ cDNA (lane 5) and water only control (lane 6). Lanes 1 and 7 contain 1 kb + ladder (Gibco BRL).

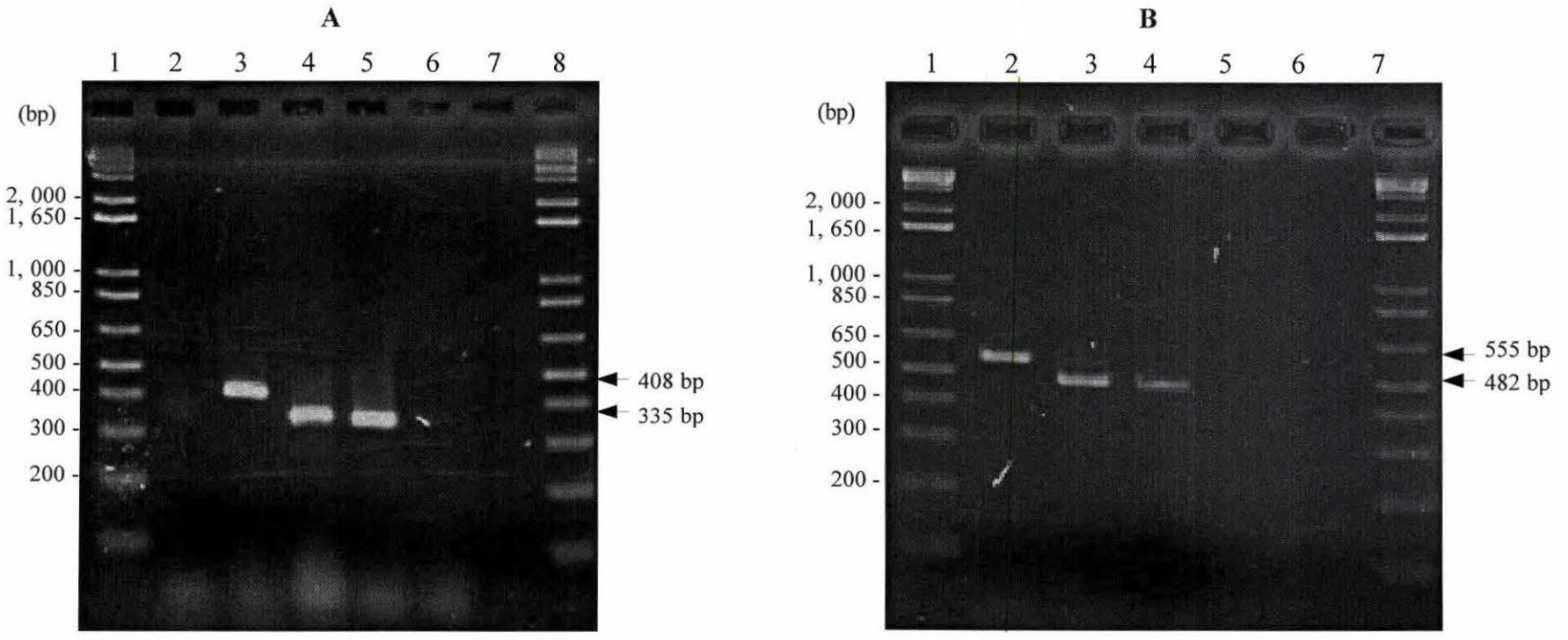


Figure 3.21 Gene expression in Lp19 cultures grown in CD salts supplemented with mannitol and glutamate.

A: Expression of hmg CoA reductase gene using primers hmg29 and hmg30: no RT control (lane 2), Lp19 genomic DNA (lane 3), $^{1}/_{10}$ cDNA dilution (lane 4), $^{1}/_{100}$ cDNA dilution (lane 5), $^{1}/_{1000}$ cDNA dilution (lane 6) and water only control (lane 7). Lanes 1 and 8 contain 1 kb + ladder (GibcoBRL).

B. Expression of the *prt*1 gene using primers MM2 and MM5: Lp19 genomic DNA (lane 2), $^{1}/_{10}$ cDNA (lane 3), $^{1}/_{100}$ cDNA (lane 4), $^{1}/_{1000}$ cDNA (lane 5) and water only control (lane 6). Lanes 1 and 7 contain 1 kb + ladder (Gibco BRL).

CHAPTER FOUR

ANALYSIS OF THE LP19 *PRT2* GENE

4.1 SOUTHERN BLOTTING OF ENDOPHYTE DNA

A 600bp *Eco*RI/*Sst*I fragment from pGH3 (Appendix A2.3) was [α^{32}P]dCTP-labelled and used to probe a Southern blot containing endophyte genomic DNA (Figure 4.1). The sizes of hybridising fragments are shown in Table 4.1. Lp1 contained at least two hybridising fragments for each digest (Figure 4.1, lanes 2-4). Lp19 shares an 8.6 kb *Eco*RI and 8.1 kb *Sst*I (lanes 5-7) hybridising fragments with the Lp1 digests. The E8 digests also share bands in common with Lp1: 4 kb *Eco*RI bands (lanes 2 and 8) and 9.2 kb *Sst*I bands (lanes 4 and 10). The pGH3 probe contains a *Hin*dIII site, which means two hybridising fragments would be expected for each *Hin*dIII restriction digest. Lp1, Lp19 and E8 all shared a small hybridising *Hin*dIII fragment (lanes 3, 6 and 9).

The blot shown in Figure 3.2 was reprobed with an [α^{32}P]dCTP-labelled *Eco*RI/*Sst*I fragment from pGH3 (Figure 4.2). Sizes of hybridising fragments are shown in Table 4.2. This blot showed that all of the endophytes screened contained a small *Hin*dIII hybridising fragment, which is conserved between all the taxonomic groups (lanes 2, 3, 6, 7, 8 and 9). A hybridising fragment is difficult to detect in the E8 genomic DNA (Figure 4.2B lane 4), but in Figure 4.1B (lane 9) the 1.2 kb *Hin*dIII hybridising fragment seen for all the other endophyte taxonomic groups is evident. This result with all the endophyte groups containing the same sized hybridising fragment was very different to the result seen for the *prt*1 gene (Section 3.1), which showed a lot of variation in *prt*1 *Hin*dIII fragment size between the endophytes.

4.2 LIBRARY SCREENING AND MAPPING

4.2.1 Library screening

A λGEM-12 library containing Lp19 genomic DNA was screened as described in Section 3.2.1, but probed with an [α^{32}P]dCTP-labelled *Eco*RI-*Sst*I fragment from pGH3. After the first round of library screening two further rounds of screening were used to purify positive λ clones. Four positive clones, λMM3.1, λMM3.3, λMM3.5 and λMM3.6 were selected for further analysis.

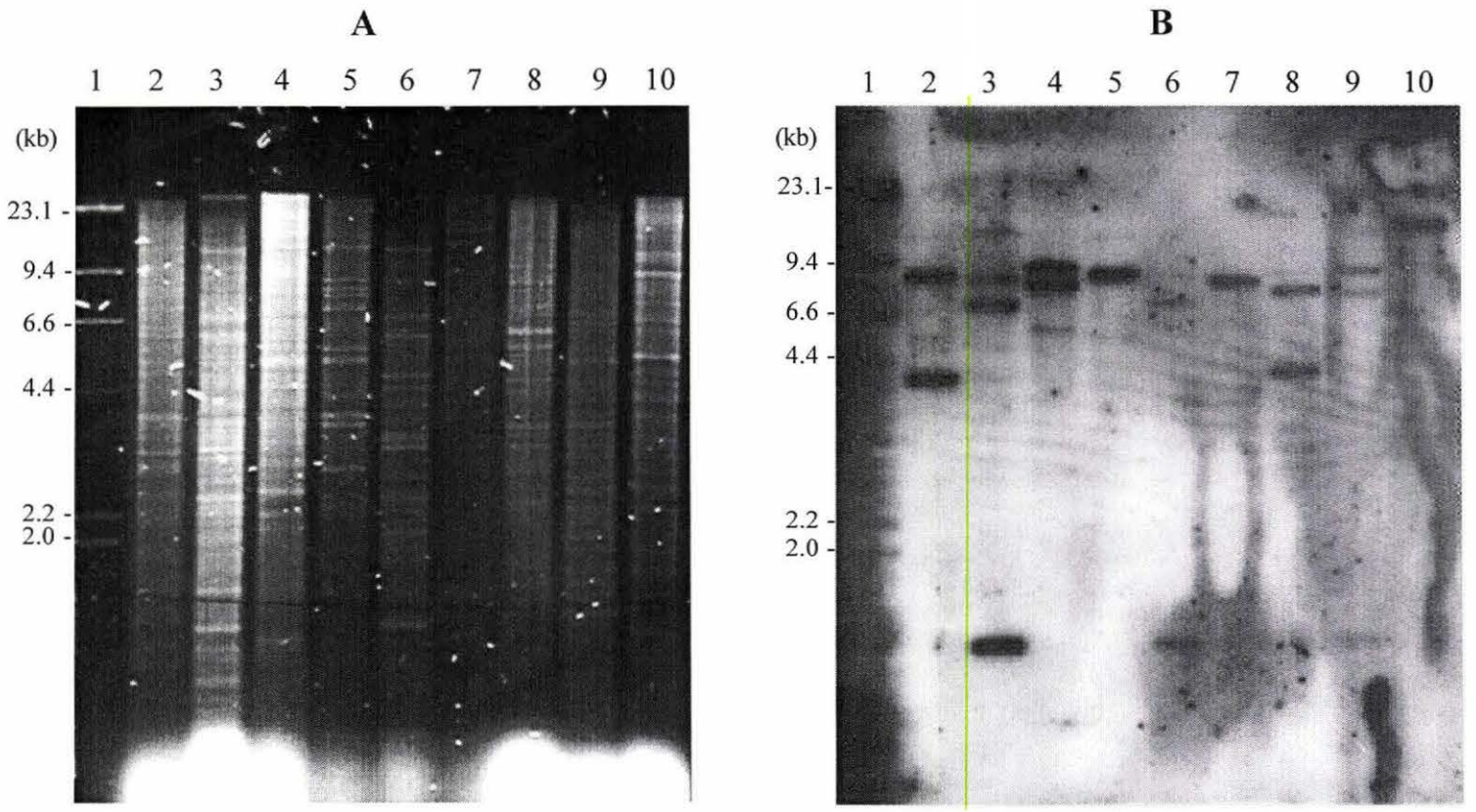


Figure 4.1 Southern blot analysis of the *prt*2 gene.

A. Southern blot of 2 µg of Lp1 (lanes 2-4), Lp19 (lanes 5-7) and E8 (lanes 8-10) genomic DNA cut with *EcoRI* (lanes 2, 5 and 8), *Hin*dIII (lanes 3, 6 and 9) and *Sst*I (lanes 4, 7 and 10). Lane 1 contains λ DNA digested with *Hin*dIII.

B. Autoradiograph of Figure 4.1A probed with [α^{32}P]dCTP-labelled pGH3 insert.

Table 4.1 Genomic DNA fragments hybridising to the pGH3 insert.

Lp1			Lp19			E8		
*Eco*RI	*Hin*dIII	*Sst*I	*Eco*RI	*Hin*dIII	*Sst*I	*Eco*RI	*Hin*dIII	*Sst*I
	14.3		13.8				14.0	14.9
8.6	8.5	9.2 8.1	8.8		8.1	8.6	8.7	9.2
	7.1			7.2			7.7	
5.3		5.7						
4.0						4.0		
	~1.2			~1.2			~1.2	

Fragment sizes are shown in kb.

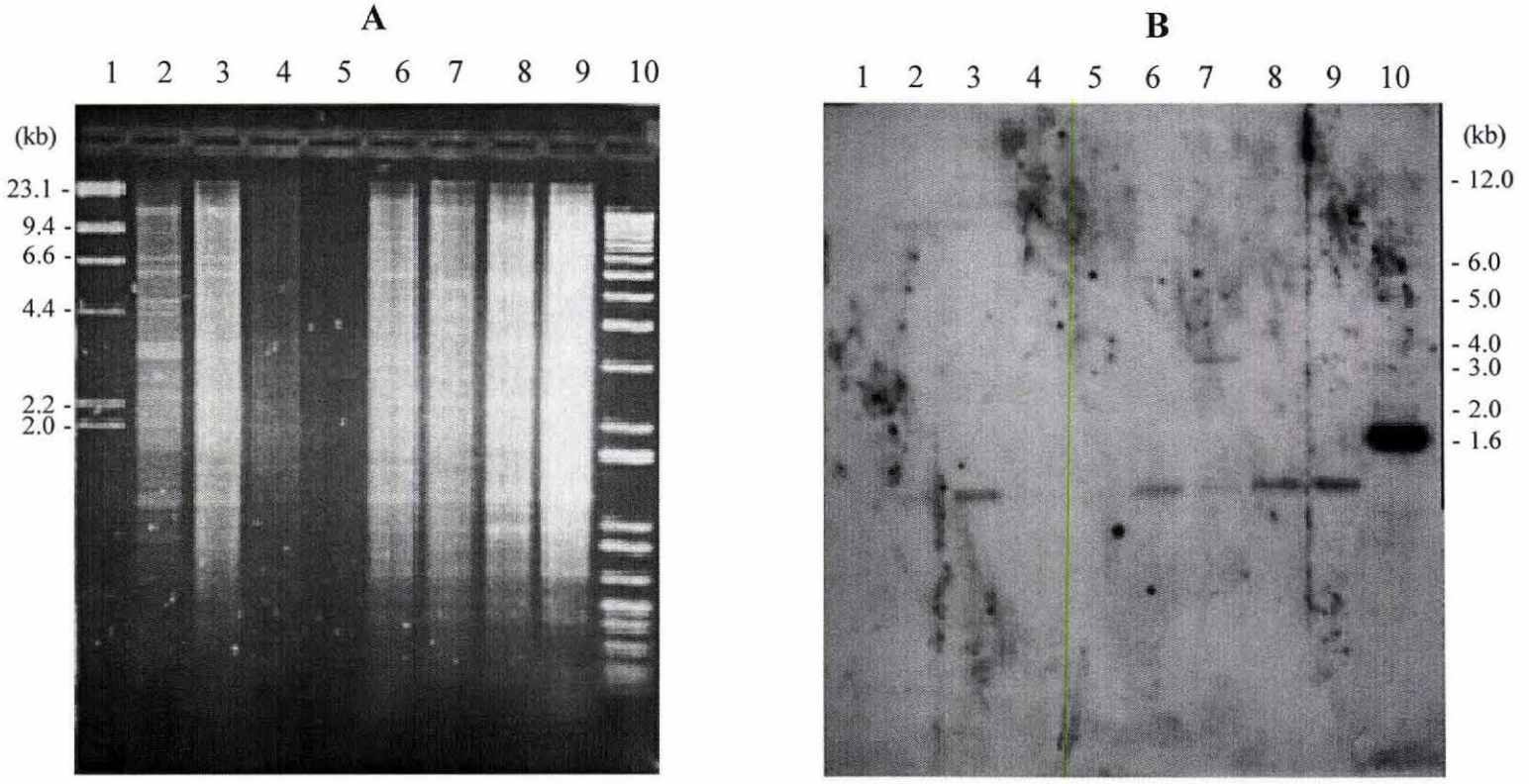


Figure 4.2 Southern blot analysis of endophyte genomic DNA homologous to the pGH3 insert

A. Southern blot of 1 µg of Lp19 (lane 2), Lp1 (lane 3), E8 (lane 4), AR510 (lane 6), Fp5 (lane 7), Tf2 (lane 8) and Tf13 (lane 9) genomic DNA cut with *Hin*dIII. Lane 1 contains λ DNA digested with *Hin*dIII, and lane 10 contains 1kb + ladder (GibcoBRL). No DNA was loaded in lane 5.

B. Autoradiograph of this blot probed with [α^{32}P]dCTP-labelled pGH3 insert.

Table 4.2 **Sizes of endophyte genomic *Hin*dIII fragments homologous to the pGH3 insert.**

	***Hin*dIII digest**
Lp19 (*N. lolii*)	1.2
Lp1 (LpTG-2)	1.2
E8 (*E. typhina*)	1.2
AR510 (FaTG-3)	1.2
Fp5 (*N. uncinatum*)	1.2, 3.7
Tf2 (*N. coenophialum*)	1.2
Tf13 (FaTG-2)	1.2

Fragment sizes are shown in kb.

4.2.2 Mapping of positive clones

4.2.2.1 Restriction digestion of positive λ clones

As in Section 3.2.2.1 four positive λ clones were digested with *Eco*RI, *Hin*dIII, or with both enzymes. All four clones (λMM3.1, λMM3.3, λMM3.5 and λMM3.6) had identical restriction patterns (Figure 4.3A). Fragment sizes for the restriction digests are shown in Table 4.3. All contained a large *Eco*RI fragment of approximately 20 kb (lanes 3, 6, 9 and 12). The *Hin*dIII digest contained a large fragment, a 4.3 kb fragment from the right λ arm, and a 1.2 kb fragment (lanes 5, 8, 11 and 14). Southern blotting with the same probe as in Section 4.1 showed the *Eco*RI hybridising fragment was large for all the λ clones (Figure 4.3B, lanes 3, 6, 9, and 12). All of the clones shared a 1.2 kb *Hin*dIII hybridising fragment (lanes 5, 8, 11 and 14) which was also seen in the *Eco*RI/*Hin*dIII digests (lanes 4, 7, 10 and 13). The 1.2 kb *Hin*dIII fragment is the same size as one of the hybridising bands seen in the genomic blots (Figures 4.1 and 4.2). λMM3 also contained a large *Hin*dIII fragment that hybridised to the probe. The size of this fragment is not consistent with the size of the fragments hybridising to genomic digests (Figure 4.1, lane 6). This suggests that these λ clones are either incompletely digested or rearranged. Further work will be required to distinguish these two options. As all of the clones appeared to be identical by means of their restriction digests, λMM3.3 was selected for further characterisation.

4.2.2.2 Restriction mapping of λMM3.3

λMM3.3 was digested with the same restriction enzymes as λMM30.2 and λMM30.4 (Section 3.2.2.2, Figure 4.4). Fragment sizes are shown in Table 4.4. A 2.0 kb hybridising *Bam*HI fragment from λMM3.3 (Figure 4.4B, lane 2), which was also seen in the *Bam*HI/*Eco*RI digest (lane 3), hybridised to the [α^{32}P]dCTP-labelled *Eco*RI/*Sst*I insert from pGH3. λMM3 appears to contain a large *Eco*RI fragment (lane 4), and also a large *Sst*I fragment (lane 8), both of which hybridised to the probe. For the *Hin*dIII digest, two fragments hybridised to the probe. One fragment was 1.2 kb in size, while the other fragment was very large. The same bands were also seen in the *Eco*RI/*Hin*dIII digest (lane 5). The overall map for λMM3.3 showed a large insert of at least 20 kb. The insert contains at least one *Bam*HI site, and two *Hin*dIII sites. The enzymes chosen for mapping were not particularly informative for this clone.

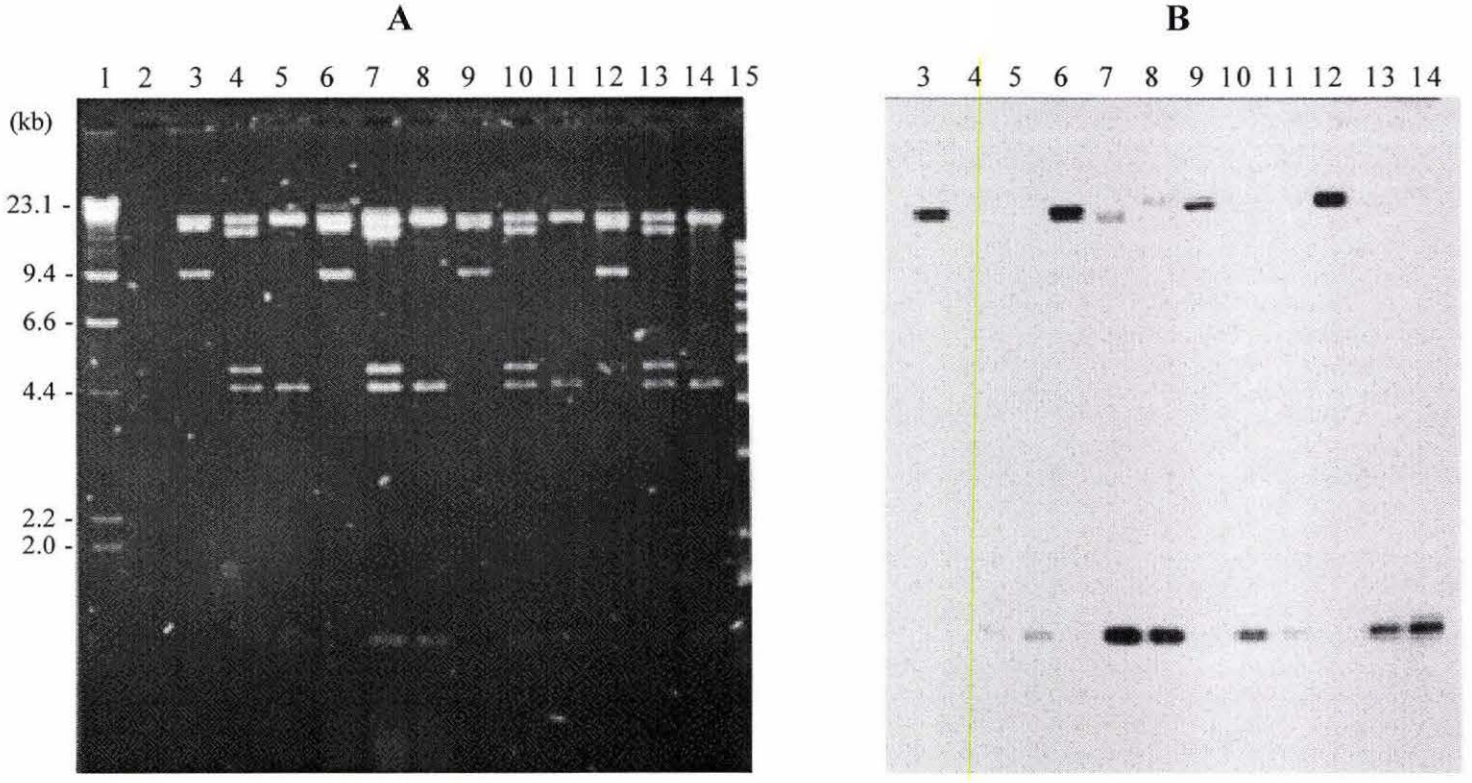


Figure 4.3 Mapping of positive λ clones and fragments homologous to the pGH3 insert.

