MOLECULAR CHARACTERISATION OF THE HMG CoA REDUCTASE GENE 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.

Joanne Margaret Dobson

1997

ABSTRACT

3-Hydroxy-3-methylglutaryl Coenzyme A reductase (Hmg) catalyses the conversion of HMG CoA to mevalonic acid; the first step of the isoprenoid biosynthetic pathway. This pathway produces a wide variety of primary metabolites which are involved in many different cellular processes. *Neotyphodium* endophytes in association with the grass host are known to produce a range of secondary metabolites including the indole diterpenoids (eg paxilline and lolitrem) and the ergopeptine alkaloids (eg ergovaline). Given that these pathways are upregulated *in planta* the availability of mevalonic acid, be it from fungal or plant source, will be important in controlling the levels of the different toxins synthesised. The aim of this work was to clone the fungal endophyte *hmg* and characterise the promoter to enable study of its regulation *in planta* via reporter gene studies.

Using degenerate primers designed against conserved regions of other *hmg* genes a 359 bp fragment was amplified from the *Neotyphodium lolii* isolate Lp19, which grows in perennial ryegrass (*Lolium perenne*). DNA sequencing confirmed that the sequence amplified was part of a unique *hmg* gene. Southern hybridisations suggest that there is a single copy of *hmg* in strain Lp19 (a haploid) but two copies in strain Lp1 (an interspecific hybrid; Schardl *et al.* 1994). The fragment of *N. lolii hmg* was used to screen a λGEM-12 genomic library of Lp19 and four positive overlapping clones were isolated.

Fragments of one clone, λJD12, were subcloned, sequenced and a physical map of this region of the genome was constructed. The entire sequence of *hmg* was determined using primer walking and was found to encode a 1188 amino acid polypeptide. From comparison to other Hmg proteins the catalytic domain has been shown to be highly conserved while the amino-terminal domain, containing transmembrane regions is divergent with very little sequence similarity near the translation start site and promoter region.

Using RT-PCR analysis the *hmg* gene was shown to consist of two open reading frames separated by a 73 bp intron. RT-PCR was also used to determine the location of the transcriptional start site and this is supported by the presence of putative CAAT and TATA consensus sequences. With the promoter region identified and characterised further analysis of the regulation of *hmg in planta* can be undertaken.

ACKNOWLEDGEMENTS

Firstly, I would like to convey my thanks and appreciation to my supervisor, Barry Scott, for his kind words and encouragement throughout my time in the lab. As an excellent supervisor you have enabled me to grow in confidence as a scientist. I also would like to thank Carolyn Young, without whose assistance and support I would not have achieved this goal. Thank you for taking time to help me with my work, personal life and in general offering me a shoulder to lean on.

I would also like to thank all the Scott Base members, past and present, who have made my time in the lab an enjoyable experience. I have appreciated your friendship and advice. I would like to acknowledge Austen Ganley, Lisa McMillan and David McSweeney for their computing assistance, Mike Collett for his guidance during the first year of this project, and Christina Moon for her technical advice. I would also like to offer thanks and appreciation to my good friends Bek and Em, whose attempts at keeping me sane over the past year and a half have not gone unnoticed. I would also like to thank Bek, Em and Carolyn for taking the time to proof read my work, and Grant Hotter for his work on defining the transcription start site location.

A big thank you must go to my parents for their support throughout my many years at Massey and also thanks heaps to my friends for their patience and encouragement. I would like to offer special thanks to my friend Selina. You have always been there for me, providing me with a 'science-free' environment to escape to. I appreciated it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 GRASS ENDOPHYTES

The filamentous fungi of the Ascomycete family Clavicipitaceae (tribe Balansieae) form symbiotic associations with grasses of the Pooideae subfamily. *Epichloë* endophytes are natural agents of biological protection, and infection can be important for fitness, competitiveness and persistence of their grass hosts under conditions of biotic or abiotic stress. These symbionts inhabit reproductive and aerial host tissue, growing as sparsely branched hyphae in intercellular spaces. Nutrients are absorbed by these organisms from freely available materials (Clay 1990).

1.1.1 Life Cycle

Neotyphodium (formerly *Acremonium*) endophytes are anamorphic (asexual) forms of the ascomycetous *Epichloë* spp. and share many features with them. *Epichloë* spp. are the causative agents of "choke" disease of grasses and can be seed transmitted when the "choke" stage is not exhibited. Chokes are sporogenous stromata produced in association with host inflorescences, preventing seed production by arresting development of the florets. At this stage, *E. typhina* can undergo sexual crosses, producing ascospores which are, presumably, capable of infecting new host plants. Following fertilisation of *E. typhina* stromata by conidia of the opposite mating type, perithecia develop on this structure and release haploid ascospores when mature. The ascospores germinate to produce phialides and conidia, which in turn germinate to produce hyphae and more conidiophores. After the last cycle of this conidiation, the mycelia are thought to infect the host via the stigmata, leaf stem stomata and/or ancillary buds. In addition, meristematic tissue must become infected for survival of the fungus (Siegel and Schardl 1991). It is this spore forming stage which the asexual endophytes lack. The life cycle of the *Neotyphodium* endophytes occurs within organs of the host grass without eliciting any symptoms of disease and the maternal-line transmission due to seed dissemination of the endophytes is a highly efficient process, removing the need for spore dissemination (Schardl and An 1993).