A: Southern blot of *Eco*RI, *Eco*RI/*Hin*dIII and *Hin*dIII digested λ clones. Clones λMM3.1 (lanes 3-5), λMM3.3 (lanes 6-8), λMM3.5 (lanes 9-11), and λMM3.6 (lanes 12-14). Lane 1 contains DNA/*Hin*dIII ladder; lane 15 contains 1 kb ladder (GibcoBRL).

B: Autoradiograph of the Southern blot shown in Fig. 4.3A probed with [α^{32}P]dCTP-labelled pGH3 insert.

Table 4.3 **Restriction fragments of λ clones hybridising to the pGH3 insert.**

	*Eco*RI	*Eco*RI/*Hin*dIII	*Hin*dIII
λMM3.1	>20.0*	>20.0	>20.0
	20.0	20.0	
	9.0	5.0	
		4.5	4.5
		1.2*	1.2*
λMM3.3	>20.0*	>20.0*	>20.0*
	20.0	20.0	
	9.0	5.0	
		4.5	4.5
		1.2*	1.2*
λMM3.5	>20.0*	>20.0	>20.0
	20.0	20.0	
	9.0	5.0	
		4.5	4.5
		1.2*	1.2*
λMM3.6	>20.0*	>20.0	>20.0
	20.0	20.0	
	9.0	5.0	
		4.5	4.5
		1.2 *	1.2*

Fragment sizes (in kb) generated by single and double enzyme digestions of λ clones obtained by library screening with the pGH3 insert. Fragments that hybridised to the pGH3 insert are indicated by an asterisk.

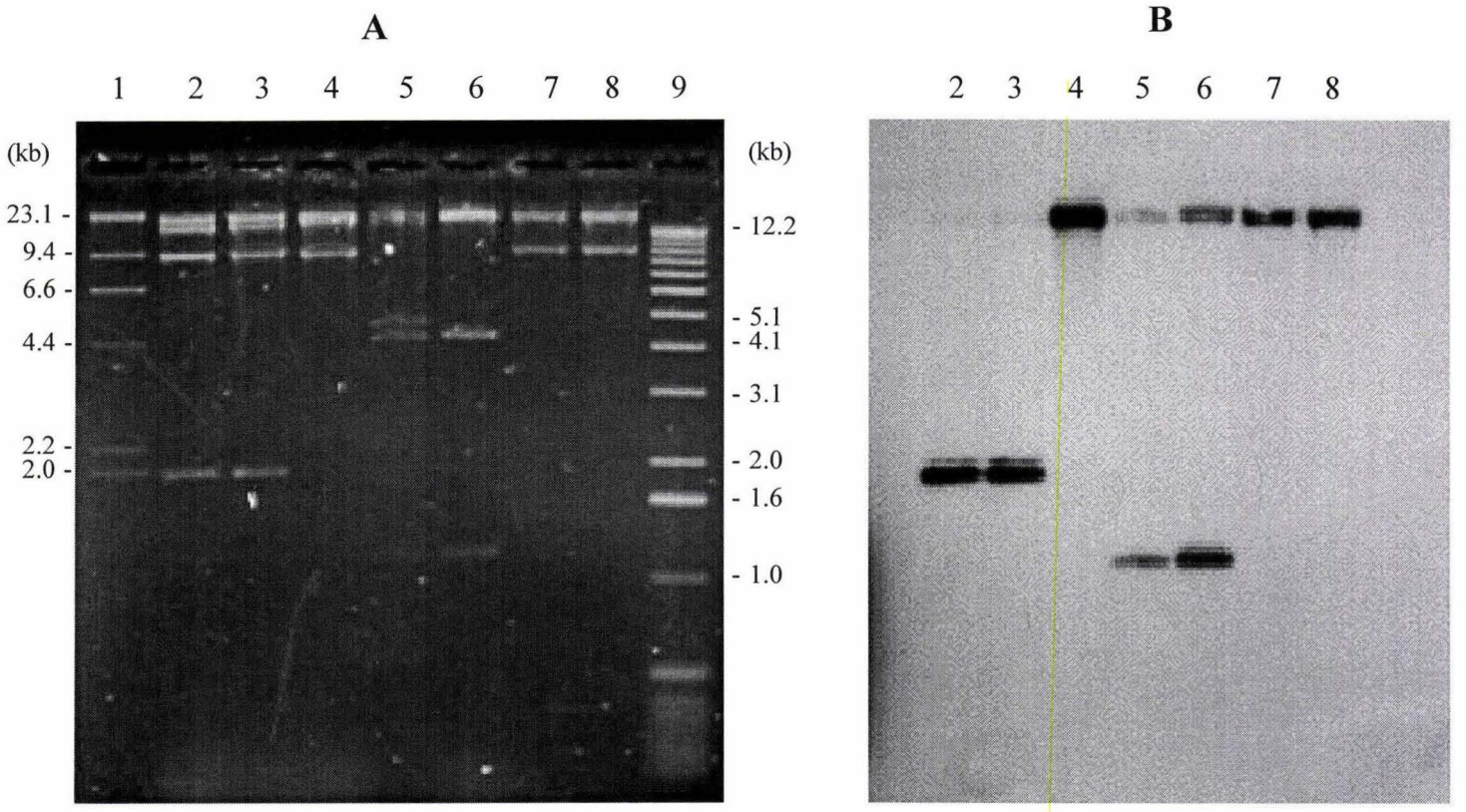


Figure 4.4 Southern analysis of λMM3.3

A: Southern blot of λMM3.3 digested with *Bam*HI (lane 2), *Bam*HI/*Eco*RI (lane 3), *Eco*RI (lane 4), *Eco*RI/*Hin*dIII (lane 5), *Hin*dIII (lane 6), *Eco*RI/*Sst*I (lane 7) and *Sst*I (lane 8). Lane 1 contains λ DNA/*Hin*dIII ladder; lane 9 contains 1 kb ladder (GibcoBRL).

B: Autoradiograph of the Southern blot shown in Fig. 4.4A probed with [α^{32}P]dCTP-labelled pGH3 insert.

Table 4.4 Fragment sizes of λMM3.3 restriction digests

*Bam*HI	*Bam*HI/ *Eco*RI	*Eco*RI	*Eco*RI/ *Hin*dIII	*Hin*dIII	*Eco*RI/ *Sst*I	*Sst*I
> 20.0	> 20.0	> 20.0*	> 20.0*	> 20.0*	> 20.0*	> 20.0*
20.0	20.0	20.0	20.0	20.0	20.0	20.0
9.0	9.0	9.0			9.0	9.0
			5.0			
			4.5	4.5		
2.0*	2.0*					
			1.2*	1.2*		

Fragment sizes are shown in kb. Fragments hybridising to the pGH3 insert are indicated by an asterisk.

4.2.2.3 Subcloning

After looking at the restriction patterns of λMM3.3, it was decided to subclone the 2.0 kb *Bam*HI hybridising fragment. The 2 kb *Bam*HI fragment from λMM3.3 was purified from a Seaplaque agarose gel as described in Section 2.5.1, before ligation into pUC118 (Appendix A1.5) to create the pMM7 vector (Figure 4.5B). The pMM7 insert was analysed by restriction mapping, revealing two *Hin*dIII sites in the insert.

4.3 SEQUENCING

4.3.1 Sequencing of pMM7

The pMM7 plasmid was sequenced initially with the forward and reverse primer. Sequence from the forward primer showed some homology to the protease sequence, while sequence from the reverse primer showed no notable homology. Based on homology with other subtilisin-like protease genes, not all of the *prt*2 coding region is contained the 2kb BamHI in pMM7. The sequence stops towards the 3' end of the coding region of *prt*2 (Figures 4.5A and 4.6). The 5' region of the *prt*2 gene has been difficult to sequence, which could be due to poor primer design or repetitive sequences. Despite these difficulties, approximately 1 kb of the *prt*2 coding region has been sequenced. Sequence data is shown in Appendix A3.2.

4.3.2 Analysis of the *prt*2 coding region

The coding region of *prt*2 coding region so far suggests the presence of at least three potential introns; one from before the start of the continuous sequence to 66, one from 253 to 335, and one from 849 to 919 (Figure 4.6). The two complete introns are 83 and 71 bp respectively, similar in size to the 73 bp intron seen in the *prt*1 gene. The position of all three introns is supported by homology with other fungal subtilisin-like proteases, and the position of introns 2 and 3 is also supported by frame shifts. Both introns 2 and 3 have similar 5' splice sites, with GTATGT for the intron 2 and GTAAGT for intron 3. All three introns appear to have the same 3' splice site, CAG.

Introns 1 and 2 contain good matches for the lariat consensus sequence; TACTAACT for intron 1 and AACTAACC for intron 2 compared with the fungal consensus sequence of YGCTAACN. The closest match to a lariat consensus sequence in the second intron is

Figure 4.5 Subcloning and sequencing of the *prt*2 gene

A. Sequencing strategy for the *prt*2 gene

The pMM7 vector contained a 2 kb hybridising *Bam*HI fragment from λMM3.3. The region where the pGH3 probe anneals is shown in blue. Restriction sites are shown in purple text. Regions which share homology with the coding regions of other subtilisin-like protease genes are shown in by the green arrows and boxes. Exon numbers are assigned based on homology with other fungal subtilisin-like protease genes.

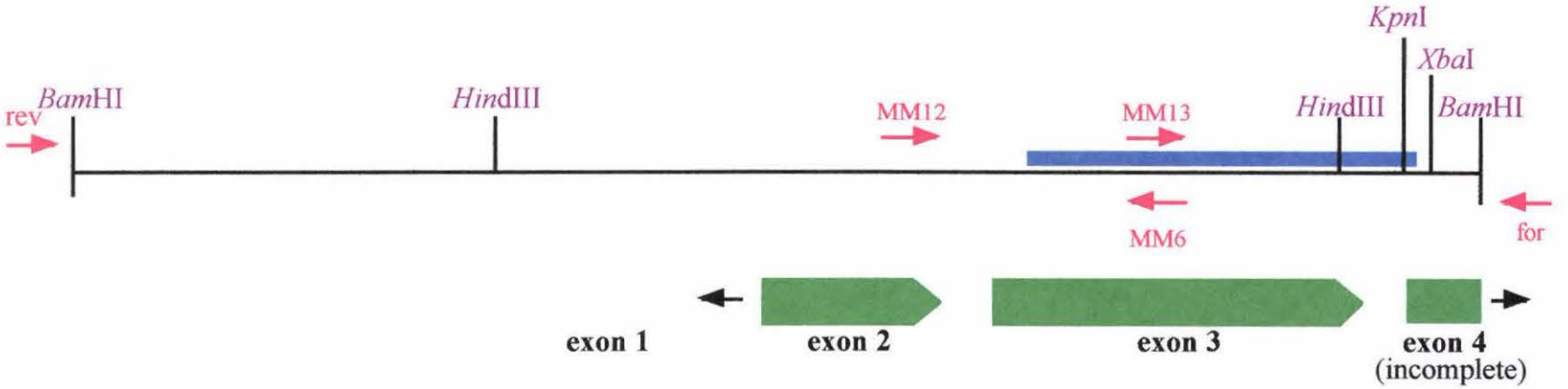


Figure 4.5 Subcloning and sequencing of the *prt*2 gene

B. Restriction map of pMM7

Restriction map of pMM7, which contains a 2 kb *Bam*HI fragment from λMM3.3 subcloned into pUC118.

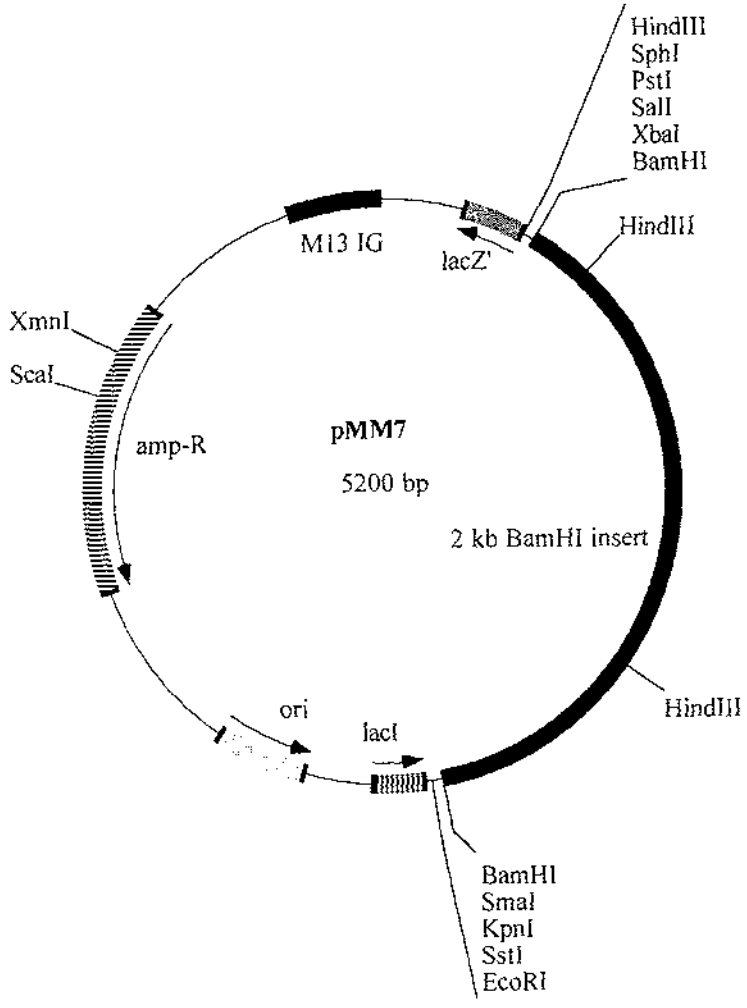
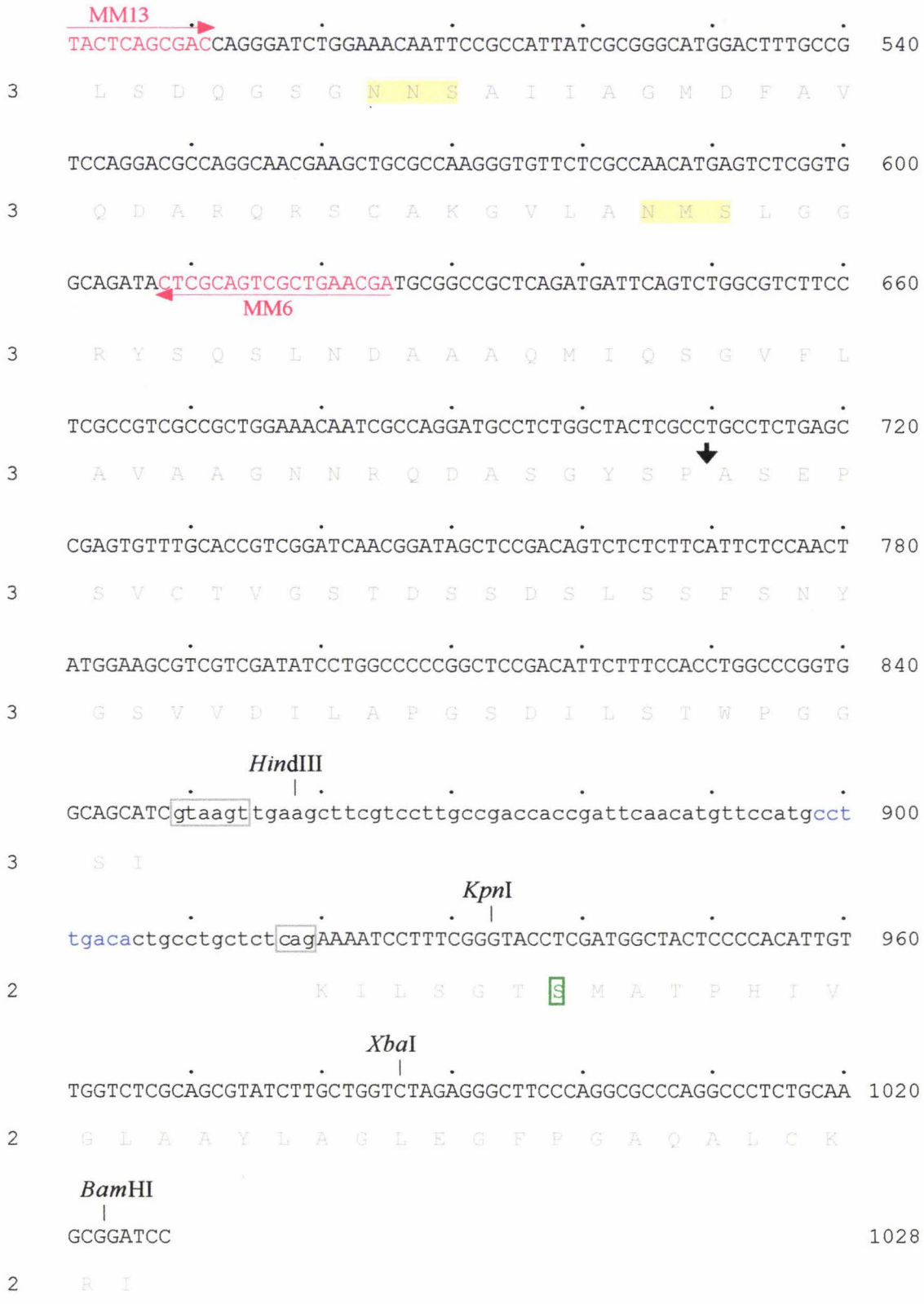


Figure 4.6 Partial sequence of the *prt2* gene

Currently 1028 bases of the *prt2* gene have been sequenced, all of which were shown to be within the coding region by Blast-X searches and homology to other subtilisin-like proteases. Primers are shown in red text, with red arrows indicating the direction of primer extension. Proposed intron sequences are shown in lower case letters, with 5' and 3' splice sites in grey boxes. Matches for lariat consensus sequences are shown in blue text.

The translation in grey uppercase text directly below the nucleotide sequence, with the frame indicated at the beginning of each line of the translation. Conserved active site residues are shown in green text within green boxes, and putative N-glycosylation sites are highlighted in yellow. The conserved autolysis site is indicated by a black arrow.


MM13
TACTCAGCGACCAGGGATCTGGAAACAATTCCGCCATTATCGCGGGCATGGACTTTGCCG 540
3 L S D Q G S G N N S A I I A G M D F A V
TCCAGGACGCCAGGCAACGAAGCTGCGCCAAGGGTGTTCTCGCCAACATGAGTCTCGGTG 600
3 Q D A R Q R S C A K G V L A N M S L G G
GCAGATACTCGCAGTCGCTGAACGATGCGGCCGCTCAGATGATTCAGTCTGGCGTCTTCC 660
MM6
3 R Y S Q S L N D A A A Q M I Q S G V F L
TCGCCGTCGCCGCTGGAAACAATCGCCAGGATGCCTCTGGCTACTCGCCTGCCTCTGAGC 720
3 A V A A G N N R Q D A S G Y S P A S E P
CGAGTGTTTGCACCGTCGGATCAACGGATAGCTCCGACAGTCTCTCTTCATTCTCCAACT 780
3 S V C T V G S T D S S D S L S S F S N Y
ATGGAAGCGTCGTCGATATCCTGGCCCCCGGCTCCGACATTCTTTCCACCTGGCCCGGTG 840
3 G S V V D I L A P G S D I L S T W P G G
HindIII
GCAGCATCgtaagttgaagcttcgtccttgccgaccaccgattcaacatgttccatgcct 900
3 S I
KpnI
tgacactgcctgctctcagAAAATCCTTTCGGGTACCTCGATGGCTACTCCCCACATTGT 960
2 K I L S G T S M A T P H I V
XbaI
TGGTCTCGCAGCGTATCTTGCTGGTCTAGAGGGCTTCCCAGGCGCCCAGGCCCTCTGCAA 1020
2 G L A A Y L A G L E G F P G A Q A L C K
BamHI
GCGGATCC 1028
2 R I

CCTTGACA, which is similar to lariat sequences from other fungal species (Gurr et al., 1987).

4.3.3 Features of the Prt2 peptide

The Prt2 peptide sequence, based on Blast-X searches and homology to other subtilisin-like proteases, currently consists of 269 amino acid residues (Figure 4.6). The first and fourth exons are currently incomplete (Figure 4.5A). Based on homology with other subtilisin-like proteases, the region encoding the signal and pro-peptides has not yet been sequenced; all of the sequence shown in Figure 4.6 encodes the mature region of the peptide sequence. Prt2 contains all of the conserved active site residues necessary for enzyme function, along with the conserved autolysis site (PASEPS) that was also found in Prt1. Prt2 shared 61% identity with the peptide sequence of Pr1, a subtilisin-like protease from the entomopathogenic fugnus *Metarhizium anisopliae*.

4.3.3.1 Comparison of Prt2 to other fungal subtilisin-like proteases

The 269 amino acid residue sequence of Prt2 was aligned with closely related fungal subtilisin-like protease peptide sequences (Figure 4.7). This alignment, which was based on Clustal-W alignments, showed that Prt2 shared large regions of identity with similar proteins.