Genetic analysis based on isozyme profiles and DNA sequence comparisons supports the theory of being multiple evolutionary origins of the asexual endophytes from *Epichloë* species, and even isolates from the same grass species can show considerable genetic diversity. Perennial ryegrass is a known host of three species or taxa in the *Epichloë* group. The two known taxa of asexual, seed-borne endophytes are *Neotyphodium lolii* (Latch *et al.* 1984) also classified as taxonomic grouping one from *Lolium perenne* (LpTG-1) and LpTG-2 (Christensen *et al.* 1993). LpTG-2 is a heteroploid derived from an interspecific hybridisation whose most likely ancestors were *N. lolii* and *E. typhina* (Schardl *et al.* 1994).

Extensively studied grass-endophyte interactions include the tall fescue (*Festuca arundinacea*) and perennial ryegrass (*Lolium perenne*) with the endophytes *Neotyphodium coenophialum* and *Neotyphodium lolii* respectively. These two grass species are major forage crops in grazing agro-ecosystems, with endophyte infected ryegrass and tall fescue being widespread in New Zealand and the United States, respectively. The importance of the endophytes to these grasses has recently led to their intense investigation.

1.1.2 Biological Benefits of Endophyte/Plant Interaction

Endophytic fungi have dramatic effects on the physiology, ecology and reproductive biology of host plants. The presence of the endophyte increases drought tolerance and leads to changes of grass morphology, growth rate and growth yield. The production of toxic alkaloids enable endophytic fungi to defend their host against mammalian and insect herbivory, and offers resistance to fungal pathogens and nematodes. This contributes to the excellent field persistence of these grasses with little metabolic cost to the plant (Clay 1990).

Endophyte infection alters the host plant physiology, providing greater vegetative vigour. More inflorescences and seeds are produced and the seeds are able to germinate more rapidly, with faster growing seedlings. Infected seeds have a higher survival rate as they contain high concentrations of alkaloids and so are less likely to be eaten. Endophyte infected tall fescue has been found to maintain a higher productivity than uninfected plants under drought conditions. This is due to the infected plants maintaining a higher net photosynthetic rate, as well as having greater stomatal resistance (Clay 1990).

A number of alkaloids are produced by endophytes within plant tissues and these are active against insects, fungal pathogens, nematodes and mammals. These include peramine, ergovaline, lolines and the indole diterpenoids, paxilline and the lolitrems.

The pyrrolizidine alkaloids (*N*-formyl loline and *N*-acetyl loline), detected in *A. coenophialum*-infected tall fescue, are the most abundant secondary metabolites that have been identified and the most important bioprotective alkaloids known to be involved in many grass-endophyte associations. These compounds have been detected in very high levels in infected plants (> 0.8% plant dry weight) and have been shown in culture to be fungal metabolites (Wilkensen *et al.* 1997). The production of these alkaloids is dependent on the fungal strain (Siegel *et al.* 1990) and are known to be potent insecticides. Their contribution to host fitness may be due to their ability to disrupt physiological processes controlled by biogenic amines (Scott and Schardl 1993).

Peramine, a pyrrolopyrazine alkaloid, is produced by a number of endophytes including *N. lolii*, *N. coenophialum* and *E. typhina* strains. This is a potent insect feeding deterrent which is thought to be derived from proline and arginine. This alkaloid has been found to provide resistance to the Argentine stem weevil (*Listronotus bonariensis*), a major pest of pastures in New Zealand. While the mechanism by which this is achieved remains unclear, the alkaloid has not been shown to be a mammalian toxin. Peramine has not been detected in fungal cultures or in uninfected plants, and therefore may be unique to the mutualistic association (Rowan and Gaynor 1986).

Endophyte infection has also been shown to negatively influence populations of certain nematodes, thus offering the plant protection against nematode infection (Kimmons *et al.* 1990). There are contrasting reports regarding the effects of endophytes on plant parasitic nematodes but this variation in effect may reflect the relative sensitivities of different species to the toxins produced. Anti-fungal activity has also been reported in endophyte infected grasses. Resistance to pathogens has not been widely demonstrated, however isolation of endophytes which confer resistance to important fungal diseases would provide significant potential for improving the productivity of perennial ryegrass pastures (van Heeswijck and McDonald 1992).