Prt2 was also aligned with the two other endophyte subtilisin-like proteases, Prt1 and At1. All three introns proposed for the *prt*2 gene are found in the same position in the At1 gene, but only the first intron is found in the *prt*1 gene. The intron found in the region of the *prt*1 gene encoding the proprotein region is conserved with the At1 gene; at the moment, not enough sequence is available to determine if it is also conserved in *prt*2. Overall, although all of the endophyte proteases have very high levels of similarity, the Prt2 peptide sequence appears to be more similar to At1 than to Prt1 (Figure 4.8). FastA searches using Wisconsin Package 9.1 (GCG) that were performed to confirmed this showed that the At protease shared 53.9 % identity with Prt2, but only 34.8% identity with Prt1. These results may be slightly misleading as only the more conserved regions of the *prt*2 coding regions have been sequenced to date.

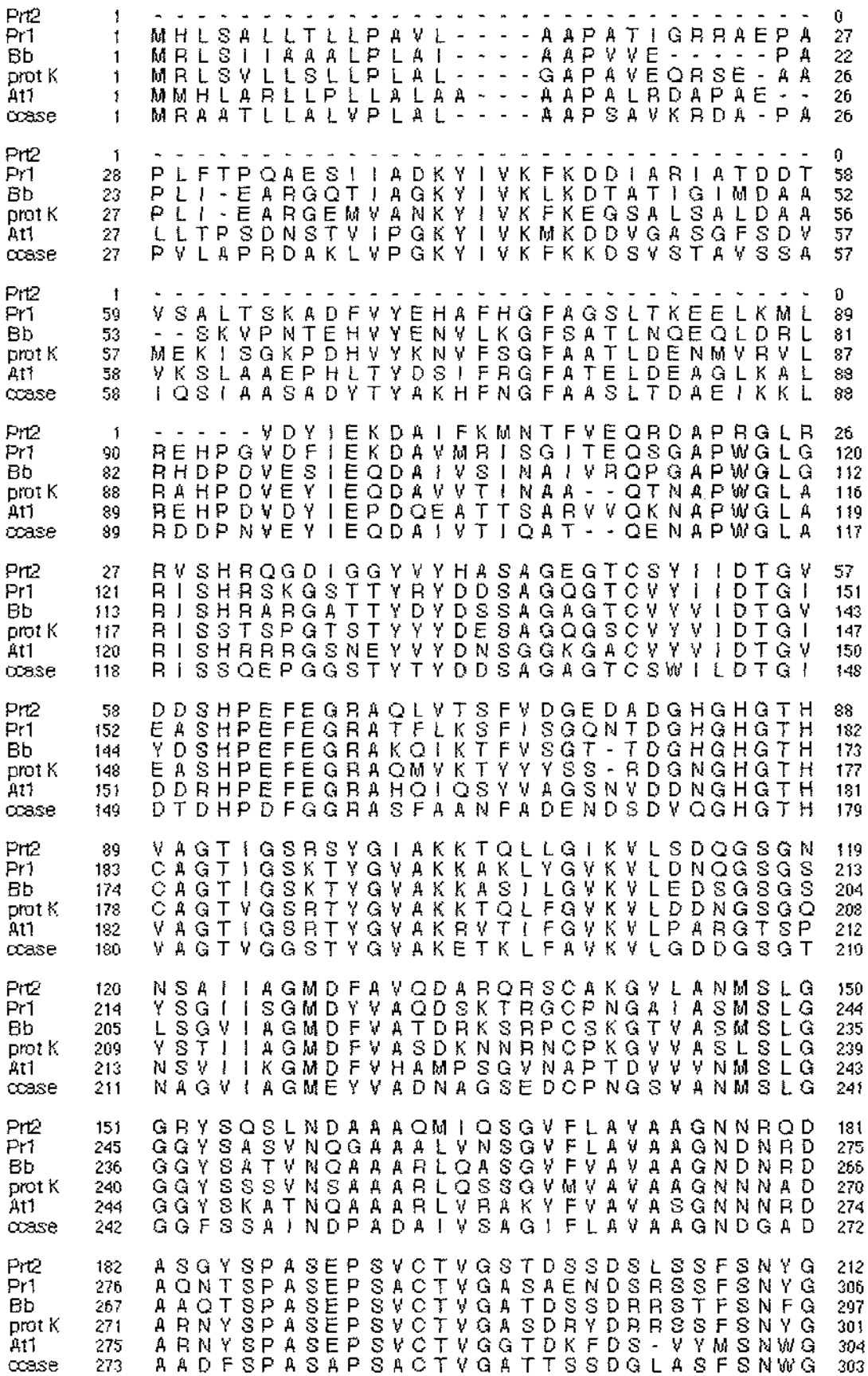
Figure 4.7 Alignment of Prt2 with other fungal subtilisin-like genes

The Prt2 peptide sequence was aligned with peptide sequences from other fungi to which it showed high degrees of similarity in Blast-P searches. Amino acid residues are indicated in grey boxes where 60% of the sequences have an identical amino acid at that position.

A key to the naming is given below:

Name	Species	Accession number
Prt2	*N. lolii* Lp19	this study (Figure 4.6)
Pr1	*Metarhizium anisopliae*	Swiss Prot accession P29138
Bb	*Beauveria bassiana*	GenBank accession AAD29255.1
protK	*Tritirachium album* Limber	SwissProt accession P06873
At1	*Epichloë typhina*	GenBank accession L76740.1
ccase[1]	*Acremonium chrysogenum*	EMBL accession AJ238108.1

[1] This sequence encodes an enzyme very similar to subtilisin-like proteases, but this enzyme functions as a non-specific esterase, cephalosporin C acetohydrolase.


Prt2 1 ---------- 0
Pr1 1 MHLSALLTLLPAVL----AAPATIGRRAEPA 27
Bb 1 MRLSIIAAALPLAI----AAPVVE-----PA 22
prot K 1 MRLSVLLSLLPLAL----GAPAVEQRSE-AA 26
At1 1 MMHLARLLPLLALAA---AAPALRDAPAE-- 26
ccase 1 MRAATLLALVPLAL----AAPSAVKRDA-PA 26
Prt2 1 ---------- 0
Pr1 28 PLFTPQAESIIADKYIVKFKDDIARIATDDT 58
Bb 23 PLI-EARGQTIAGKYIVKLKDTATIGIMDAA 52
prot K 27 PLI-EARGEMVANKYIVKFKEGSALSALDAA 56
At1 27 LLTPSDNSTVIPGKYIVKMKDDVGASGFSDV 57
ccase 27 PVLAPRDAKLVPGKYIVKFKKDSVSTAVSSA 57
Prt2 1 ---------- 0
Pr1 59 VSALTSKADFVYEHAFHGFAGSLTKEELKML 89
Bb 53 --SKVPNTEHVYENVLKGFSATLNQEQLDRL 81
prot K 57 MEKISGKPDHVYKNVFSGFAATLDENMVRVL 87
At1 58 VKSLAAEPHLTYDSIFRGFATELDEAGLKAL 88
ccase 58 IQSIAASADYTYAKHFNGFAASLTDAEIKKL 88
Prt2 1 -----VDYIEKDAIFKMNTFVEQRDAPRGLR 26
Pr1 90 REHPGVDFIEKDAVMRISGITEQSGAPWGLG 120
Bb 82 RHDPDVESIEQDAIVSINAIVRQPGAPWGLG 112
prot K 88 RAHPDVEYIEQDAVVTINAA--QTNAPWGLA 116
At1 89 REHPDVDYIEPDQEATTSARVVQKNAPWGLA 119
ccase 89 RDDPNVEYIEQDAIVTIQAT--QENAPWGLA 117
Prt2 27 RVSHRQGDIGGYVYHASAGEGTCSYIIDTGV 57
Pr1 121 RISHRSKGSTTYRYDDSAGQGTCVYIIDTGI 151
Bb 113 RISHRARGATTYDYDSSAGAGTCVYVIDTGV 143
prot K 117 RISSTSPGTSTYYYDESAGQGSCVYVIDTGI 147
At1 120 RISHRRRGSNEYVYDNSGGKGACVYVIDTGV 150
ccase 118 RISSQEPGGSTYTYDDSAGAGTCSWILDTGI 148
Prt2 58 DDSHPEFEGRAQLVTSFVDGEDADGHGHGTH 88
Pr1 152 EASHPEFEGRATFLKSFISGQNTDGHGHGTH 182
Bb 144 YDSHPEFEGRAKQIKTFVSGT-TDGHGHGTH 173
prot K 148 EASHPEFEGRAQMVKTYYYSS-RDGNGHGTH 177
At1 151 DDRHPEFEGRAHQIQSYVAGSNVDDNGHGTH 181
ccase 149 DTDHPDFGGRASFAANFADENDSDVQGHGTH 179
Prt2 89 VAGTIGSRSYGIAKKTQLLGIKVLSDQGSGN 119
Pr1 183 CAGTIGSKTYGVAKKAKLYGVKVLDNQGSGS 213
Bb 174 CAGTIGSKTYGVAKKASILGVKVLEDSGSGS 204
prot K 178 CAGTVGSRTYGVAKKTQLFGVKVLDDNGSGQ 208
At1 182 VAGTIGSRTYGVAKRVTIFGVKVLPARGTSP 212
ccase 180 VAGTVGGSTYGVAKETKLFAVKVLGDDGSGT 210
Prt2 120 NSAIIAGMDFAVQDARQRSCAKGVLANMSLG 150
Pr1 214 YSGIISGMDYVAQDSKTRGCPNGAIASMSLG 244
Bb 205 LSGVIAGMDFVATDRKSRPCSKGTVASMSLG 235
prot K 209 YSTIIAGMDFVASDKNNRNCPKGVVASLSLG 239
At1 213 NSVIIKGMDFVHAMPSGVNAPTDVVVNMSLG 243
ccase 211 NAGVIAGMEYVADNAGSEDCPNGSVANMSLG 241
Prt2 151 GRYSQSLNDAAAQMIQSGVFLAVAAGNNRQD 181
Pr1 245 GGYSASVNQGAAALVNSGVFLAVAAGNDNRD 275
Bb 236 GGYSATVNQAAAARLQASGVFVAVAAGNDNRD 266
prot K 240 GGYSSSVNSAAAARLQSSGVMVAVAAGNNNAD 270
At1 244 GGYSKATNQAAAARLVRAKYFVAVASGNNNRD 274
ccase 242 GGFSSAINDPADAIVSAGIFLAVAAGNDGAD 272
Prt2 182 ASGYSPASEPSVCTVGSTDSSDSLSSFSNYG 212
Pr1 276 AQNTSPASEPSACTVGASAENDSRSSFSNYG 306
Bb 267 AAQTSPASEPSVCTVGATDSSDRRSTFSNFG 297
prot K 271 ARNYSPASEPSVCTVGASDRYDRRSSFSNYG 301
At1 275 ARNYSPASEPSVCTVGGTDKFDS-VYMSNWG 304
ccase 273 AADFSPASAPSACTVGATTSSDGLASFSNWG 303

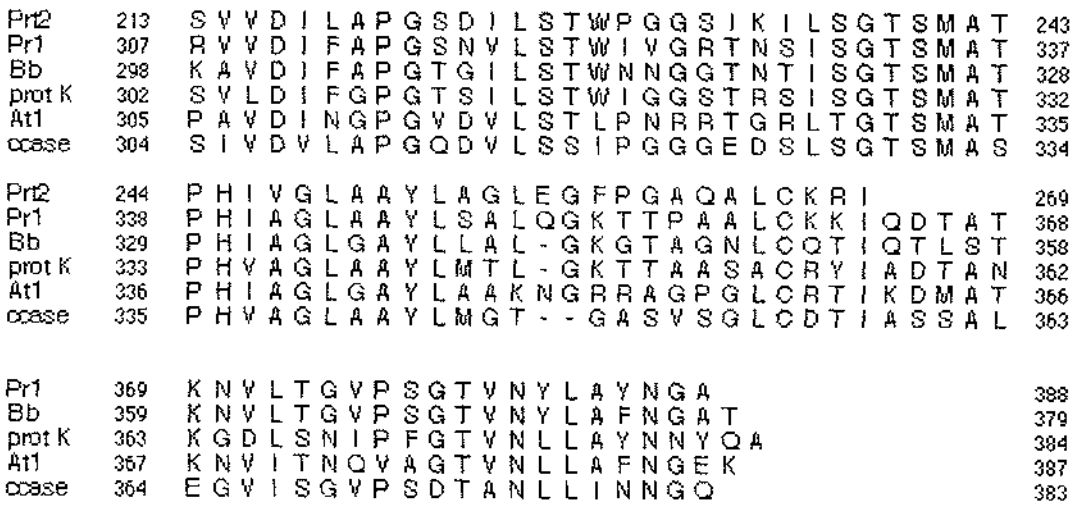

Prt2 213 S V V D I L A P G S D I L S T W P G G S I K I L S G T S M A T 243
Pr1 307 R V V D I F A P G S N V L S T W I V G R T N S I S G T S M A T 337
Bb 298 K A V D I F A P G T G I L S T W N N G G T N T I S G T S M A T 328
prot K 302 S V L D I F G P G T S I L S T W I G G S T R S I S G T S M A T 332
At1 305 P A V D I N G P G V D V L S T L P N R R T G R L T G T S M A T 335
ccase 304 S I V D V L A P G Q D V L S S I P G G G E D S L S G T S M A S 334
Prt2 244 P H I V G L A A Y L A G L E G F P G A Q A L C K R I 269
Pr1 338 P H I A G L A A Y L S A L Q G K T T P A A L C K K I Q D T A T 368
Bb 329 P H I A G L G A Y L L A L - G K G T A G N L C Q T I Q T L S T 358
prot K 333 P H V A G L A A Y L M T L - G K T T A A S A C R Y I A D T A N 362
At1 336 P H I A G L G A Y L A A K N G R R A G P G L C R T I K D M A T 366
ccase 335 P H V A G L A A Y L M G T - - G A S V S G L C D T I A S S A L 363
Pr1 369 K N V L T G V P S G T V N Y L A Y N G A 388
Bb 359 K N V L T G V P S G T V N Y L A F N G A T 379
prot K 363 K G D L S N I P F G T V N L L A Y N N Y Q A 384
At1 367 K N V I T N Q V A G T V N L L A F N G E K 387
ccase 364 E G V I S G V P S D T A N L L I N N G Q 383

Figure 4.8 Alignment of At1, Prt1 and Prt2 subtilisin-like proteases

The Prt1 and Prt2 peptide sequences from *N. lolii* Lp19 were aligned with the At1 peptide sequence (*Epichloë typhina,* GenBank locus APEPROT accession L76740.1). Amino acid residues shaded in grey show where 60% of the residues at a given site share homology.

```
prot1 (N. lolii)      1  -MLNYKNLYLTAAAALASQAIAAPTGPDAGNAKIQAAQ  37
prot2 (N. lolii)      1  -----------------------------------------  0
At1 (E. typhina)      1  MMHLARLLPLLALAAAAPALRDAP------AELLTPSD  32
prot1 (N. lolii)      1  -MLNYKNLYLTAAAALASQAIAAPTGPDAGNAKIQAAQ  37

prot1 (N. lolii)     38  GGQVIPGKFIYTLKPGSKPAYLESHMRWYNGVHARASG  75
prot2 (N. lolii)      1  -----------------------------------------  0
At1 (E. typhina)     33  NSTYIPGKYIYKMKDD----YGASGFSDYYKSLAAEPH  66
prot1 (N. lolii)     38  GGQVIPGKFIYTLKPGSKPAYLESHMRWYNGVHARASG  75

prot1 (N. lolii)     76  DEAIKGYETMLDGIYGFMGYYGSFSEAVLAQIKAHPDV  113
prot2 (N. lolii)      1  ----------------------------------------V  1
At1 (E. typhina)     67  LT--------YDSI--FRGFATELDEAGLKALREHPDV  94
prot1 (N. lolii)     76  DEAIKGYETMLDGIYGFMGYYGSFSEAVLAQIKAHPDV  113

prot1 (N. lolii)    114  EAYEQDKIWTLDWITDDQQLEARDDDKEPPSSGGGSNF  151
prot2 (N. lolii)      2  DYIEKDAIFKMN----------------------TF  15
At1 (E. typhina)     95  DYIEPDQEATT----------------------SAR  108
prot1 (N. lolii)    114  EAYEQDKIWTLDWITDDQQLEARDDDKEPPSSGGGSNF  151

prot1 (N. lolii)    152  IQQKNATWGLGSISHRAPYATEYGYQESAGKDTYAYYI  189
prot2 (N. lolii)     16  YEQRDAPRGLRRYSHRQGDIGGYYYHASAGEGTCSYII  53
At1 (E. typhina)    109  YYQKNAPWGLARISHRRRGSNEYYYDNSGGKGACYYYI  146
prot1 (N. lolii)    152  IQQKNATWGLGSISHRAPYATEYGYQESAGKDTYAYYI  189

prot1 (N. lolii)    190  DTGIRTTHEEFEGRASHAWSAYLTR-TDNYGHGTHYAG  226
prot2 (N. lolii)     54  XTGYDDSHPEFEGRAQLYTSFYDGEDADGHGHGTHYAG  91
At1 (E. typhina)    147  DTGYDDRHPEFEGRAHQIQSYYAGSNYDDNGHGTHYAG  184
prot1 (N. lolii)    190  DTGIRTTHEEFEGRASHAWSAYLTR-TDNYGHGTHYAG  226

prot1 (N. lolii)    227  TIGGKTYGYAKNAKLLAYKIFNS-RSSSTSYILAGYNW  263
prot2 (N. lolii)     92  TIGSRSYGIAKKTQLLGIKYLSDQGSGNNSAIIAGMDF  129
At1 (E. typhina)    185  TIGSRTYGYAKRYTIFGYKYLPARGTSPNSYIIKGMDF  222
prot1 (N. lolii)    227  TIGGKTYGYAKNAKLLAYKIFNS-RSSSTSYILAGYNW  263

prot1 (N. lolii)    264  AYNDIYRKGRTKRAAINMSLGGPKSTAFNTAYERASAS  301
prot2 (N. lolii)    130  AYQDARQRSCAKGYLANMSLGGRYSQSLNDAAAQMIQS  167
At1 (E. typhina)    223  YHAMPSGYNAPTDYYYNMSLGGGYSKATNQAAARLYRA  260
prot1 (N. lolii)    264  AYNDIYRKGRTKRAAINMSLGGPKSTAFNTAYERASAS  301

prot1 (N. lolii)    302  GYLSIIAAGNEAQDASNYSPASAPSAITYAAINRDWTL  339
prot2 (N. lolii)    168  GYFLAYAAGNNRQDASGYSPASEPSYCTYGSTDSSDSL  205
At1 (E. typhina)    261  KYFYAYASGNNNRDARNYSPASEPSYCTYGGTDKFDS-  297
prot1 (N. lolii)    302  GYLSIIAAGNEAQDASNYSPASAPSAITYAAINRDWTL  339

prot1 (N. lolii)    340  ASYSNFGSYYDICAPGSNITSAWNTGDSSEKTISGTSM  377
prot2 (N. lolii)    206  SSFSNYGSYYDILAPGSDILSTWPGG--SIKILSGTSM  241
At1 (E. typhina)    298  YYMSNWGPAYDINGPGYDYLSTLPNR--RTGRLTGTSM  333
prot1 (N. lolii)    340  ASYSNFGSYYDICAPGSNITSAWNTGDSSEKTISGTSM  377

prot1 (N. lolii)    378  ATPHYYGLALYAISYDGATGYDGYTKHLLSTATKDKYA  415
prot2 (N. lolii)    242  ATPHIYGLAAYLAGLEGFPGAQALCKRI            269
At1 (E. typhina)    334  ATPHIAGLGAYLAAKNGRRAGPGLCRTIKDMATKNYIT  371
prot1 (N. lolii)    378  ATPHYYGLALYAISYDGATGYDGYTKHLLSTATKDKYA  415

prot1 (N. lolii)    416  GDTRGSPNLIGNNNNSYQK                      434

At1 (E. typhina)    372  NQYAGTYNLLAFNGEK                         387
prot1 (N. lolii)    416  GDTRGSPNLIGNNNNSYQK                      434
```

CHAPTER FIVE

DISCUSSION

5.1 SUBTILISIN-LIKE PROTEASES IN ENDOPHYTES

The first subtilisin-like protease isolated from a grass endophyte was isolated by Reddy et al., 1996. In this study, the At1 protease constituted 1 to 2 % of the total leaf-sheath protein. The high levels of At1 expression suggested subtilisin-like proteases could play an important role in the interaction between endophytes and their hosts. Northern analysis also suggested *N. lolii* endophytes growing in perennial ryegrass (*Lolium perenne*) could express subtilisin-like proteases at lower levels than the *E. typhina* in it's interaction with *Poa ampla*. The At1 transcript could be detected in total RNA from the *E. typhina/P. ampla* interaction, but could only be detected in poly(A)$^+$ mRNA from the *N. lolii/L. perenne* interaction.

Using degenerate primers based on highly conserved regions of the At1 gene, Hotter (Section 1.5.2) found two different PCR products with homology to subtilisin-like proteases. The first, GH3, showed homology to proteases from *Paecilomyces lilacinus*, an entomopathogenic fungus. The second PCR product, GH30, showed homology to an *Aspergillus fumigatus* subtilisin-like protease.

Initially, Southern blots confirmed the presence of sequences homologous to the pGH3 and pGH30 inserts in Lp19 genomic DNA, and also in Lp1 and E8 genomic DNA (Section 3.1 and 4.1). The pGH30 insert hybridised to 3.3 kb *Eco*RI, 10.7 kb *Hin*dIII, and 20.7 kb *Sst*I fragments respectively. The pGH3 insert hybridised to 8.8 kb *Eco*RI fragment, to 7.2 kb and <1.6 kb *Eco*RI fragments, and an 8.2 kb *Sst*I fragment.

Southern analysis confirmed that sequences homologous to these PCR products were found in endophytes from all taxonomic groups (Sections 3.1 and 4.1). The *Hin*dIII hybridising pattern for *prt*2 (GH3) was identical for all of the endophytes. For *prt*1 (GH30) there was considerable size variation in *Hin*dIII hybridising fragments, which could be due to variation in the minisatellite downstream of the *prt*1 gene (Section 3.3.7). Even under stringent hybridisation conditions, there were other weakly hybridising bands for both the GH3 and GH30 probes (Section 3.1 and 4.1). This could indicate either the probes hybridised weakly to the other subtilisin-like protease gene, or there are other subtilisin-like proteases in the endophyte strains examined.

Southern blots probed with the GH30 and GH3 probe supported the evolution of asexual endophytes by interspecific hybridisation. Southern results for the *prt*1 gene (Figures 3.1 and

3.2) supported hybrid origins for *N. coenophialum* (FaTG-1), FaTG-2, FaTG-3 and LpTG-2 taxonomic groups, which are all thought to be interspecific hybrids (Schardl et al., 1994; Tsai et al., 1994). The hybrid origin of LpTG-2 was also supported by *prt*2 results (Figure 4.1).

AR510 from FaTG-3, which is a hybrid between *E. typhina* and *E. baconii*, contains a hybridising fragment identical in size to the E8 fragment. This suggests three possibilities: either one of the *E. typhina* or *E. baconii* hybridising fragments has been lost, or both fragments are the same size. Tf 2, an interspecific hybrid between *N. uncinatum*, *E. baconii* and *E. festucae*, contains hybridising fragments similar in size to those seen in AR510 and Lp19. As AR510 and Tf2 both share *E. baconii* ancestry, this suggests the *E. baconii* and *E. typhina* bands could be the same size. Both Tf2 and Lp19 have *E. festucae* ancestry, which could account for their similar-sized hybridising bands. The *N. uncinatum* hybridising band may have been lost or altered in Tf2. Finally Tf13, an interspecific hybrid between *E. baconii* and *E. festucae*, contains a band equivalent in size to the Lp19 band, along with two smaller bands. Tf13 and Lp19 also share *E. festucae* ancestry, which again explains the presence of a shared hybridising fragment. The smaller bands seen in Tf13 are probably the result of a mutation introducing a *Hin*dIII restriction site to the *E. baconii* fragment.

5.2 SUBTILISIN-LIKE PROTEASES IN MULTI-GENE FAMILIES

Multiple copies of subtilisin-like protease genes are found in many filamentous fungal species. In saprophytic species, these proteases would be used to break down proteins in the environment. A varying range of protease activities could expand the number of environmental niches available to a saprophyte. Although multiple copies of subtilisin-like proteases have no apparent benefit, minor differences in protein composition could lead to differences in protease activity (Segers et al., 1999).

Southern blots using a fragment of the Prl gene from *Metarhizium anisopliae* found that *Verticillium chlamydosporium* and *V. lecanii* contained multiple different copies of subtilisin-like proteases (Segers et al., 1999). St Leger and co-workers (1992) also found multiple copies of subtilisin-like genes in *V. lecanii*. *V. chlamydosporium* strains showed wide variation in the amount of subtilisin-like proteases produced, although Southern blotting showed that all strains had at least one copy of the gene. Strain 8 produced at least

two different subtilisin-like proteins, while strain 10 produced four different subtilisin-like proteins. The proteins differed in both isoelectric point and molecular weight.

Tritirachium album, a saprophytic fungus, also produces multiple subtilisin-like proteases. As for *V. chlamydosporium*, there is some variation between strains. Some strains of *T. album* produces proteinase K and proteinase T (Gunkel and Gassen, 1989; Samal et al., 1989). Proteinase R, which is closely related to proteinase K, is produced by other strains of *T. album* (Samal et al., 1990). *Metarhizium anisopliae*, an entomopathogenic fungus, is currently thought to contain at least three subtilisin-like protease genes (Joshi et al., 1997; St Leger et al., 1994). *Beauveria bassiana* (Joshi et al., 1995; Segers et al., 1994) also appears to contain multiple subtilisin-like protease genes. *Magnaporthe poae,* a fungal pathogen of *Poaceae* grass species, also contains multiple copies of sequences homologous to subtilisin-like proteases (Sreedhar et al., 1999). Southern blot analysis showed there were at least two copies of sequences homologous to the gene. Searches of databases also revealed that *Acremonium chrysogenum* contained two sequences related to subtilisin-like proteases. One of these sequences, despite its high level of identity to subtilisin-like proteases, is thought to act as an unspecific esterase in cephalosporin C biosynthesis (GenBank accession CAB87194).

Most work on fungal subtilisin-like proteases has been done on the basis of subtilisin-like activity, rather than expression of a particular subtilisin-like protease gene. Many of the fungal species currently thought to contain only one subtilisin-like protease gene may contain others, which have not been isolated at the present time. A lack of information on the relationships between ascomycete fungi and their deuteromycete relatives also complicates the analysis of multi-gene families of subtilisin-like proteases.

The presence of at least two subtilisin-like protease sequences in all of the endophyte taxonomic groups suggests that the *prt*1 and *prt*2 genes from Lp19 have not arisen by gene duplication in *N. lolii* endophytes. Unlike most of the asexual endophyte taxonomic groups, *N. lolii* is haploid and has evolved directly from the asexual stage of *E. festucae*. This suggests that interspecific hybridisation is not responsible for the two subtilisin-like protease genes in Lp19. Duplications of subtilisin-like protease genes early in fungal evolution may have given rise to the multi-gene families found in many ascomycetes.

5.3 FEATURES OF THE *PRT1* GENE

5.3.1 Coding region of the *prt1* gene

The *prt1* gene appears to contain only one putative intron, which is yet to be confirmed by sequencing cDNA (Section 3.3.5). Most other fungal subtilisin-like protease genes contain two or more introns. The *Aspergillus oryzae* subtilisin-like protease gene contains two introns (Cheevadhanarak et al., 1991), while other sequences such as Atl (*E. typhina*; (Reddy et al., 1996)) and Mpl (*Magnaporthe poae*; (Sreedhar et al., 1999)) contain three introns. The position of the single intron in *prt1* is highly conserved with related fungal genes.

The *prt1* gene encodes a putative peptide of 434 amino acid residues, with significant homology to subtilisin-like proteases from *Aspergillus* species, *Magnaporthe poae*, and other ascomycete fungi. Prt1 contains all of the conserved active site residues required for subtilisin-like protease activity (Section 3.3.5). Based on homology with other subtilisin-like proteases, Prt1 contains a 24 amino acid residue signal peptide and a 101 amino acid residue propeptide, which are cleaved to give a mature protein of 309 amino acid residues (Section 3.3.8). Current evidence suggests that the signal peptide of proteinase K family subtilisin-like proteases is removed by a signal peptidase, while the propeptide is cleaved from the mature protease by autolysis (Abraham and Breuil, 1995; Terada et al., 1990).

There is a pronounced bias in the codon frequencies of the *prt1* gene (Section 3.3.5). Generally, there is a preference for cytosine at the third position of each codon, followed by guanine when a purine base is present. Similar codon biases are also seen in highly expressed genes in *Aspergillus nidulans* and *Neurospora crassa* (Gurr et al., 1987). The Lp19 hmg CoA reductase gene also demonstrates a similar codon bias (Dobson, 1997). Hmg CoA reductase is a key enzyme in cell metabolism, and is expressed constitutively. Based on this evidence, *prt1* may also be a highly expressed gene.

5.3.2 Regulation of *prt1*

Analysis of the promoter region revealed potential binding sites for global nutrient regulators CreA and AreA (Section 3.3.3). CreA represses gene expression in the presence of easily used carbon sources such as glucose (Felenbok and Kelly, 1996). AreA is a positively acting

factor controlling response to nitrogen limitation: when levels of nitrogen are low, AreA binds to promoters, enhancing gene expression. Most potential AreA binding sites in the *prt*1 promoter occurs in close pairs; for instance between –550 and–530 and between –380 and –350 (Figure 3.9).

The promoter region also contained a binding site for PacC, a global regulator of gene expression in response to pH in filamentous fungi (Tilburn et al., 1995). This pH regulator is thought to activate expression of subtilisin like proteases under alkaline conditions. The PacC binding site is very close to two GATT AreA binding sites. In *Neurospora crassa*, GATT sites are only 50% as efficient at binding Nit2 (an AreA homologue) as GATA sites (Chiang et al., 1994). As the PacC binding site is closer to its consensus sequence than the AreA homologue is to GATT, it is possible that PacC is more likely than AreA to bind in this region. Further work needs to be done on the effect of pH on *prt*1 expression.

Expression studies using different carbon and nitrogen sources revealed that expression of *prt*1 was responsive to nutrient limitation. Sucrose and mannitol were selected as carbon sources, while nitrate and glutamate were selected as nitrogen sources. Growth studies *using Neotyphodium coenophialum* showed that the fungus could utilise sucrose and mannitol as carbon sources (Kulkarni and Nielsen 1986). *N. coenophialum* could also use glutamate as a nitrogen source, but could not utilise nitrate.

Both mannitol and sucrose are non-reducible sugars that are transported in the phloem of the grass host. Endophytes tend to grow within regions of the plant that are net importers of photosynthates, so it is probable that most of the photosynthates supplied by the host for fungal growth are transported in the phloem. Sucrose and mannitol processed by endophytes in different ways. Experimental evidence suggests that endophytes have a cell wall invertase, which allows the uptake of sucrose as glucose and fructose by separate transporters. Mechanisms for mannitol uptake have not been elucidated. Other filamentous fungi have a pentose phosphate system that allows them to utilise mannitol as a carbon source (Hult et al., 1980).

In the presence of sucrose as the sole carbon source or in unsupplemented cultures, the *prt*1 gene was expressed. If sucrose and nitrate were present, *prt*1 expression was particularly strong. When sucrose and glutamate were present, *prt*1 expression was repressed

(Section 3.6). These results suggested that an easily utilised nitrogen source was capable of repressing *prt*1 expression.

However, when endophyte was grown in mannitol and glutamate, expression of *prt*1 appeared similar to that in unsupplemented cultures. All cultures containing mannitol grew very quickly, so it is possible that this was due to the rapid growth leading to nutrient depletion. Mannitol is known to be an excellent carbon source for supporting endophyte growth (Kulkarni and Nielsen 1986). When cultures were only supplemented with glutamate as the sole carbon and nitrogen source *prt*1 was expressed.

The nutrient limitation studies carried out in this study supported the role of carbon and nitrogen availability in regulating *prt*1 expression. These preliminary experiments do not give any clear indication of how this regulation is achieved. It is probable that more informative results might be obtained by making some alterations to the experimental design. For the experiments described endophytes were first grown in PD broth for 9 days at 22°C before subculturing to CD. Expression of *prt*1 should have been examined at the end of the initial growth period to compare with *prt*1 expression in the supplemented CD media. It is also possible that two days is an excessively long time period for inducing subtilisin-like protease expression. *Metarhizium anisopliae*, a fungal entomopathogen, produces a subtilisin-like polypeptide within 1 hour of nutrient deprivation (St Leger et al., 1991).

Transforming an expression vector containing the *prt*1 promoter into *Penicillium paxilli* provided further evidence that *prt*1 expression is controlled by global regulatory mechanisms that are common to filamentous fungal species, or that *prt*1 is constitutively expression on PD agar. Although transformation frequencies were low, at least half of the transformants were expressing GUS (Section 3.5.2). As *Penicillium paxilli* is not closely related to *N. lolii*, this suggests that factors common to both fungal species regulate *prt*1 expression.

5.3.3 The 3'UTR and repetitive DNA downstream of *prt*1

The polyadenylation site for the *prt*1 gene was confirmed by 3' RACE. This site was similar to polyadenylation sites for yeast and mammalian genes. In yeast genes, the polyadenylation site is $Py(A)_n$ while in mammalian genes it is usually between a pyrimidine and an adenine base (Zhao et al., 1999). The polyadenylation recognition sequence, AAUAAA in higher eukaryotes and related sequences in yeast, are necessary for both cleavage and polyadenylation. The closest sequences to this for the *prt*1 gene were AATAAT (1453-1458,

58 bp upstream of the polyadenylation site) and AATATA (1500-1505, 11 bp upstream). In most eukaryotes, the distance between the AAUAAA sequence and the polyadenylation site is between 11 and 30 base pairs. The AT-rich sequence in the complex microsatellite found in the 3' UTR could also contribute to the efficient polyadenylation of *prt*1 mRNA.

Microsatellite and minisatellite DNA is found in most eukaryotic genomes. Kokoska and colleagues (1999) defined microsatellites as repeat units from 1-13 base pairs in arrays of 10-60 base pairs, and minisatellites as repeat units of less than 15 base pairs, often in arrays of several hundred base pairs. Based on this definition, a complex microsatellite was found in the 3' untranslated region of the *prt*1 gene just downstream of the polyadenylation recognition site. The sequence of the microsatellite is very AT rich, which suggests RNA transcribed in this region would anneal very weakly to the DNA template. The second part of the complex microsatellite, which largely consists of TA repeats, could also form a hairpin secondary structure that may be important in the termination of transcription.

The minisatellite downstream of the *prt*1 gene is thought to be the first minisatellite discovered in endophyte DNA. This minisatellite array consists of 40 repeats of a 9 base pair unit. In the centre of the minisatellite, the core sequence is highly conserved, with 17 identical tandem repeats (Figure 3.11). Towards the outside of the array, there is more variation in the repeats. Haber and Louis (1998) suggested that short direct repeats of 5 to 10 base pairs that flank yeast and human minisatellites may be involved in the origin of minisatellite sequences. They hypothesised that replication slippage or unequal crossing over could lead to the expansion of minisatellite repeats. For this minisatellite, direct repeats of AAAATC were found 42 bases upstream of the minisatellite and flanking the 3' side of the minisatellite. This suggests this minisatellite may have arisen by a similar mechanism.

5.4 FEATURES OF THE *PRT*2 GENE

Southern blotting showed that the *prt*2 gene was located on 8.6 kb *Eco*RI and 9.2 kb *Ssi*I fragments respectively. After library screening, the pGH3 insert used to screen the Southern blot hybridised to a 2.0 kb *Bam*HI fragment, which was subcloned for sequencing. The nucleotide sequence of the *prt*2 gene has proved difficult to obtain due to difficulties designing primers for sequencing. The 5' region of the *prt*2 gene appears to contain sequences that are mildly repetitive. Oligonucleotide primers often anneal to different regions of the gene other than that to which they were designed.

The subcloned *Bam*HI fragment does not contain the entire *prt*2 gene, which has complicated sequencing at the 3' region of the gene. This region may need to be isolated by completing another round of library screening, which should provide λ clones that overlap with λMM3. The initial library screen only resulted in one unique λ clone. This is probably due to a lack of positive clones during the first round of library screening, and poor sampling of clones during the second round of screening.

The sequence currently available shows that the *prt*2 coding region encodes a gene similar to but not identical to *prt*1. Unlike the *prt*1 gene, which contained only one intron, the *prt*2 gene appears to contain at least three introns. The potential intron splice sites and the lariat sequence are good matches for fungal consensus sequences (Gurr et al., 1987). Difficulties in designing primers have meant that it has been difficult to check for the presence of these introns by RT-PCR.

The partial protein sequence of Prt2 shows that all of the conserved active site residues are present. Prt2 also shows high levels of similarity to other fungal subtilisin-like proteases (Figures 4.7 and 4.8). At present, the Prt2 peptide sequence includes only mature regions of the protease. Further nucleotide sequencing needs to be done to elucidate the signal peptide and propeptide sequences.

Difficulties in primer design have created difficulties in sequencing the *prt*2 gene. Some primers have not annealed to where they were designed. Initially, I thought this was due to repetitive sequences in the DNA, but there appears to be no obvious repetition in the coding region of the *prt*2 gene. Problems with primers could be due to the formation of primer secondary structure, or the formation of primer dimers.

5.5 FUTURE WORK

Many further experiments are needed to finish characterising the *prt*1 and *prt*2 genes. Further sequencing is needed to confirm the transcription start sites and the sequences and positions of introns in both genes. The *prt*2 gene sequence is incomplete, and must be finished to determine a putative amino acid sequence for the *prt*2 subtilisin-like protease. The polyadenylation site for the *prt*2 gene also needs to be confirmed by 3' RACE.

Ideally, further sequence will need to be collected for the promoter regions of the *prt*1 and *prt*2 gene. To date, there are only approximately 500 base pairs of *prt*1 promoter sequence. Although 500 base pairs of the *prt*1 promoter sequence was sufficient for expression in *P. paxilli* (Sections 3.4-3.5), it is probable that there are more upstream regulatory elements that control *prt*1 expression. Complete promoter sequence is critical to getting meaningful results using expression vectors, and to checking if the *prt*1 and *prt*2 promoters contain similar regulatory elements or consensus sequences.

At present, the *prt*1-GUS translational fusion vector pMM9 is the only expression vector constructed. Similar constructs containing the *prt*1 or *prt*2 promoters in a GFP expression vector could be made, along with a GUS expression vector containing the *prt*2 promoter. These expression constructs could be transformed into *N. lolii*, and stable transformants used to artificially infect perennial ryegrass. GUS expression vectors are ideal for checking the level of expression within infected plant tissues (Saunders, 1997). GFP expression vectors, with the use of confocal microscopy, could be used to examine the localisation of *prt*1 and *prt*2 expression within endophyte-infected plants.

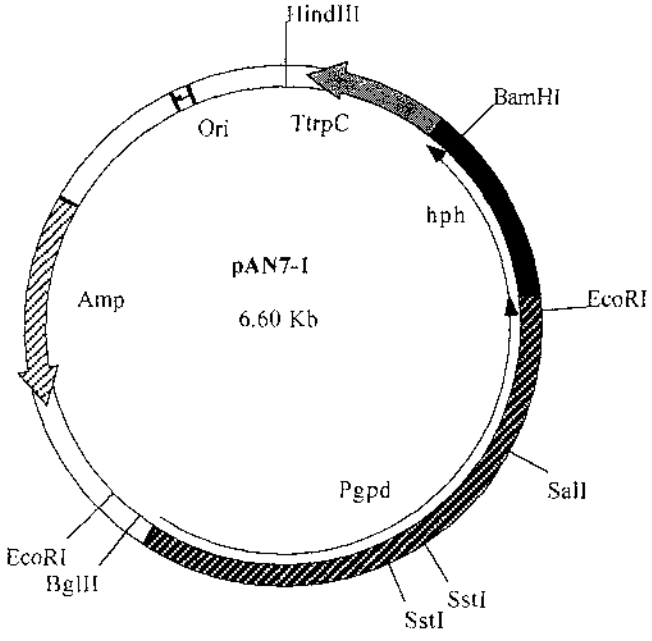
Expression of *prt*2 within cultures also needs to be completed. The same RNA samples used in Section 3.6 could be examined for *prt*2 expression. RT-PCR using RNA samples from grass-endophyte interactions should be completed to confirm whether *prt*1 and *prt*2 are expressed by the endophyte within infected plants.

APPENDIX A1

PLASMID MAPS

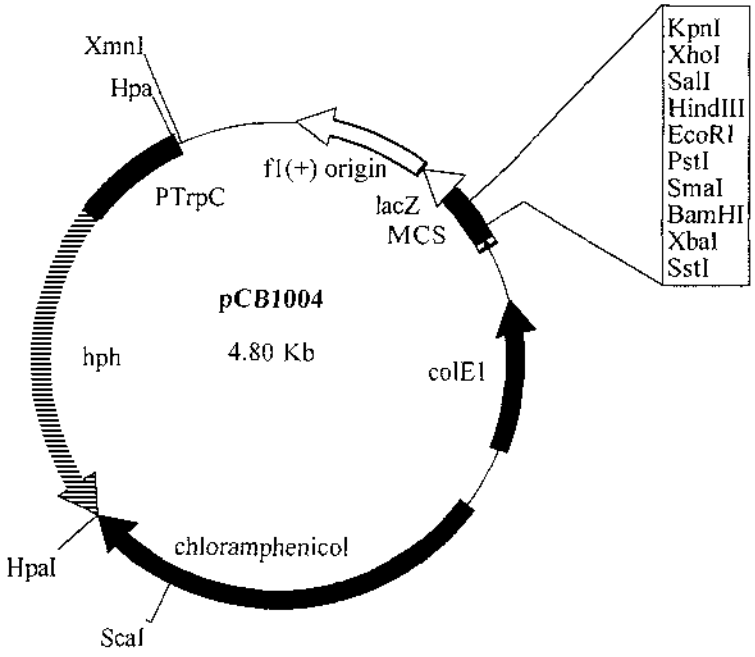
Appendix A1.1 pAN7-1

Restriction map of pAN7-1.



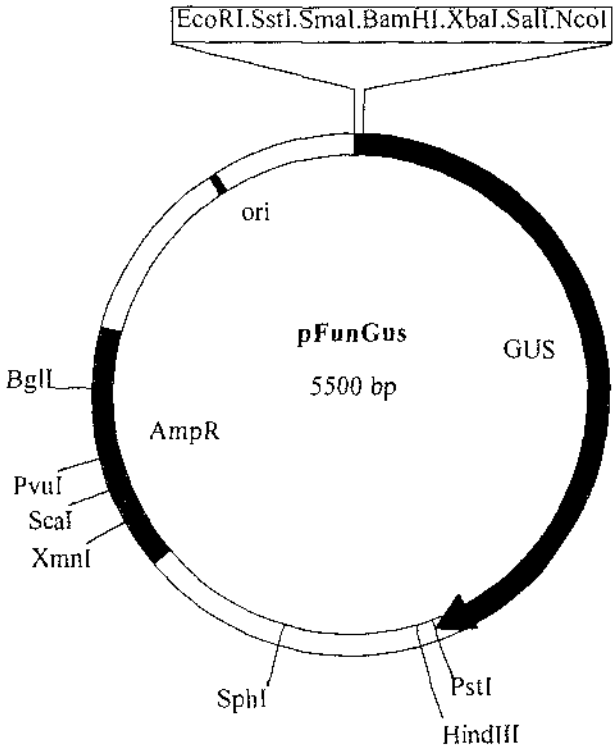
Appendix A1.2 pCB1004

Restriction map of pCB1004.