A. coenophialum infection of tall fescue produces ergopeptine alkaloids, in particular ergovaline, which is thought to be responsible for fescue toxicosis. Cattle grazing on infected tall fescue present a number of symptoms which are clinically identical with ergot poisoning. This condition is usually observed in summer, with cattle showing a range of symptoms which include, reduced weight gain, decreased milk production, lower feed intake, rough hair-coat, excessive salivation, increased respiration and high

temperatures. Fescue foot is another syndrome observed in cattle that have been feeding on infected tall fescue. This occurs most frequently in winter causing lameness and dry gangrene of extremities which can, in severe cases, lead to loss of hooves (van Heeswijck and McDonald 1992). The biochemical mechanisms responsible for these syndromes are not known, but the ergot alkaloids are known to interact with the α-adreno and dopamine receptors in the brain, and these compounds have vascoconstrictive effects. Ergovaline is thought to be composed of a tricyclic peptide and an ergolene ring, synthesised from the precursors mevalonic acid and tryptophan. Ergopeptines are found in a number of grass-endophyte interactions, including perennial ryegrass associations (Christensen *et al.* 1993).

Cattle grazing on *N. lolii* infected perennial ryegrass suffer from the neurological disorder 'ryegrass staggers'. The potent neurotoxin lolitrem A, and other indole diterpenoid lolitrems, have been isolated from perennial ryegrass (*Lolium perenne*) pastures on which this livestock disease occurs. These compounds have been shown to produce tremors in livestock and therefore may be the causative agents of the disease (Gallagher *et al.* 1982; Gallagher *et al.* 1981). These compounds are tremorgenic mycotoxins which are thought to bind to the $GABA_A$ receptor in the brain and inhibit its function (Gant *et al.* 1987), thus causing affected animals to display muscle spasms and a hypersensitivity to external stimuli. Little is known about the biochemical pathways leading to the synthesis of these compounds, although putative intermediates paxilline, paspaline and lolitriol have been identified and the outlines of a biosynthetic pathway suggested (Miles *et al.* 1992; Munday-Finch *et al.* 1996). Paxilline, a known tremorgen, is thought to be derived from tryptophan and mevalonic acid, and produced by a series of modifications of geranyl geranyl pyrophosphate. Cultures of *N. lolii* produce a reduced amount of lolitrem compared to amounts detected *in planta*, suggesting that the host plant may have a role in biosynthesis.

Endophyte concentrations vary throughout the host plant, with higher concentrations of hyphae and metabolic activity in lower parts. Distribution is also high in the leaf sheath in comparison to the leaf blade. There are also concentration differences between leaves of different ages: younger leaves contain lower *N. lolii* concentrations, perhaps reflecting the pattern and stage of development of fungal and plant tissue (Herd *et al.* 1997; Keogh *et al.* 1996). The distribution of alkaloids also varies throughout the host plant, with lolitrem B being present in higher concentrations in the leaf sheath than the blade and in particular, the outer leaf sheath. High lolitrem B concentrations have also been found in the older leaves, mostly in the upper portion of the leaf sheath, where the concentration of *N. lolii* was the highest. This distribution pattern correlates with livestock grazing patterns, the acquisition of the neurotoxin and development of

ryegrass staggers. Peramine was found in concentrations greater than the 10 ppm threshold which deters feeding by adult and larval Argentine stem weevil, in all portions of the grass except the tip of the older blades. As peramine is produced by *N. lolii*, it is apparent that this fungal metabolite may be mobile within the ryegrass tiller. Argentine stem weevil have been noted to often feed only on the top portion of ryegrass leaves, but it is not clear whether this behaviour is affected by the presence of *N. lolii* (Keogh *et al.* 1996).

As the above compounds are not often detected in fungal culture or in endophyte free grass, it seems likely that molecular interactions occur between the grass host and endophytic fungus which result in the increased expression of these compounds within the symbiotic association. Thus these alkaloid production pathways provide an area for investigation into the symbiotic relationship between grasses and endophytes. The carbon skeleton for both indole diterpenoids, paxilline and lolitrem B, and the ergot alkaloid ergovaline, is derived from long chain isoprenoid units which are in turn derived from mevalonic acid. High level synthesis of these compounds with therefore be dependant on an abundant supply of mevalonic acid. The synthesis of mevalonic acid is catalysed by 3-hydroxy-3-methylglutaryl Coenzyme A reductase (Hmg); a key primary metabolic step in isoprenoid biosynthesis and therefore toxin production (Figure 1).