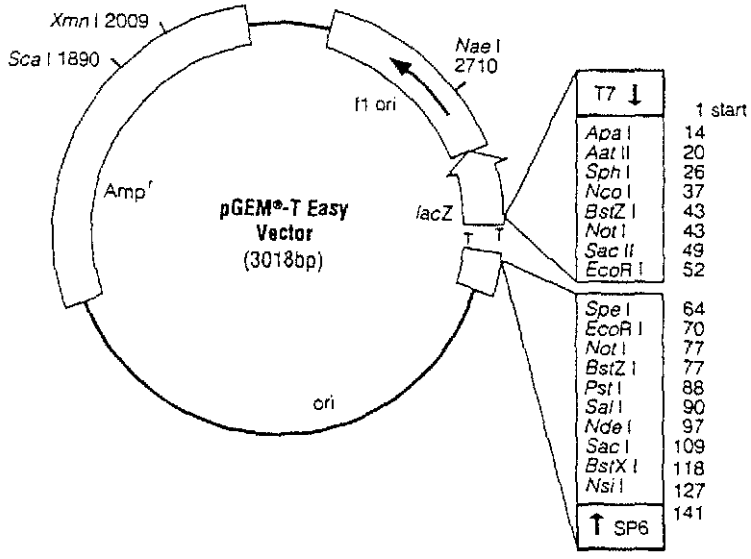
Appendix A1.3 pFunGus

Restriction map of pFunGus.



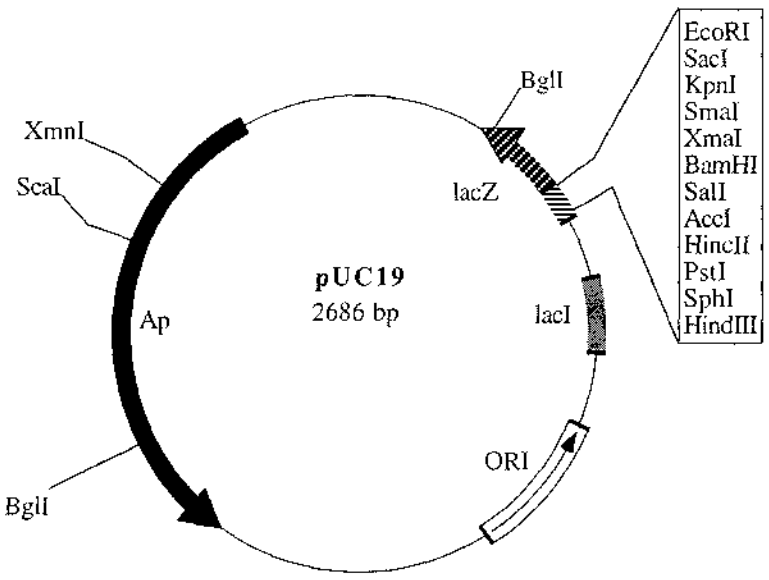
Appendix A1.4 pGEM-T Easy

Restriction map of pGEM-T Easy.

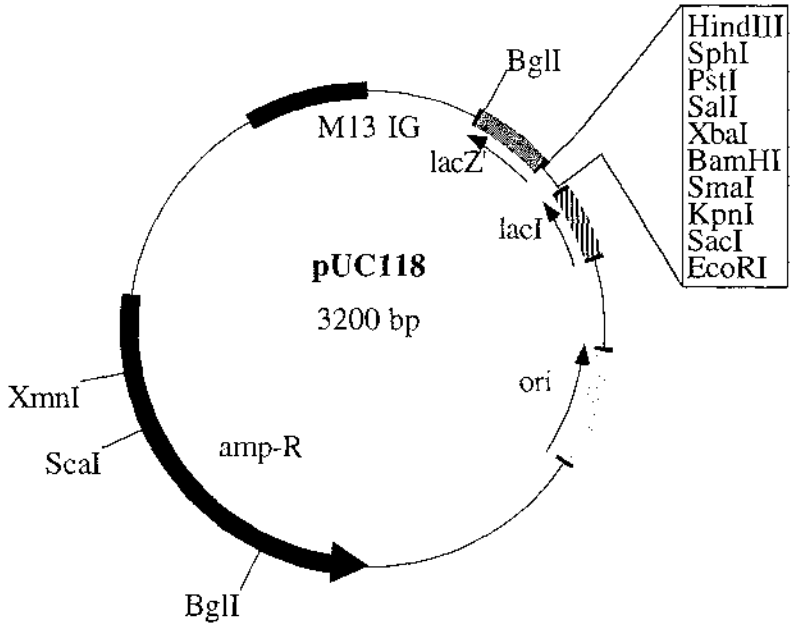


Appendix A1.5 pUC vectors

A. Restriction map of pUC19.

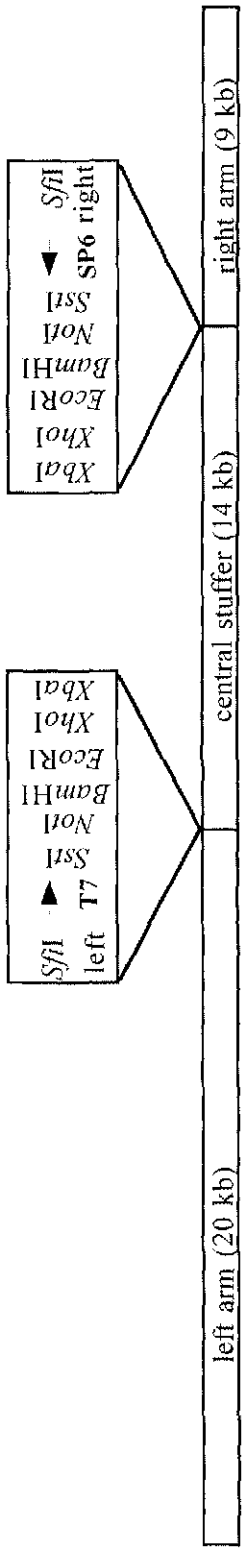


B. Restriction map of pUC118.



Appendix A1.6 λGEM®-12 cloning vector

Restriction map of the λGEM®-12 vector



APPENDIX A2

PROBE INFORMATION

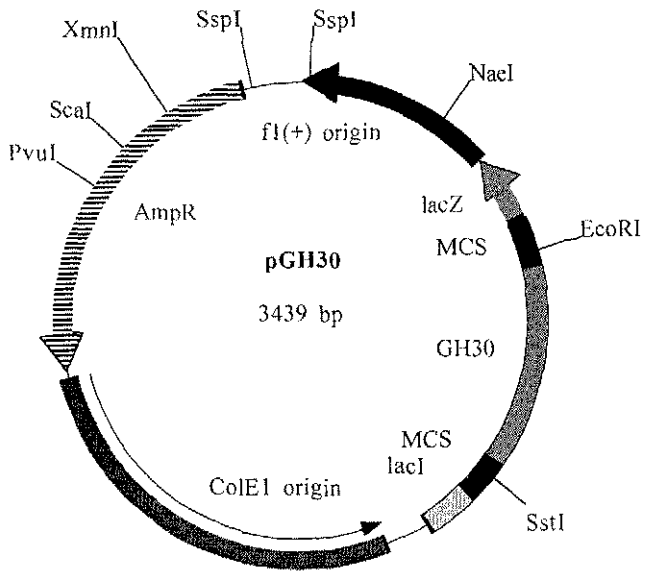
Appendix A2.1 Amplification of protease genes using degenerate primers

Partial nucleotide sequence of the At1 protease is shown below. Introns are shown in lower case text, and the corresponding amino acid sequence is shown in grey uppercase text. Annealing sites of degenerate primers are indicated in coloured text: ACREL-R1 (ACT CAC GTC GCC GGT ACI AT) is indicated in blue, ACREL-R2 (CGC ACC TAC GGT GTI GCI AA) in green, and ACREL-L1 (TGI GGA GTA GCC ATG GAI GT) in red. The direction of extension from the primer is shown by an arrow in the same colour as the text. All of these primers are designed to highly conserved regions of the coding sequence (Reddy et al., 1996). Both the GH30 and GH3 PCR products were amplified using the ACREL-R1 and ACREL-L1 primers.

```
GCCAGAGTCGTGCAGAAGAATGCTCCATGGGGCCTAGCCCGTATATCCCACCGA   453
 A  R  V  V  Q  K  N  A  P  W  G  L  A  R  I  S  H  R    124
CGACGCGGCTCGAACGAGTACGTCTACGACAATAGTGGCGGCAAAGGTGCTTGC   507
 R  R  G  S  N  E  Y  V  Y  D  N  S  G  G  K  G  A  C    142
GTCTATGTCATCGACACGGGCGTAGATGATCGCCACCCGgtgagaaacaccctt   561
 V  Y  V  I  D  T  G  V  D  D  R  H  P                   155
cttgtccctttttttccacaactcactcggcccggttcacccgagcgcggaacta   615

acagcatccagGAGTTCGAAGGCCGGGCGCACCAGATCCAGTCCTACGTCGCCG   669
           E  F  E  G  R  A  H  Q  I  Q  S  Y  V  A  G      170
GATCCAACGTCGATGACAACGGCCACGGCACCCACGTCGCCGGCACAATCGGCA   723
  S  N  V  D  D  N  G  H  G  T  H  V  A  G  T  I  G  S    188
GCCGCACCTACGGCGTAGCGAAGCGGGTGACCATCTTCGGCGTCAAGGTCCTCC   777
  R  T  Y  G  V  A  K  R  V  T  I  F  G  V  K  V  L  P    206
CTGCCCGCGGCACGAGCCCCAATTCCGTCATCATCAAGGGCATGGATTTCGTGC   831
  A  R  G  T  S  P  N  S  V  I  I  K  G  M  D  F  V  H    224
ACGCGATGCCCAGCGGCGTAAATGCCCCCACGGACGTCGTGGTCAACATGTCCC   885
  A  M  P  S  G  V  N  A  P  T  D  V  V  V  N  M  S  L    242
TCGGCGGAGGCTACTCCAAGGCCACAAACCAAGCCGCCGCCCGCCTCGAGCCAG   939
  G  G  G  Y  S  K  A  T  N  Q  A  A  A  R  L  V  R  A    260
GTACTTCGTCGCCGTGGCCTCGGGCAACAATAACAGAGACGCCCTCAGGGAACT   993
  K  Y  F  V  A  V  A  S  G  N  N  N  R  D  A  R  N  Y    278
ACTCACCCGCCTCGGAACCATCCGTCTGCACTGTCGGCGGCACGGACAAGTTCG  1047
  S  P  A  S  E  P  S  V  C  T  V  G  G  T  D  K  F  D    296
ACAGCGTATACATGTCGAACTGGGGGCCTGCCGTCGACATCAACGGTCCCGGCG  1101
  S  V  Y  M  S  N  W  G  P  A  V  D  I  N  G  P  G  V    314
TCGATGTCCTGTCCACTCTCCCCAACCGCCGGACTgtatgtttttttttctta  1155
  D  V  L  S  T  L  P  N  R  R  T                        325
taaaaatccccgcttggcgagcagaggaactgacatgcatgatgcagGGCCGCT  1209
                                                 G  R  L    328
TGACGGGAACGTCCATGGCTACCCCGCACATTGCGGGACTGGGCGCGTACCTCG  1263
  T  G  T  S  M  A  T  P  H  I  A  G  L  G  A  Y  L  A    346
CTGCTAAAAACGGCCGGCGCGCTGGTCCCGGGTTGTGCCGGACGATAAAGGACA  1317
  A  K  N  G  R  R  A  G  P  G  L  C  R  T  I  K  D  K    364
TGGCCACTAAAAATGTCATCACGAACCAGGTGGCTGGCACGGTCAATCTGCTGG  1371
  A  T  K  N  V  I  T  N  Q  V  A  G  T  V  N  L  L  A    382
CATTCAACGGCGAGAAGTAG                                    1391
  F  N  G  E  K  *                                        387
```

Appendix 2.2 pGH30 plasmid containing the GH30 PCR product

Part A: Restriction map of pGH30



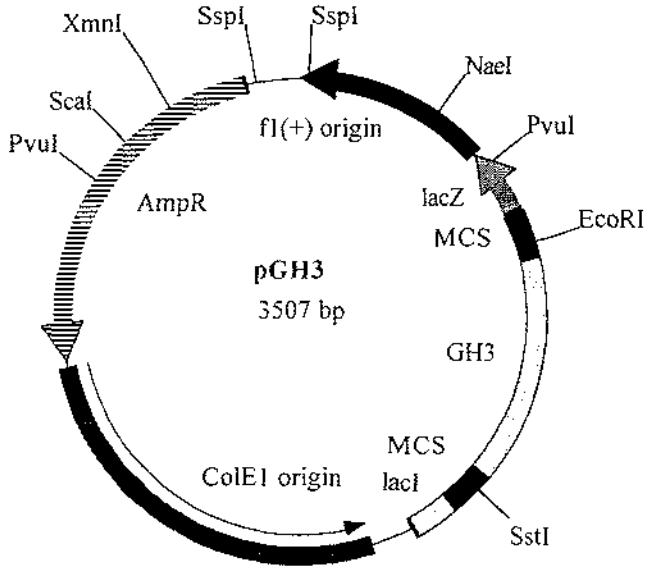
Appendix 2.2 pGH30 plasmid containing the GH30 PCR product

Part B: Nucleotide sequence of GH30 PCR product

```
CTCACGTCGCCGGTACGATCGGCGGCAAGACGTACGGAGTGGCCAAGAAC   50

GCCTTGCTGCTCGCCGTCAAGATCTTCAACAGCAGGTCGTCCAGCACGTC  100

CGTCATCCTGGCCGGATACAACTGGGCGGTCAACGACATCGTCCGAAAGG  150

                                                 SalI
                                                  |
GCCGCACCAAGAGGGCCGCCATCAACATGTCCCTCGGCGGCCCCAAGTCG  200

ACCGCCTTCAACACGGCCGTCGAGAGGGCCTCGGCCTCGGGCGTCTTGTC  250

CATCATCGCCGCCGGCAACGAGGCCCAGGATGCCTCCAACGTGTCTCCCG  300

CGTCGGCCCCGAGCGCCATCACCGTCGCCGCCATCAATCGCGACTGGACC  350

CTCGCCTCGTACAGCAACTTTGGCTCCGTCGTGGACATTTGCGCCCCTGG  400

ATCGAACATCACCTCTGCCTGGAACACGGGAGACTCGTCCGAGAAGACCA  450

TCTCGGGCACCTCCATGGCTACTCCCCA                         478
```

Appendix 2.3 pGH3 plasmid containing the GH3 PCR product

Part A: Restriction map of pGH3



Appendix A2.3 pGH3 plasmid containing the GH3 PCR product

Part B: Nucleotide sequence of the GH3 PCR product used as a probe.

Intron is depicted by lower case text.

```
CTCACGTCGCCGGTACGATCGGTAGCCGTAGCTACGGCATCGCCAAGAAG   50

ACTCAGCTGCTTGACATCAAGGTACTCAGCGACCAGGGATCTGGAAACAA  100

TTCCGCCATTATCGCGGGCATGGACTTTGCCGTCCAGGACGCCAGGCAAC  150

GAAGCTGCGCCAAGGGTGTTCTCGCCAACATGAGTCTCGGCGGCAGATAC  200

TCGCAGTCGCTGAACGATGCGGCCGCTCAGATGATTCAGTCTGGCGTCTT  250

CCTCGCCGTCGCCGCTGGAAACAATCGCCAGGATGCCTCTGGCTACTCGC  300

CTGCCTCTGAGCCGAGTGTTTGCACCGTCGGATCAACGGATAGCTCCGAC  350

AGTCTCTCTTCATTCTCCAACTATGGAAGCGTCGTCGATATCCTGGCCCC  400
                                                 HindIII
                                                    |
CGGCTCCGACATTCTTTCCACCTGGCCCGGTGGCAGCATCgtaagttgaa  450

gcttcgtccttgccgaccaccgattcaccatgttccatgccttgacactg  500
                                KpnI
                                 |
cctgctctcagAAAATCCTTTCGGGTACCTCCATGGCTACTCCCCA      546
```

Appendix A2.4 Comparison of GH30 with *prt*1 and GH3 with *prt*2

A. Comparison of GH30 with *prt*1 sequence

The probe sequence is shown in black, with the corresponding *prt*1 sequence in green text. Identical bases are linked by a line.

```
CTCACGTCGCCGGTACGATCGGCGGCAAGACGTACGGAGTGGCCAAGAAC  50
|| |||||||||||  || |||||||||||||||||||||||||||||||||
CCCACGTCGCCGGCACCATCGGCGGCAAGACGTACGGAGTGGCCAAGAAC

GCCTTGCTGCTCGCCGTCAAGATCTTCAACAGCAGGTCGTCCAGCACGTC 100
|||  |||||||||||||||||||||||||||||||||||||||||||||||
GCCAAGCTGCTCGCCGTCAAGATCTTCAACAGCAGGTCGTCCAGCACGTC

CGTCATCCTGGCCGGATACAACTGGGCGGTCAACGACATCGTCCGAAAGG 150
|||||||||||||||||||||||||||||||||||||||||||||||||
CGTCATCCTGGCCGGATACAACTGGGCGGTCAACGACATCGTCCGAAAGG

GCCGCACCAAGAGGGCCGCCATCAACATGTCCCTCGGCGGCCCCAAGTCG 200
|||||||||||||||||||||||||||||||||||||||||||||||||
GCCGCACCAAGAGGGCCGCCATCAACATGTCCCTCGGCGGCCCCAAGTCG

ACCGCCTTCAACACGGCCGTCGAGAGGGCCTCGGCCTCGGGCGTCTTGTC 250
|||||||||||||||||||||||||||||||||||||||||||||||||
ACCGCCTTCAACACGGCCGTCGAGAGGGCCTCGGCCTCGGGCGTCTTGTC

CATCATCGCCGCCGGCAACGAGGCCCAGGATGCCTCCAACGTGTCTCCCG 300
|||||||||||||||||||||||||||||||||||||||||||||||||
CATCATCGCCGCCGGCAACGAGGCCCAGGATGCCTCCAACGTGTCTCCCG

CGTCGGCCCCGAGCGCCATCACCGTCGCCGCCATCAATCGCGACTGGACC 350
|||||||||||||||||||||||||||||||||||||||||||||||||
CGTCGGCCCCGAGCGCCATCACCGTCGCCGCCATCAATCGCGACTGGACC

CTCGCCTCGTACAGCAACTTTGGCTCCGTCGTGGACATTTGCGCCCCTGG 400
|||||||||||||||||||||||||||||||||||||||||||||||||
CTCGCCTCGTACAGCAACTTTGGCTCCGTCGTGGACATTTGCGCCCCTGG

ATCGAACATCACCTCTGCCTGGAACACGGGAGACTCGTCCGAGAAGACCA 450
|||||||||||||||||||||||||||||||||||||||||||||||||
ATCGAACATCACCTCTGCCTGGAACACGGGAGACTCGTCCGAGAAGACCA

TCTCGGGCACCTCCATGGCTACTCCCCA                        478
|||||||||||||||||| |||||  ||
TCTCGGGCACCTCCATGGCGACTCCTCA
```

Appendix A2.4 Comparison of GH30 with *prt*1 and GH3 with *prt*2

B. Comparison of GH3 with *prt*2 sequence

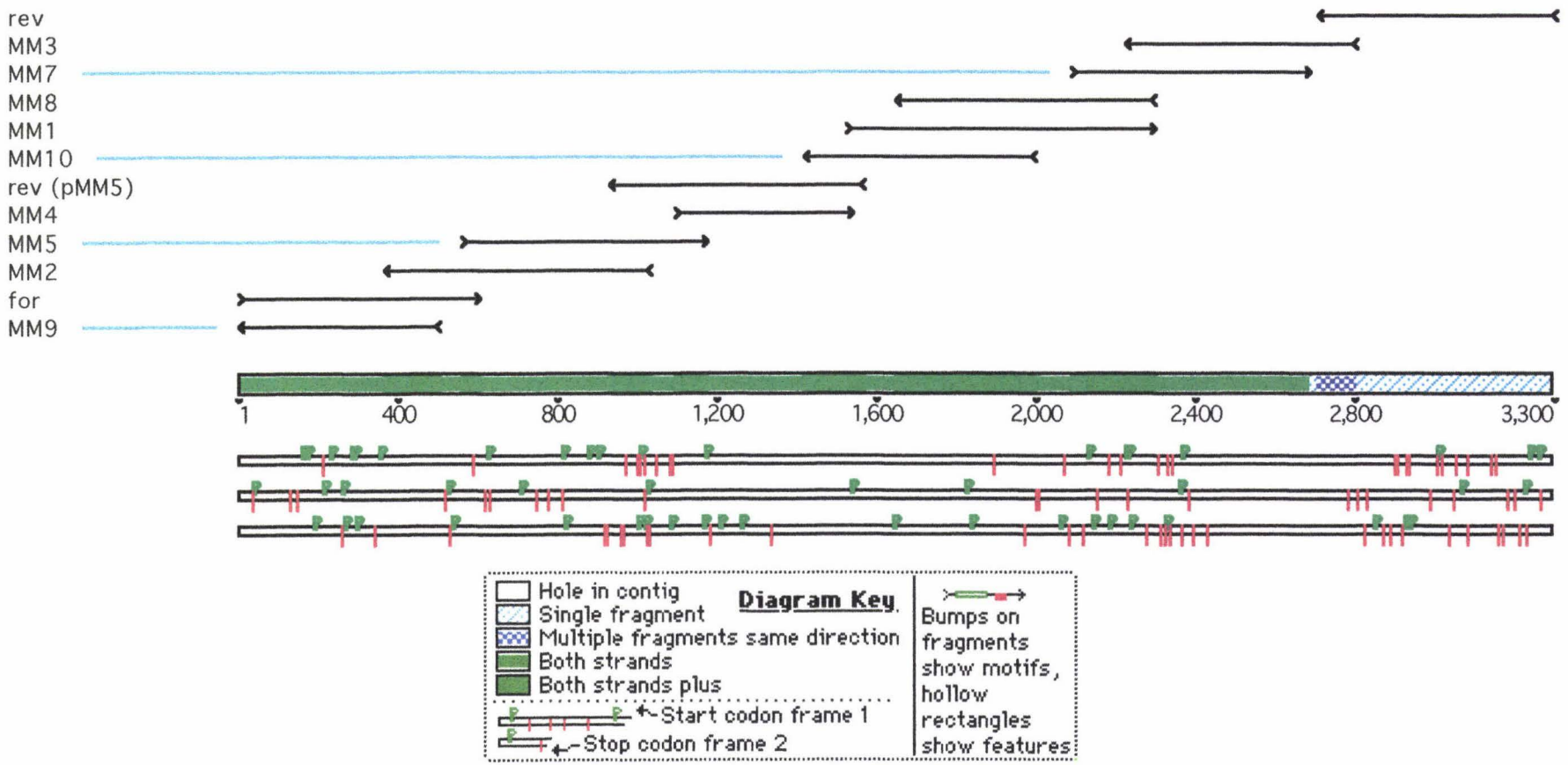
The probe sequence is shown in black, with the corresponding *prt*2 sequence in blue text. Lower case text indicates a potential intron. Identical bases are linked by a line.

```
CTCACGTCGCCGGTACGATCGGTAGCCGTAGCTACGGCATCGCCAAGAAG  50
CTCACGTCGCTGGCACCATCGGTAGCCGTAGCTACGGCATCGCCAAGAAG

ACTCAGCTGCTTGACATCAAGGTACTCAGCGACCAGGGATCTGGAAACAA 100
ACTCAGCTGCTTGGCATCAAGGTACTCAGCGACCAGGGATCTGGAAACAA

TTCCGCCATTATCGCGGGCATGGACTTTGCCGTCCAGGACGCCAGGCAAC 150
TTCCGCCATTATCGCGGGCATGGACTTTGCCGTCCAGGACGCCAGGCAAC

GAAGCTGCGCCAAGGGTGTTCTCGCCAACATGAGTCTCGGCGGCAGATAC 200
GAAGCTGCGCCAAGGGTGTTCTCGCCAACATGAGTCTCGGTGGCAGATAC

TCGCAGTCGCTGAACGATGCGGCCGCTCAGATGATTCAGTCTGGCGTCTT 250
TCGCAGTCGCTGAACGATGCGGCCGCTCAGATGATTCAGTCTGGCGTCTT

CCTCGCCGTCGCCGCTGGAAACAATCGCCAGGATGCCTCTGGCTACTCGC 300
CCTCGCCGTCGCCGCTGGAAACAATCGCCAGGATGCCTCTGGCTACTCGC

CTGCCTCTGAGCCGAGTGTTTGCACCGTCGGATCAACGGATAGCTCCGAC 350
CTGCCTCTGAGCCGAGTGTTTGCACCGTCGGATCAACGGATAGCTCCGAC

AGTCTCTCTTCATTCTCCAACTATGGAAGCGTCGTCGATATCCTGGCCCC 400
AGTCTCTCTTCATTCTCCAACTATGGAAGCGTCGTCGATATCCTGGCCCC

CGGCTCCGACATTCTTTCCACCTGGCCCGGTGGCAGCATCgtaagttgaa 450
CGGCTCCGACATTCTTTCCACCTGGCCCGGTGGCAGCATCgtaagttgaa

gcttcgtccttgccgaccaccgattcaccatgttccatgccttgacactg 500
gcttcgtccttgccgaccaccgattcaacatgttccatgccttgacactg

cctgctctcagAAAATCCTTTCGGGTACCTCCATGGCTACTCCCCA      546
cctgctctcagAAAATCCTTTCGGGTACCTCGATGGCTACTCCCCA
```

APPENDIX A3

SEQUENCE DATA

Contig[0008]
Appendix A3.1 prt1 contig



Contig[0008]
prt1 contig

```
for     GAATTCGCGGAAAGGGTTCTGCAACGGACGCTAGATGTGGAGCCAGCTTGGTGGCTTT
MM9     GAATTCGCGGAAAGGGTTCTGCAACGGACGCTAGATGTGGAGCCAGCTTGGTGGCTTT
  #1    ------------------------------------------------------------

        GAATTCGCGGAAAGGGTTCTGCAACGGACGCTAGATGTGGAGCCAGCTTGGTGGCTTT

for     GTCTCTTTCCTTTCGTTGGATTCCTTTCAATCCTCGCCAAGAGCTCGGGCTCAAAGAA
MM9     GTCTCTTTCCTTTCGTTGGATTCCTTTCAATCCTCGCCAAGAGCTCGGGCTCAAAGAA
  #59    ------------------------------------------------------------

        GTCTCTTTCCTTTCGTTGGATTCCTTTCAATCCTCGCCAAGAGCTCGGGCTCAAAGAA

for     GTCAGGACTGAGCGAGCAACCGTGGTGGTTGAGTTTCGCGACGATGCACATCATGGCA
MM9     GTCAGGACTGAGCGAGCAACCGTGGTGGTTGAGTTTCGCGACGATGCACATCATGGCA
  #117   ------------------------------------------------------------

        GTCAGGACTGAGCGAGCAACCGTGGTGGTTGAGTTTCGCGACGATGCACATCATGGCA

for     GCCTCTTGGTGGCCATGTTGTCCGGCAGCGCTTTCCTAATATGGTCGCGAGGAGGGTA
MM9     GCCTCTTGGTGGCCATGTTGTCCGGCAGCGCTTTCCTAATATGGTCGCGAGGAGGGTA
  #175   ------------------------------------------------------------

        GCCTCTTGGTGGCCATGTTGTCCGGCAGCGCTTTCCTAATATGGTCGCGAGGAGGGTA

for     TGGAAACCTTGTTGCAGATACATCGTGATGGATGCGATCCTATCCACCCGATGGGAAT
MM9     TGGAAACCTTGTTGCAGATACATCGTGATGGATGCGATCCTATCCACCCGATGGGAAT
  #233   ------------------------------------------------------------

        TGGAAACCTTGTTGCAGATACATCGTGATGGATGCGATCCTATCCACCCGATGGGAAT

for     GGGATGGGGTTGGCCAGACTCCGAACTTGTTCGAGAGCACAAGTCGGATTGTAGCGCG
MM9     GGGATGGGGTTGGCCAGACTCCGAACTTGTTCGAGAGCACAAGTCGGATTGTAGCGCG
  #291   ------------------------------------------------------------

        GGGATGGGGTTGGCCAGACTCCGAACTTGTTCGAGAGCACAAGTCGGATTGTAGCGCG

MM2                          CCCTTCTGCTCGATCCAGTCTGTGCATTTGTTCGCTATTCT
for     CGGCAGATGTTCACGTTCCCTTCTGCTCGATCCAGTCTGTGCATTTGTTCGCTATTCT
MM9     CGGCAGATGTTCACGTTCCCTTCTGCTCGATCCAGTCTGTGCATTTGTTCGCTATTCT
  #349   ------------------------------------------------------------

        CGGCAGATGTTCACGTTCCCTTCTGCTCGATCCAGTCTGTGCATTTGTTCGCTATTCT
```

Contig[0008]
prt1 contig

```
MM2       TCTTCCAGGCGGGCACACGGGAAAACGCCCCGAGCCAAATCTACTCTTGGGGGCAGCG
for       TCTTCCAGGCGGGCACACGGGAAAACGCCCCGAGCCAAATCTACTCTTGGGGGCAGCG
MM9       TCTTCCAGGCGGGCACACGGGAAAACGCCCCGAGCCAAATCTACTCTTGGGGGCAGCG
  #407    ------------------------------------------------------------

          TCTTCCAGGCGGGCACACGGGAAAACGCCCCGAGCCAAATCTACTCTTGGGGGCAGCG

MM2       GCCCGATCTCACAATATTTGCTCTTGTTGGGAGAGAAAAAAAAAAGTACTTAAGGGCC
for       GCCCGATCTCACAATATTTGCTCTTGTTGGGAGAGAAAAAAAAAAGTACTTAAGGGCC
MM9       GCCCGATCTCACAATATTTGCTCTTGTTGGGAGAG
  #465    ---------------------------------------------------------

          GCCCGATCTCACAATATTTGCTCTTGTTGGGAGAGAAAAAAAAAAGTACTTAAGGGCC

MM5                                                 ATCAAACTCGGCTTCCTTCGC
MM2       CATGATGCCTGGACATGTTGGTTTGGAAGTTTTCTGCATCAAACTCGGCTTCCTTCGC
for       CATGATGCCTGGACATGTTGGTTTGGAAGTTTTCTGCATCAAACTCGGCTTCCTTCGC
  #523    ---------------------------------------------------------

          CATGATGCCTGGACATGTTGGTTTGGAAGTTTTCTGCATCAAACTCGGCTTCCTTCGC

MM5       ATCACTAAAACAAGAGCACCTCCTCCCAGCAGTTGAGAACCAGAATGTTGAACGTCAA
MM2       ATCACTAAAACAAGAGCACCTCCTCCCAGCAGTTGAGAACCAGAATGTTGAACGTCAA
for       ATCACTAAAACAAGAGCACCTC
  #581    ---------------------------------------------------------

          ATCACTAAAACAAGAGCACCTCCTCCCAGCAGTTGAGAACCAGAATGTTGAACGTCAA

MM5       GAACCTTGTTCTCACGGCGGCGGCGGCGCTTGCTTCGCAGGCCATCGCGGCACCGACT
MM2       GAACCTTGTTCTCACGGCGGCGGCGGCGCTTGCTTCGCAGGCCATCGCGGCACCGACT
  #639    ---------------------------------------------------------

          GAACCTTGTTCTCACGGCGGCGGCGGCGCTTGCTTCGCAGGCCATCGCGGCACCGACT

MM5       GGGCCCGATGCCGGAAACGCCAAGATCCAAGCAGCACAGGGTGGCCAGGTGATTCCTG
MM2       GGGCCCGATGCCGGAAACGCCAAGATCCAAGCAGCACAGGGTGGCCAGGTGATTCCTG
  #697    ---------------------------------------------------------

          GGGCCCGATGCCGGAAACGCCAAGATCCAAGCAGCACAGGGTGGCCAGGTGATTCCTG

MM5       GCAAGTTCATCGTCACGCTGAAGCCCGGCTCCAAGCCAGCAGTGCTCGAGAGCCATAT
MM2       GCAAGTTCATCGTCACGCTGAAGCCCGGCTCCAAGCCAGCAGTGCTCGAGAGCCATAT
  #755    ---------------------------------------------------------

          GCAAGTTCATCGTCACGCTGAAGCCCGGCTCCAAGCCAGCAGTGCTCGAGAGCCATAT
```

Contig[0008]
prt1 contig

```
MM5       GAGATGGGTCAACGGGGTTCACGCAAGAGCATCTGGCGACGAGGCCATCAAGGGAGTG
MM2       GAGATGGGTCAACGGGGTTCACGCAAGAGCATCTGGCGACGAGGCCATCAAGGGAGTG
  #813    ..........................................................
          GAGATGGGTCAACGGGGTTCACGCAAGAGCATCTGGCGACGAGGCCATCAAGGGAGTG

MM5       GAGACCATGTTGGACGGAATTTACGGCTTCATGGGCTACGTCGGTAGTTTCAGTGAGG
MM2       GAGACCATGTTGGACGGAATTTACGGCTTCATGGGCTACGTCGGTAGTTTCAGTGAGG
  #871    ..........................................................
          GAGACCATGTTGGACGGAATTTACGGCTTCATGGGCTACGTCGGTAGTTTCAGTGAGG

rev (p...  GGTTCTGGCCCAGATCAAAGCTCATCCTGACGTTAGTTGAGACTTTTTTTTTTTTTC
MM5       CGGTTCTGGCCCAGATCAAAGCTCATCCTGACGTTAGTTGAGACTTTTTTTTTTTTTC
MM2       CGGTTCTGGCCCAGATCAAAGCTCATCCTGACGTTAGTTGAGACTTTTTTTTTTTTTC
  #929    ..........................................................
          CGGTTCTGGCCCAGATCAAAGCTCATCCTGACGTTAGTTGAGACTTTTTTTTTTTTTC

rev (p...  TTCTCCCCATTCATGATGAGGCATGCTAACATGATGTGATGACTCAGGTCGAGGCTGT
MM5       TTCTCCCCATTCATGATGAGGCATGCTAACATGATGTGATGACTCAGGTCGAGGCTGT
MM2       TTCTCCCCATTCATGATGAGGCATGCTAACATGATGTGATGA
  #987    ..........................................
          TTCTCCCCATTCATGATGAGGCATGCTAACATGATGTGATGACTCAGGTCGAGGCTGT

rev (p...  TGAGCAAGACAAAATCTGGACTCTCGACTGGATCACTGATGACCAGCAACTCGAAGCA
MM4                                                       GAAGCA
MM5       TGAGCAAGACAAAATCTGGACTCTCGACTGGATCACTGATGACCAGCAACTCGAAGCA
  #1045   ..........................................................
          TGAGCAAGACAAAATCTGGACTCTCGACTGGATCACTGATGACCAGCAACTCGAAGCA

rev (p...  AGAGACGACGACAAGGAGCCACCTTCCAGCGGCGGGGGCAGCAACTTCATCCAACAGA
MM4       AGAGACGACGACAAGGAGCCACCTTCCAGCGGCGGGGGCAGCAACTTCATCCAACAGA
MM5       AGAGACGACGACAAGGAGCCACCTTCCAGCGGCGGGGGCAGCAACTTCATCCAACAGA
  #1103   ..........................................................
          AGAGACGACGACAAGGAGCCACCTTCCAGCGGCGGGGGCAGCAACTTCATCCAACAGA

rev (p...  AAAATGCGACATGGGGACTAGGAAGCATCTCTCACCGGGCCCCATATGCCACCGAGTA
MM4       AAAATGCGACATGGGGACTAGGAAGCATCTCTCACCGGGCCCCATATGCCACCGAGTA
MM5       AAAATGCGACATGG
  #1161   ..............
          AAAATGCGACATGGGGACTAGGAAGCATCTCTCACCGGGCCCCATATGCCACCGAGTA
```

Contig[0008]
prt1 contig

```
rev (p...  CGGCTATCAGGAATCTGCCGGGAAGGACACGTACGCCTATGTCATCGACACGGGCATC
MM4        CGGCTATCAGGAATCTGCCGGGAAGGACACGTACGCCTATGTCATCGACACGGGCATC
  #1219    ------------------------------------------------------------

           CGGCTATCAGGAATCTGCCGGGAAGGACACGTACGCCTATGTCATCGACACGGGCATC

rev (p...  CGAACCACGCACGAGGAGTTCGAGGGCCGCGCATCTCACGCCTGGAGCGCGTATCTGA
MM4        CGAACCACGCACGAGGAGTTCGAGGGCCGCGCATCTCACGCCTGGAGCGCGTATCTGA
  #1277    ------------------------------------------------------------

           CGAACCACGCACGAGGAGTTCGAGGGCCGCGCATCTCACGCCTGGAGCGCGTATCTGA

rev (p...  CGAGGACGGACAACGTCGGCCACGGCACCCACGTCGCCGGCACCATCGGCGGCAAGAC
MM4        CGAGGACGGACAACGTCGGCCACGGCACCCACGTCGCCGGCACCATCGGCGGCAAGAC
  #1335    ------------------------------------------------------------

           CGAGGACGGACAACGTCGGCCACGGCACCCACGTCGCCGGCACCATCGGCGGCAAGAC

MM10                                AGCTGCTCGCCGTCAAGATCTTCAACAGCAGGTCG
rev (p...  GTACGGAGTGGCCAAGAACGCCAAGCTGCTCGCCGTCAAGATCTTCAACAGCAGGTCG
MM4        GTACGGAGTGGCCAAGAACGCCAAGCTGCTCGCCGTCAAGATCTTCAACAGCAGGTCG
  #1393    ------------------------------------------------------------

           GTACGGAGTGGCCAAGAACGCCAAGCTGCTCGCCGTCAAGATCTTCAACAGCAGGTCG

MM10       TCCAGCACGTCCGTCATCCTGGCCGGATACAACTGGGCGGTCAACGACATCGTCCGAA
rev (p...  TCCAGCACGTCCGTCATCCTGGCCGGATACAACTGGGCGGTCAACGACATCGTCCGAA
MM4        TCCAGCACGTCCGTCATCCTGGCCGGATACAACTGGGCGGTCAACGACATCGTCCGAA
  #1451    ------------------------------------------------------------

           TCCAGCACGTCCGTCATCCTGGCCGGATACAACTGGGCGGTCAACGACATCGTCCGAA

MM1                           GGGCCGCCATCAACATGTCCCTCGGCGGCCCCAAGTCGACCGC
MM10       AGGGCCGCACCAAGAGGGCCGCCATCAACATGTCCCTCGGCGGCCCCAAGTCGACCGC
rev (p...  AGGGCCGCACCAAGAGGGCCGCCATCAACATGTCCCTCGGCGGCCCCAAGTCGACC
MM4        AGGGCCGCACCAAGAGGGCCGCCATCAAC
  #1509    ------------------------------------------------------------

           AGGGCCGCACCAAGAGGGCCGCCATCAACATGTCCCTCGGCGGCCCCAAGTCGACCGC

MM1        CTTCAACACGGCCGTCGAGAGGGCCTCGGCCTCGGGCGTCTTGTCCATCATCGCCGCC
MM10       CTTCAACACGGCCGTCGAGAGGGCCTCGGCCTCGGGCGTCTTGTCCATCATCGCCGCC
  #1567    ------------------------------------------------------------

           CTTCAACACGGCCGTCGAGAGGGCCTCGGCCTCGGGCGTCTTGTCCATCATCGCCGCC
```

Contig[0008]
prt1 contig

```
MM8                                     ACGTGTCTCCCGCGTCGGCCCCGAGCGCCATCA
MM1       GGCAACGAGGCCCAGGATGCCTCCAACGTGTCTCCCGCGTCGGCCCCGAGCGCCATCA
MM10      GGCAACGAGGCCCAGGATGCCTCCAACGTGTCTCCCGCGTCGGCCCCGAGCGCCATCA
  #1625   . . . . . .. . . . . .. . . . . .. . . . . .. . . . . .. . .

          GGCAACGAGGCCCAGGATGCCTCCAACGTGTCTCCCGCGTCGGCCCCGAGCGCCATCA

MM8       CCGTCGCCGCCATCAATCGCGACTGGACCCTCGCCTCGTACAGCAACTTTGGCTCCGT
MM1       CCGTCGCCGCCATCAATCGCGACTGGACCCTCGCCTCGTACAGCAACTTTGGCTCCGT
MM10      CCGTCGCCGCCATCAATCGCGACTGGACCCTCGCCTCGTACAGCAACTTTGGCTCCGT
  #1683   . . . . . .. . . . . .. . . . . .. . . . . .. . . . . .. . .

          CCGTCGCCGCCATCAATCGCGACTGGACCCTCGCCTCGTACAGCAACTTTGGCTCCGT

MM8       CGTGGACATTTGCGCCCCTGGATCGAACATCACCTCTGCCTGGAACACGGGAGACTCG
MM1       CGTGGACATTTGCGCCCCTGGATCGAACATCACCTCTGCCTGGAACACGGGAGACTCG
MM10      CGTGGACATTTGCGCCCCTGGATCGAACATCACCTCTGCCTGGAACACGGGAGACTCG
  #1741   . . . . . .. . . . . .. . . . . .. . . . . .. . . . . .. . .

          CGTGGACATTTGCGCCCCTGGATCGAACATCACCTCTGCCTGGAACACGGGAGACTCG

MM8       TCCGAGAAGACCATCTCGGGCACCTCCATGGCGACTCCTCATGTTGTCGGCCTCGCTC
MM1       TCCGAGAAGACCATCTCGGGCACCTCCATGGCGACTCCTCATGTTGTCGGCCTCGCTC
MM10      TCCGAGAAGACCATCTCGGGCACCTCCATGGCGACTCCTCATGTTGTCGGCCTCGCTC
  #1799   . . . . . .. . . . . .. . . . . .. . . . . .. . . . . .. . .

          TCCGAGAAGACCATCTCGGGCACCTCCATGGCGACTCCTCATGTTGTCGGCCTCGCTC

MM8       TTTACGCCATCTCCGTGGACGGCGCTACCGGCGTTGACGGCGTCACCAAGCATCTTCT
MM1       TTTACGCCATCTCCGTGGACGGCGCTACCGGCGTTGACGGCGTCACCAAGCATCTTCT
MM10      TTTACGCCATCTCCGTGGACGGCGCTACCGGCGTTGACGGCGTCACCAAGCATCTTCT
  #1857   . . . . . .. . . . . .. . . . . .. . . . . .. . . . . .. . .

          TTTACGCCATCTCCGTGGACGGCGCTACCGGCGTTGACGGCGTCACCAAGCATCTTCT

MM8       GTCAACCGCCACAAAGGACAAAGTTGCCGGCGACACGCGCGGGTCGCCCAATCTGATT
MM1       GTCAACCGCCACAAAGGACAAAGTTGCCGGCGACACGCGCGGGTCGCCCAATCTGATT
MM10      GTCAACCGCCACAAAGGACAAAGTTGCCGGCGACACGCGCGGGTCGCCCAATCTGATT
  #1915   . . . . . .. . . . . .. . . . . .. . . . . .. . . . . .. . .

          GTCAACCGCCACAAAGGACAAAGTTGCCGGCGACACGCGCGGGTCGCCCAATCTGATT
```

Contig[0008]
prt1 contig

```
MM8       GGCAACAACAACAATTCTTACCAGAAGTAGTAAAGCAGTCAGTCAGTCACACGCGTCC
MM1       GGCAACAACAACAATTCTTACCAGAAGTAGTAAAGCAGTCAGTCAGTCACACGCGTCC
MM10      GGCAACAACAACAATTCTTA
  #1973
          GGCAACAACAACAATTCTTACCAGAAGTAGTAAAGCAGTCAGTCAGTCACACGCGTCC

MM7                                                               TTC
MM8       GACTTGGGATCGTGGGCAACGACAAGGATGGCAATTGTAGAGGACCAATAATTTCTTC
MM1       GACTTGGGATCGTGGGCAACGACAAGGATGGCAATTGTAGAGGACCAATAATTTCTTC
  #2031
          GACTTGGGATCGTGGGCAACGACAAGGATGGCAATTGTAGAGGACCAATAATTTCTTC

MM7       TTCTTTATATATACATACATATATATACATAGCACAATATACATGCACCCTCAATGCT
MM8       TTCTTTATATATACATACATATATATACATAGCACAATATACATGCACCCTCAATGCT
MM1       TTCTTTATATATACATACATATATATACATAGCACAATATACATGCACCCTCAATGCT
  #2089
          TTCTTTATATATACATACATATATATACATAGCACAATATACATGCACCCTCAATGCT

MM7       GGTCTCTAAATCGTCAAGTCGTCCACGTTTCGTCGTGATGCAATTTGGCGGGGCGGGG
MM8       GGTCTCTAAATCGTCAAGTCGTCCACGTTTCGTCGTGATGCAATTTGGCGGGGCGGGG
MM1       GGTCTCTAAATCGTCAAGTCGTCCACGTTTCGTCGTGATGCAATTTGGCGGGGCGGGG
  #2147
          GGTCTCTAAATCGTCAAGTCGTCCACGTTTCGTCGTGATGCAATTTGGCGGGGCGGGG

MM3                                 AGATGATGCTATGCCAATCTTTGCTGTCAGGCATTTTC
MM7       AGGTTGATCGGCCGAAAGCGAGATGATGCTATGCCAATCTTTGCTGTCAGGCATTTTC
MM8       AGGTTGATCGGCCGAAAGCGAGATGATGCTATGCCAATCTTTGCTGTCAGGCATTTTC
MM1       AGGTTGATCGGCCGAAAGCGAGATGATGCTATGCCAATCTTTGCTGTCAGGCATTTTC
  #2205
          AGGTTGATCGGCCGAAAGCGAGATGATGCTATGCCAATCTTTGCTGTCAGGCATTTTC

MM3       TTTGGAAACGGTAGATCTCGGCTCCAAGAAGGTTCGTCTCGTTAGGGTGAAGCCTCTG
MM7       TTTGGAAACGGTAGATCTCGGCTCCAAGAAGGTTCGTCTCGTTAGGGTGAAGCCTCTG
MM8       TTTGGAAACGGTAGATCTCGGCTCCAAGAAGGT
MM1       TTTGGAAACGGTAGATCTCGGCTCCAAGAA
  #2263
          TTTGGAAACGGTAGATCTCGGCTCCAAGAAGGTTCGTCTCGTTAGGGTGAAGCCTCTG
```

Contig[0008]
prt1 contig

```
MM3     AATGCTAACTGGGTAGGCGGTGAAAAATCACTGCATATGTTAAGCATATGCCAAAGAC
MM7     AATGCTAACTGGGTAGGCGGTGAAAAATCACTGCATATGTTAAGCATATGCCAAAGAC
 #2321
        AATGCTAACTGGGTAGGCGGTGAAAAATCACTGCATATGTTAAGCATATGCCAAAGAC

MM3     CATAGTGTATTCGTTTAATCCGTTCCATTCGTTCCATTCGTTCCATTCGTTTAATCCG
MM7     CATAGTGTATTCGTTTAATCCGTTCCATTCGTTCCATTCGTTCCATTCGTTTAATCCG
 #2379
        CATAGTGTATTCGTTTAATCCGTTCCATTCGTTCCATTCGTTCCATTCGTTTAATCCG

MM3     TTCTATCCGTTCTATTTGTTTTATTCGTTCCATTTATTCTATTTATTCTATTTATTCT
MM7     TTCTATCCGTTCTATTTGTTTTATTCGCTCCATTTATTCTATTTATTCTATTTATTCT
 #2437
        TTCTATCCGTTCTATTTGTTTTATTCGYTCCATTTATTCTATTTATTCTATTTATTCT

MM3     ATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTT
MM7     ATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTT
 #2495
        ATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTT

MM3     ATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTC
MM7     ATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTC
 #2553
        ATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTCTATTTATTC

MM3     TATTTATTCTATTTATTCCATTTATTCTATTTGTTTTATTTATTCTATTTATTCTATT
MM7     TATTTATTCTATTTATTCCATTTATTCTATTTGCTCTATTTATTCTATTTATTCTATT
 #2611
        TATTTATTCTATTTATTCCATTTATTCTATTTGYTYTATTTATTCTATTTATTCTATT

rev                                                TCATTTGTTCNATTTGTT
MM3     TATTCCATTTGTTTTATCTATTCCATTTGTTCCATTTGTTTCATTTGTTCTATTTGTT
MM7     TATTCCATTTGT
 #2669
        TATTCCATTTGTTTTATCTATTCCATTTGTTCCATTTGTTTCATTTGTTCTATTTGTT
```

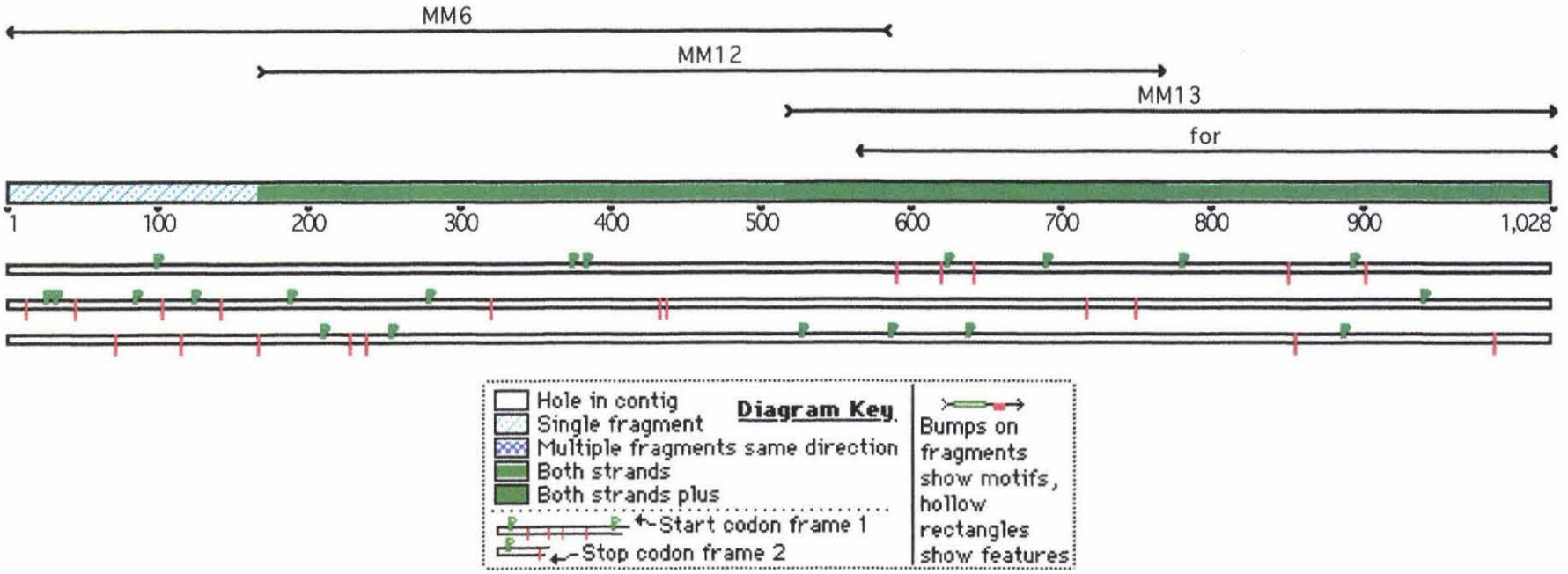
Contig[0008]
prt1 contig

```
rev     CCATTTGTTCAATTATTGTCATATTCGCCCGGAAGAACCTGGCCAGTGGGGANCTGTG
MM3     CCATTTGTTCAATTATTGTCATAATCGCCCGGAGGAACCTGGCCAGTGGGGACCTGTG
 #2727  ..... ..... ..... ..... ..... ..... ..... ..... ..... .....

        CCATTTGTTCAATTATTGTCATAWTCGCCCGGARGAACCTGGCCAGTGGGGACCTGTG

rev     ACACCCACTCCGCAACGGTTAACTTGGCACTTNTTTGTTAAGTTAAAAGTGGCCATTT
MM3     ACACCCACTCCGCAAC
 #2785  ..... ..... ..... ..... ..... ..... ..... ..... ..... .....

        ACACCCACTCCGCAACGGTTAACTTGGCACTTNTTTGTTAAGTTAAAAGTGGCCATTT

rev     GGTGGTCATGGCGAGACTTCGTCGGTAGGCTCAGCNCCGCCAGACATAGCAGCTAGTT
 #2843  ..... ..... ..... ..... ..... ..... ..... ..... ..... .....

        GGTGGTCATGGCGAGACTTCGTCGGTAGGCTCAGCNCCGCCAGACATAGCAGCTAGTT

rev     GTACTAGAGCCCACTTGAATTTTTATGACAGTGATGAATTTGGGCGCGTCGATTTCAT
 #2901  ..... ..... ..... ..... ..... ..... ..... ..... ..... .....

        GTACTAGAGCCCACTTGAATTTTTATGACAGTGATGAATTTGGGCGCGTCGATTTCAT

rev     CCAGAGCCACGTCAAAAGTCTGGCCTGATAAGATTGGCCGGTTAGATGGGCACACGTT
 #2959  ..... ..... ..... ..... ..... ..... ..... ..... ..... .....

        CCAGAGCCACGTCAAAAGTCTGGCCTGATAAGATTGGCCGGTTAGATGGGCACACGTT

rev     AGCTGCCACAGCAGGTTGAGGCGCCTCGAATAGCTTAGCTCGGAAACCATGCTGCTGT
 #3017  ..... ..... ..... ..... ..... ..... ..... ..... ..... .....

        AGCTGCCACAGCAGGTTGAGGCGCCTCGAATAGCTTAGCTCGGAAACCATGCTGCTGT

rev     TGCTGATTGAGAGTATCGGTCGCCTTGGAGTGCTTGTTCTGCTTCAGCCAGCGAAATT
 #3075  ..... ..... ..... ..... ..... ..... ..... ..... ..... .....

        TGCTGATTGAGAGTATCGGTCGCCTTGGAGTGCTTGTTCTGCTTCAGCCAGCGAAATT

rev     CGAGCGTGACGATTACCTTAATCTGAGCGTTATCAGACTGAAAGGGCATTGAGCCCAG
 #3133  ..... ..... ..... ..... ..... ..... ..... ..... ..... .....

        CGAGCGTGACGATTACCTTAATCTGAGCGTTATCAGACTGAAAGGGCATTGAGCCCAG

rev     ATATTTTAGGCCAAAGTAGGCCTTGCCCCACCACGTATGAAGCGATGCGCTCAGACAA
 #3191  ..... ..... ..... ..... ..... ..... ..... ..... ..... .....

        ATATTTTAGGCCAAAGTAGGCCTTGCCCCACCACGTATGAAGCGATGCGCTCAGACAA
```

Contig[0008]
prt1 contig

```
rev       ATCATATTCCATGACTTATCCAAATTGCCAGATACAAAACCGGCGCGAATTC
 #3249    ....................................................
          ATCATATTCCATGACTTATCCAAATTGCCAGATACAAAACCGGCGCGAATTC
```

Contig[0058]
Appendix A3.2 prt2 contig



Contig[0058]
Contig of sequences for the prt2 gene

```
MM6      CTTGGCCAAGTAGTGGACNCTGGACATGGCCATGGCAAATTACTAACTTTGTGCTCCCC
  #1     ------------------------------------------------------------
         CTTGGCCAAGTAGTGGACNCTGGACATGGCCATGGCAAATTACTAACTTTGTGCTCCCC

MM6      AAAACAGGTTGACTACATCGAGAAGGATGCCATCTTCAAGATGAACACTTTTGTTGAGC
  #60     ------------------------------------------------------------
         AAAACAGGTTGACTACATCGAGAAGGATGCCATCTTCAAGATGAACACTTTTGTTGAGC

MM6      AGCGCGATGCTCCTCGGGGATTGAGACGTGTTTCTCACCGCCAGGGTGACATTGGTGGC
MM12                                                   CATTGGTGGC
  #119    ------------------------------------------------------------
         AGCGCGATGCTCCTCGGGGATTGAGACGTGTTTCTCACCGCCAGGGTGACATTGGTGGC

MM6      TACGTTTATCATGCGAGTGCCGGCGAGGGCACATGCTCCTACATTATTGACACTGGAGT
MM12     TACGTTTATCATGCGAGTGCCGGCGAGGGCACATGCTCCTACATTATTGACACTGGAGT
  #178    ------------------------------------------------------------
         TACGTTTATCATGCGAGTGCCGGCGAGGGCACATGCTCCTACATTATTGACACTGGAGT

MM6      TGACGACTCCCACCCTGTATGTCATTTCGTCCAAGTCGATCCCGATGTGCCCAGGTTCT
MM12     TGACGACTCCCACCCTGTATGTCATTTCGTCCAAGTCGATCCCGATGTGCCCAGGTTCT
  #237    ------------------------------------------------------------
         TGACGACTCCCACCCTGTATGTCATTTCGTCCAAGTCGATCCCGATGTGCCCAGGTTCT

MM6      CGCTGGCAAGGCGGACATCCCAACTAACCCGGAGTCGCAGGAGTTCGAGGGTCGCGCTC
MM12     CGCTGGCAAGGCGGACATCCCAACTAACCCGGAGTCGCAGGAGTTCGAGGGTCGCGCTC
  #296    ------------------------------------------------------------
         CGCTGGCAAGGCGGACATCCCAACTAACCCGGAGTCGCAGGAGTTCGAGGGTCGCGCTC

MM6      AGCTCGTCACATCCTTTGTCGATGGGGAGGATGCCGACGGCCACGGTCACGGCACTCAC
MM12     AGCTCGTCACATCCTTTGTCGATGGGGAGGATGCCGACGGCCACGGTCACGGCACTCAC
  #355    ------------------------------------------------------------
         AGCTCGTCACATCCTTTGTCGATGGGGAGGATGCCGACGGCCACGGTCACGGCACTCAC

MM6      GTCGCTGGCACCATCGGTAGCCGTAGCTACGGCATCGCCAAGAAGACTCAGCTGCTTGG
MM12     GTCGCTGGCACCATCGGTAGCCGTAGCTACGGCATCGCCAAGAAGACTCAGCTGCTTGG
  #414    ------------------------------------------------------------
         GTCGCTGGCACCATCGGTAGCCGTAGCTACGGCATCGCCAAGAAGACTCAGCTGCTTGG
```

Contig[0058]
Contig of sequences for the prt2 gene

```
MM6      CATCAAGGTACTCAGCGACCAGGGATCTGGAAACAATTCCGCCATTATCGCGGGCATGG
MM12     CATCAAGGTACTCAGCGACCAGGGATCTGGAAACAATTCCGCCATTATCGCGGGCATGG
MM13                                                 TATCGCGGGCATGG
  #473   ----------------------------------------------------------

         CATCAAGGTACTCAGCGACCAGGGATCTGGAAACAATTCCGCCATTATCGCGGGCATGG

MM6      ACTTTGCCGTCCAGGACGCCAGGCAACGAAGCTGCGCCAAGGGTGTTCTCGCCA
MM12     ACTTTGCCGTCCAGGACGCCAGGCAACGAAGCTGCGCCAAGGGTGTTCTCGCCAACATG
MM13     ACTTTGCCGTCCAGGACGCCAGGCAACGAAGCTGCGCCAAGGGTGTTCTCGCCAACATG
for                                       GCGCCAAGGGTGTTCTCGCCAACATG
  #532   ----------------------------------------------------------

         ACTTTGCCGTCCAGGACGCCAGGCAACGAAGCTGCGCCAAGGGTGTTCTCGCCAACATG

MM12     AGTCTCGGTGGCAGATACTCGCAGTCGCTGAACGATGCGGCCGCTCAGATGATTCAGTC
MM13     AGTCTCGGTGGCAGATACTCGCAGTCGCTGAACGATGCGGCCGCTCAGATGATTCAGTC
for      AGTCTCGGTGGCAGATACTCGCAGTCGCTGAACGATGCGGCCGCTCAGATGATTCAGTC
  #591   ----------------------------------------------------------

         AGTCTCGGTGGCAGATACTCGCAGTCGCTGAACGATGCGGCCGCTCAGATGATTCAGTC

MM12     TGGCGTCTTCCTCGCCGTCGCCGCTGGAAACAATCGCCAGGATGCCTCTGGCTACTCGC
MM13     TGGCGTCTTCCTCGCCGTCGCCGCTGGAAACAATCGCCAGGATGCCTCTGGCTACTCGC
for      TGGCGTCTTCCTCGCCGTCGCCGCTGGAAACAATCGCCAGGATGCCTCTGGCTACTCGC
  #650   ----------------------------------------------------------

         TGGCGTCTTCCTCGCCGTCGCCGCTGGAAACAATCGCCAGGATGCCTCTGGCTACTCGC

MM12     CTGCCTCTGAGCCGAGTGTTTGCACCGTCGGATCAACGGATAGCTCCGACAGTCTCTCT
MM13     CTGCCTCTGAGCCGAGTGTTTGCACCGTCGGATCAACGGATAGCTCCGACAGTCTCTCT
for      CTGCCTCTGAGCCGAGTGTTTGCACCGTCGGATCAACGGATAGCTCCGACAGTCTCTCT
  #709   ----------------------------------------------------------

         CTGCCTCTGAGCCGAGTGTTTGCACCGTCGGATCAACGGATAGCTCCGACAGTCTCTCT

MM12     T
MM13     TCATTCTCCAACTATGGAAGCGTCGTCGATATCCTGGCCCCCGGCTCCGACATTCTTTC
for      TCATTCTCCAACTATGGAAGCGTCGTCGATATCCTGGCCCCCGGCTCCGACATTCTTTC
  #768   ----------------------------------------------------------

         TCATTCTCCAACTATGGAAGCGTCGTCGATATCCTGGCCCCCGGCTCCGACATTCTTTC
```

Contig[0058]
Contig of sequences for the prt2 gene

```
MM13      CACCTGGCCCGGTGGCAGCATCGTAAGTTGAAGCTTCGTCCTTGCCGACCACCGATTCA
for       CACCTGGCCCGGTGGCAGCATCGTAAGTTGAAGCTTCGTCCTTGCCGACCACCGATTCA
  #827
          CACCTGGCCCGGTGGCAGCATCGTAAGTTGAAGCTTCGTCCTTGCCGACCACCGATTCA

MM13      ACATGTTCCATGCCTTGACACTGCCTGCTCTCAGAAAATCCTTTCGGGTACCTCGATGG
for       ACATGTTCCATGCCTTGACACTGCCTGCTCTCAGAAAATCCTTTCGGGTACCTCGATGG
  #886
          ACATGTTCCATGCCTTGACACTGCCTGCTCTCAGAAAATCCTTTCGGGTACCTCGATGG

MM13      CTACTCCCCACATTGTTGGTCTCGCAGCGTATCTTGCTGGTCTAGAGGGCTTCCCAGGC
for       CTACTCCCCACATTGTTGGTCTCGCAGCGTATCTTGCTGGTCTAGAGGGCTTCCCAGGC
  #945
          CTACTCCCCACATTGTTGGTCTCGCAGCGTATCTTGCTGGTCTAGAGGGCTTCCCAGGC

MM13      GCCCAGGCCCTCTGCAAGCGGATCC
for       GCCCAGGCCCTCTGCAAGCGGATCC
  #1004
          GCCCAGGCCCTCTGCAAGCGGATCC
```

Appendix A3.3 Comparison of pMM9 promoter sequence with prt1 promoter

Sequence of the pMM9 promoter is shown in black text, while prt1 promoter is shown in green text. Matching bases are linked by a grey line. The mismatches at -1 and 0 are due to mismatches introduced to the prot1Nco primer to give an *Nco*I restriction site.

```
GGAAAGGGTTCTGCAACGGACGCTAGATGTGGAGCCAGCTTGGTGGCTTTGTCTCTTTCC      60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GGAAAGGGTTCTGCAACGGACGCTAGATGTGGAGCCAGCTTGGTGGCTTTGTCTCTTTCC    -556

TTTCGTTGGATTCCTTTCAATCCTCGCCAAGAGCTCGGGCTCAAAGAAGTCAGGACTGAG      120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TTTCGTTGGATTCCTTTCAATCCTCGCCAAGAGCTCGGGCTCAAAGAAGTCAGGACTGAG    -496

CGAGCAACCGTGGTGGTTGAGTTTCGCGACGATGCACATCATGGCAGCCTCTTGGTGGCC      180
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CGAGCAACCGTGGTGGTTGAGTTTCGCGACGATGCACATCATGGCAGCCTCTTGGTGGCC    -436

ATGTTGTCCGGCAGCGCTTTCCTAATATGGTCGCGAGGAGGGTATGGAAACCTTGTTGCA      240
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATGTTGTCCGGCAGCGCTTTCCTAATATGGTCGCGAGGAGGGTATGGAAACCTTGTTGCA    -376

GATACATCGTGATGGATGCGATCCTATCCACCCGATGGGAATGGGATGGGGTTGGCCAGA      300
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GATACATCGTGATGGATGCGATCCTATCCACCCGATGGGAATGGGATGGGGTTGGCCAGA    -316

CTCCGAACTTGTTCGAGAGCACAAGTCGGATTGTAGCGCGCGGCAGATGTTCACGTTCCC      360
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CTCCGAACTTGTTCGAGAGCACAAGTCGGATTGTAGCGCGCGGCAGATGTTCACGTTCCC    -256

TTCTGCTCGATCCAGTCTGTGCATTTGTTCGCTATTCTTCTTCCAGGCGGGCACACGGGA      420
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TTCTGCTCGATCCAGTCTGTGCATTTGTTCGCTATTCTTCTTCCAGGCGGGCACACGGGA    -196

AAACGCCCCGAGCCAAATCTACTCTTGGGGGCAGCGGCCCGATCTCACAATATTTGCTCT      480
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AAACGCCCCGAGCCAAATCTACTCTTGGGGGCAGCGGCCCGATCTCACAATATTTGCTCT    -136

TGTTGGGAGAGAAAAAAAAAAGTACTTAAGGGCCCATGATGCCTGGACATGTTGGTTTGG      540
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TGTTGGGAGAGAAAAAAAAAAGTACTTAAGGGCCCATGATGCCTGGACATGTTGGTTTGG     -76
.

AAGTTTTCTGCATCAAACTCGGCTTCCTTCGCATCACTAAAACAAGAGCACCTCCTCCCA      600
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AAGTTTTCTGCATCAAACTCGGCTTCCTTCGCATCACTAAAACAAGAGCACCTCCTCCCA     -16

GCAGTTGAGAACCACCATGGG                                             622
|||||||||||||| |||
GCAGTTGAGAACCAGAATGTT                                              5
```

REFERENCES

Abraham, L. D., and Breuil, C. (1995). Factors affecting autolysis of a subtilisin-like serine proteinase secreted by *Ophiostoma piceae* and identification of the cleavage site. Biochim. Biophys. Acta *1245*, 76-84.

Åhman, J., Ek, B., Rask, L., and Tunlid, A. (1996). Sequence analysis and regulation of a gene encoding a cuticle-degrading serine protease from the nematophagous fungus *Arthrobotrys oligospora*. Microbiology *142*, 1605-1616.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (1994). Molecular Biology of the Cell, 3rd edition, B. Alberts, ed. (New York: Garland Publishing).

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. Nucl. Acids Res. *25*, 3389-3402.

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.

Bacon, C. W., and White, J. F. Jr (2000). Physiological adaptions in the evolution of endophytism in the Clavicipitaceae. *In* Microbial Endophytes, C. W. Bacon and J. F. White Jr., eds. (New York: Marcel Dekker, Inc.), pp. 237-261.

Bidochka, M. J., and Khachatourians, G. G. (1994). Protein hydrolysis in grasshopper cuticles by entomopathogenic fungal extracellular proteases. J. Invert. Pathol. *63*, 7-13.

Bonants, P. J. M., Fitters, P. F. L., Thijs, H., de Belder, E., Waalwijk, C., and Henfling, J. W. D. M. (1995). A basic serine protease from *Paecilomyces lilacinus* with biological activity against *Meloidogyne hapla* eggs. Microbiology *141*, 775-784.

Bradshaw, R. E., and Pillar, T. M. (1992). Isolation and Northern blotting of RNA from *Aspergillus nidulans*. J. Microbiol. Meth. *15*, 1-5.

Broekaert, W. F., Cammue, B. P. A., De Bolle, M. F. C., Thevissen, K., De Samblanx, G., and Osborn, R. W. (1997). Antimicrobial peptides from plants. Cri. Rev. Plant Sci. *16*, 297-323.

Bucheli, E., and Leuchtmann, A. (1996). Evidence for genetic differentiation between choke-inducing and asymptomatic strains of the *Epichloë* grass endophyte from *Brachypodium sylvaticum*. Evolution *50*, 1879-1887.

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. Jr, Bowdish, T. I., and Welch, A. M. (1998). A new kind of mutualism between fungi and insects. Mycol. Res. *102*, 235-238.

Bush, L. P., Fannin, F. F., Siegel, M. R., Dahlman, D. L., and Burton, H. R. (1993). Chemistry, occurrence and biological effects of saturated pyrrolizidine alkaloids associated with endophyte-grass interactions. Agric. Ecosyst. Environ. *44*, 1-4.

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

Carroll, A. M., Sweigard, J. A., and Valent, B. (1994). Improved vectors for selecting resistance to hygromycin. Fungal Genetics Newsletter *41*, 22.

Carroll, G. (1988). Fungal endophytes in stems and leaves: from latent pathogen to mutualistic symbiont. Ecology *69*, 2-9.

Cheevadhanarak, S., Renno, D. V., Saunders, G., and Holt, G. (1991). Cloning and selective overexpression of an alkaline protease-encoding gene from *Aspergillus oryzae*. Gene *108*, 151-155.

Cheplick, G. P., Clay, K., and Marks, S. (1989). Interactions between infection by endophytic fungi and nutrient limitation in the grasses *Lolium perenne* and *Festuca arundinacea*. New Phytol. *111*, 89-97.

Chiang, T.-Y., Rai, R., Cooper, T. G., and Marzluf, G. A. (1994). DNA binding site specificity of the *Neurospora* global nitrogen regulatory protein NIT2: Analysis with mutated binding sites. Mol. Gen. Genet. *245*, 512-516.

Christensen, M. J. (1995). Variation in the ability of *Acremonium* endophytes of *Lolium perenne*, *Festuca arundinacea* and *F. pratensis* to form compatible associations in the three grasses. Mycol. Res. *99*, 466-470.

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). Vegetative compatibility between and within *Epichloë* species. Mycologia *89*, 558-565.

Clay, K. (1988). Fungal endophytes of grasses: a defensive mutualism between plants and fungi. Ecology *69*, 10-16.

Clay, K. (1991). Fungal endophytes, grasses and herbivores. *In* Microbial mediation of plant-herbivore interactions., P. Barbosa, V. A. Krischik and C. J. Jones, eds. (New York: John Wiley & Sons, Inc.), pp. 199-226.

De Battista, J. P., Bacon, C. W., Severson, R. F., Plattner, R. D., and Bouton, J. H. (1990). Indole acetic acid production by the fungal endophyte of tall fescue. Agron. J. *82*, 878-880.

Dobson, J. M. (1997). Molecular characterisation of the hmg CoA reductase gene from *Neotyphodium lolii*. M. Sc. Thesis, Department of Microbiology and Genetics Palmerston North, Massey University, pp. 136 leaves.

Felenbok, B., and Kelly, J. M. (1996). Regulation of carbon metabolism in mycelial fungi. *In* The Mycota Vol. III Biochemistry and Molecular Biology, Brambl and G. A. Marzluf, eds. (Berlin: Springer-Verlag), pp. 369-380.

Franks, D. W., Bush, L. P., and Phillips, T. D. (1997). Aminopyrrolizidine production by *Neotyphodium*. *In* Proceedings International *Neotyphodium*/Grass Interactions Symposium (Athens, Georgia, U. S. A.

Froeliger, E. H., and Carpenter, B. E. (1996). NUT1, a major nitrogen regulatory gene in *Magnaporthe grisea*, is dispensable for pathogenicity. Mol. Gen. Genet. *251*, 647-656.

Fu, Y. H., and Marzluf, G. A. (1990). *nit*-2, the major nitrogen regulatory gene of *Neurospora crassa*, encodes a protein with a putative zinc finger DNA-binding domain. Mol. Cell. Biol. *10*, 1056-1065.

Geremia, R. A., Goldman, G. H., Jacobs, D., Ardiles, W., Vila, S. B., Van Montagu, M., and Herrera-Estrella, A. (1993). Molecular characterisation of the proteinase-encoding gene, *prb1*, related to mycoparasitism by *Trichoderma harzianum*. Mol. Microbiol. *8*, 603-613.

Glenn, A. E., Bacon, C. W., Price, R., and Hanlin, R. T. (1996). Molecular phylogeny of *Acremonium* and its taxonomic implications. Mycologia *88*, 369-383.

Gunkel, F. A., and Gassen, H. G. (1989). Proteinase K from *Tritirachium album* Limber. Characterization of the chromosomal gene and expression of the cDNA in *Escherichia coli*. Eur. J. Biochem. *179*, 185-194.

Gurr, S. J., Unkles, S. E., and Kinghorn, J. K. (1987). The structure and organisation of nuclear genes in filamentous fungi. *In* Gene structure in eukaryotic microbes, J. K. Kinghorn, ed. (Oxford: IRL Press), pp. 93-139.

Gwinn, K. D., and Gavin, A. M. (1992). Relationship between endophyte infestation level of tall fescue seed lots and *Rhizoctonia zeae* seedling disease. Plant Disease *76*, 911-914.

Haas, H., Bauer, B., Redl, B., Stöffler, G., and Marzluf, G. A. (1995). Molecular cloning and analysis of *nre*, the major nitrogen regulatory gene of *Penicillium chrysogenum*. Curr. Genet. *27*, 150-158.

Haber, J. E., and Louis, E. J. (1998). Minisatellite origins in yeast and humans. Genomics *48*, 132-135.

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

Hotter, G. S. (1996). Molecular signalling events in the activation of plant defenses against pathogens. NZ BioScience *1996*, 7-12.

Hult, K., Veide, A., and Gatenbeck, S. (1980). The distribution of the NADPH regenerating mannitol cycle among fungal species. Arch. Microbiol. *128*, 253-255.

Itoh, Y., and Scott, B. (1994). Heterologous and homologous plasmid integration at a spore-pigment locus in *Penicillium paxilli* generates large deletions. Curr. Genet. *26*, 468-476.

Jackson, A. O., and Taylor, C. P. (1996). Plant-microbe interactions: life and death at the interface. Plant Cell *8*, 1651-1668.

Joshi, L., St Leger, R. J., and Bidochka, M. J. (1995). Cloning of a cuticle-degrading protease from the entomopathogenic fungus, *Beauveria bassiana*. FEMS Microbiol. Lett. *125*, 211-217.

Joshi, L., St Leger, R. J., and Roberts, D. W. (1997). Isolation of a cDNA encoding a novel subtilisin-like protease (Pr1B) from the entomopathogenic fungus, *Metarhizium anisopliae* using differential display-RT-PCR. Gene *197*, 1-8.

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

Koga, H., Christensen, M. J., and Bennett, R. J. (1993). Incompatibility of some grass-*Acremonium* endophyte associations. Mycol. Res. *97*, 1237-1244.

Kokoska, R. J., Stefanovic, L., Buermeyer, A. B., Liskay, R. M., and Petes, T. D. (1999). A mutation of the yeast gene encoding PCNA destabilizes both microsatellite and minisatellite DNA sequences. Genetics *151*, 511-519.

Kulkarni, R. K., and Nielsen, B. D. (1986). Nutritional requirements for growth of a fungus endophyte of tall fescue grass. Mycologia *78*, 781-786.

Lam, C. K., Belanger, F. C., White, J. F., and Daie, J. (1994). Mechanism and rate of sugar uptake by *Acremonium typhinum*, and endophytic fungus infecting *Festuca rubra*: evidence for presence of a cell wall invertase in endophytic fungi. Mycologia *8*, 408-415.

Leuchtmann, A. (1994). Isozyme characterisation, persistence and compatibility of fungal and grass mutualists. *In* Biotechnology of endophytic fungi of grasses, C. W. Bacon and J. F. White Jr., eds. (London: CRC Press, Inc.), pp. 21-33.

Leuchtmann, A., and Clay, K. (1990). Isozyme variation in the *Acremonium/Epichloë* fungal endophyte complex. Phytopathol. *80*, 1133-1139.

Leuchtmann, A., and Clay, K. (1997). The population biology of grass endophytes. *In* The Mycota Volume V part A Plant Relationships. G. C. Carroll and P. Tudzynski, eds. (Berlin: Springer Verlag), pp. 185-202.

Lindstrom, J. T., and Belanger, F. C. (1994). Purification and characterisation of an endophytic fungal proteinase that is abundantly expressed in the infected host grass. Plant Physiol. *106*, 7-16.

Lindstrom, J. T., Sun, S., and Belanger, F. C. (1993). A novel fungal protease expressed in endophytic infection of *Poa* sepcies. Plant Physiol. *102*, 645-650.

Lopez-Llorca, L. V. (1990). Purification and properties of extracellular proteases produced by the nematophagous fungus *Verticillium suchlasporium*. Can. J. Microbiol. *36*, 530-537.

McGowan, T. L. (1996). Construction of a novel fungal gus expression plasmid, and its evaluation in *Aspergillus nidulans*. M. Sc. thesis, Department of Microbiology and Genetics, Palmerston North, Massey University, pp. 122 leaves.

Meijer, G., and Leuchtmann, A. (1999). Multistrain infections of the grass *Brachypodium sylvaticum* by its fungal endophyte *Epichloë sylvatica*. New Phytol. *141*, 355-368.

Morgan-Jones, G., and Gams, W. (1982). Notes on Hyphomycetes. XLI. An endophyte of *Festuca arundinacea* and the anamorph of *Epichloë typhina*, new taxa in one of two new sections of *Acremonium*. Mycotaxon *15*, 311-318.

Negrete-Urtasun, S., Reiter, W., Diez, E., Denison, S. H., Tilburn, J., Espeso, E. A., Peñalva, M. A., and Arst, H. N. Jr (1999). Ambient pH signal transduction in *Aspergillus*: completion of gene characterisation. Mol. Microbiol. *33*, 994-1003.

Norrander, J., Kempe, T., and Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene *26*, 101-106.

Orejas, M., Espeso, E. A., Tilburn, J., Sarkar, S., Arst, H. N. Jr, and Peñalva, M. A. (1995). Activation of the *Aspergillus* PacC transcription factor in response to alkaline ambient pH requires proteolysis of the carboxy-terminal moiety. Genes Dev. *9*, 1622-1632.

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

Powell, R. G., TePaske, M. R., Plattner, R. D., White, J. F., and Clement, S. L. (1994). Isolation of resveratrol from *Festuca versuta* and evidence for the widespread occurrence of this stilbene in the Poaceae. Phytochemistry *35*, 335-338

Premakumar, R., Sorger, G. J., and Gooden, D. (1979). Nitrogen metabolite repression of nitrate reductase in *Neurospora crassa*. J. Bacteriol. *137*, 1119-1126.

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

Reddy, P. V., Lam, C. K., and Belanger, F. C. (1996). Mutualistic fungal endophytes express a proteinase that is homologous to proteases suspected to be important in fungal pathogenicity. Plant Physiol. *111*, 1209-1218.

Richardson, M. D., Chapman, G. W. Jr., Hoveland, C. S., and Bacon, C. W. (1992). Sugar alcohols in endophyte-infected tall fescue under drought. Crop Sci. *32*, 1060-1061.

Roberts, C. A., Marek, S. M., Niblack, T. L., and Karr, A. L. (1992). Parasitic *Meloidogyne* and mutualistic *Acremonium* increase chitinase in tall fescue. J. Chem. Ecol. *18*, 1107-1116.

Rowan, D. D. (1993) Lolitrems, peramine and paxilline - mycotoxins of the ryegrass-endophyte interaction. Agric. Ecosyst. Environ. *44*, 103-122

Samal, B. B., Karan, B., Boone, T. C., Chen, K. K., Rohde, M. F., and Stabinsky, Y. (1989). Cloning and expression of the gene encoding a novel proteinase from *Tritirachium album* limber. Gene *85*, 329-333.

Samal, B. B., Karan, B., Boone, T. C., Osslund, T. D., Chen, K. K., and Stabinsky, Y. (1990). Isolation and characterisation of the gene encoding a novel, thermostable serine proteinase from the mould *Tritirachium album Limber*. Mol. Microbiol. *4*, 1789-1792.

Saunders, K. (1997). Development of the ß-glucuronidase reporter gene system to study *Acremonium* endophyte interactions with perennial ryegrass. M. Sc. thesis, Department of Microbiology and Genetics, Palmerston North, Massey University, pp. 117 leaves.

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

Schardl, C. L. (1997). Protective grass endophytes: where are they and where are they going? Plant Disease *81*, 430-438.

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 hybridisation of a mutualist with the ryegrass choke pathogen, *Epichloë typhina*. Genetics *136*, 1307-1317.

Schardl, C. L., Liu, J.-S., White, J. F., Jr., Finkel, R. A., An, Z.-Q., and Siegel, M. R. (1991). Molecular phylogenetic relationships of nonpathogenic grass mycosymbionts and clavicipitaceous plant pathogens. Plant Syst. Evol. *178*, 27-41.

Screen, S., Bailey, A., Charnley, K., Cooper, R., and Clarkson, J. (1998). Isolation of a nitrogen response regulator gene (*nrr*1) from *Metarhizium anisopliae*. Gene *221*, 17-24.

Segers, R., Butt, T. M., Carder, J. H., Keen, J., Kerry, B. R., and Peberdy, J. F. (1999). The subtilisins of fungal pathogens of insects, nematodes and plants: distribution and variation. Mycol. Res. *103*, 395-402.

Segers, R., Butt, T. M., Kerry, B. R., and Peberdy, J. F. (1994). The nematophagous fungus *Verticillium chlamydosporium* produces a chymoelastase-like protease which hydrolyses host nematode proteins *in situ*. Microbiology *140*, 2715-2723.

Siegel, M. R., and Bush, L. P. (1997). Toxin production in grass/endophyte associations. *In* The Mycota, Vol V part A Plant Relationships. G. C. Carroll and P. Tudzynski, eds. (Berlin: Springer-Verlag), pp. 185-207.

Siegel, M. R., Jarlfors, U., Latch, G. C. M., and Johnson, M. C. (1987). Ultrastructure of *Acremonium coenophialum, Acremonium lolii* and *Epichloë typhina* endophytes in host and nonhost *Festuca* and *Lolium* species of grasses. Can. J. Bot. *65*, 2357-2367.

Siegel, M. R., and Latch, G. C. M. (1991). Expression of antifungal activity in agar culture by isolates of grass endophytes. Mycologia *83*, 529-537.

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

Siegel, M. R., and Schardl, C. L. (1991). Fungal endophytes of grasses: detrimental and beneficial associations. *In* Microbial Ecology of Leaves, J. H. Andrew and S. S. Hirano, eds. (Berlin: Springer Verlag), pp. 198-221.

Siezen, R. J., and Leunissen, J. A. M. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. Protein Sci. *6*, 501-523.

Siezen, R. J., Leunissen, J. A. M., and Shinde, U. (1995). Homology analysis of the propeptides of subtilisin-like serine proteases (subtilases). *In* Intramolecular chaperones and protein folding, U. Shinde and M. Inouye, eds (New York: Springer), pp. 233-256.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel elctrophoresis. J. Mol. Biol. *98*, 503-517.

Sreedhar, L., Kobayashi, D. Y., Bunting, T. E., Hillman, B. I., and Belanger, F. C. (1999). Fungal proteinase expression in the interaction of the plant pathogen *Magnaporthe poae* with its host. Gene *235*, 121-129.

St Leger, R. J., Bidochka, M. J., and Roberts, D. W. (1994). Isoforms of the cuticle-degrading Pr1 proteinase and production of a metalloproteinase by *Metarhizium anisopliae*. Arch. Biochem. Biophys. *313*, 1-7.

St Leger, R. J., Frank, D. C., Roberts, D. W., and Staples, R. C. (1992). Molecular cloning and regulatory analysis of the cuticle-degrading-protease structural gene from the entomopathogenic fungus *Metarhizium anisopliae*. Eur. J. Biochem. *204*, 991-1001.

St Leger, R. J., Joshi, L., and Roberts, D. W. (1997). Adaptation of proteases and carbohydrases of saprophytic, phytopathogenic and entomopathogenic fungi to the requirements of their ecological niches. Microbiology *143*, 1983-1992.

St Leger, R. J., Staples, R. C., and Roberts, D. W. (1991). Changes in translatable mRNA species associated with nutrient deprivation and protease synthesis in *Metarhizium anisopliae*. J. Gen. Microbiol. *137*, 807-815.

Strickland, J. R., Oliver, J. W., and Cross, D. L. (1993). Fescue toxicosis and its impact on animal agriculture. Vet. Hum. Toxicol. *35*, 454-464.

Terada, I., Kwon, S. T., Miyata, Y., Matsuzawa, H., and Ohta, T. (1990). Unique precursor structure of an extracellular protease, aqualysin I, with NH2- and COOH-terminal pro-sequences and its processing in *Escherichia coli*. J. Biol. Chem. *265*, 6576-6581.

Thuring, R. W. J., Sanders, J. P. M., and Borst, P. (1975). A freeze squeeze method for recovering long DNA from agarose gels. Anal. Biochem. *66*, 213-220.

Tilburn, J., Sarkar, S., Widdick, D. A., Espeso, E. A., Orejas, M., Mungroo, J., Peñalva, M. A., and Arst, H. N. Jr. (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.

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 hybridisation with *Epichloë* species. Proc. Ntl. Aca. Sci USA *91*, 2542-2546.

Tudzynski, B., Homann, V., Feng, B., and Marzluf, G. A. (1999). Isolation, characterization and disruption of the areA nitrogen regulatory gene of *Gibberella fujikuroi*. Mol. Gen. Genet. *261*, 106-114.

Tunlid, A., Rosén, S., Ek, B., and Rask, L. (1994). Purification and characterisation of an extracellular serine protease from the nematode-trapping fungus *Arthrobotrys oligospora*. Microbiology *140*, 1687-1695.

van Heeswijck, R., and McDonald, G. (1992). *Acremonium* endophytes in perennial ryegrass and other pasture grasses in Australia and New Zealand. Aust. J. Agric. Res. *43*, 1683-1709.

Vieira, J., and Messing, J. (1987). Production of single-stranded plasmid DNA. Methods Enzymol. *153*, 3-11.

West, C. P. (1994). Physiology and drought tolerance of endophyte-infected grasses. *In* Biotechnology of endophytic fungi of grasses, C. W. Bacon and J. F. J. White, eds. (London: CRC Press, Inc.), pp. 87-99.

Wilkinson, H. H., Bush, L. P., Burton, H. R., Siegel, M. R., and Schardl, C. (1997). Lolines are fungal metabolites. *In* Proceedings International *Neotyphodium*/Grass Interactions Symposium (Athens, Georgia, U. S. A).

Yelton, M. M., Harner, J. E., and Timberlake, W. E. (1984) Transformation of *Aspergillus nidulans* by using a trpC plasmid. Pro. Ntl. Aca. Sci. USA *81*, 1470-1474.

Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. MMBR *63*, 405-445.