1.2 HMG-CoA REDUCTASE

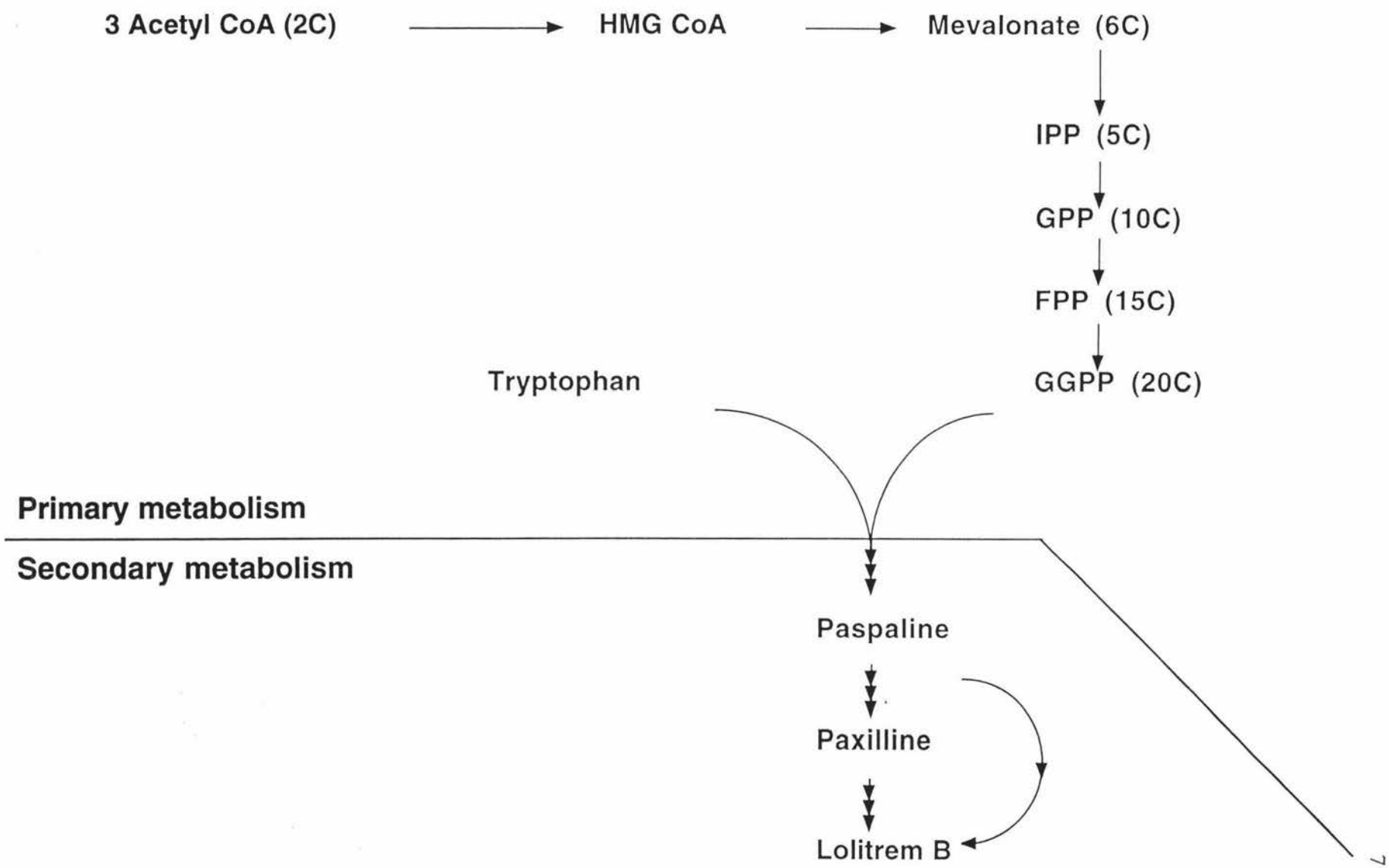
3-Hydroxy-3-methylglutaryl Coenzyme A reductase (Hmg) catalyses the synthesis of the primary metabolite, mevalonic acid. This is the first step in the isoprenoid biosynthetic pathway which leads to intermediates that serve key roles in protein synthesis, protein glycosylation, electron transport, cell cycle control and maintenance of membrane fluidity.

1.2.1 Structure and Function

Eukaryotic Hmg is an integral membrane protein of the endoplasmic reticulum (ER). It is bound to the ER via an amino-terminal membrane domain which contains a number (1-8) of hydrophobic regions or potential membrane spanning domains. A linker region

Figure 1 Biosynthesis of lolitrem B from acetyl CoA.

The isoprenoid biosynthetic pathway produces essential primary metabolites which act as precursors for synthesis of secondary metabolites, not essential for basic function of the organism. Hmg converts Hmg CoA to mevalonate, a precursor for the diverse range of secondary metabolites produced via this pathway. The pathway for production of lolitrem B from geranyl geranyl pyrophosphate has not been elucidated, except for the two metabolites, paspaline and paxilline.


3 Acetyl CoA (2C)
HMG CoA
Mevalonate (6C)
IPP (5C)
GPP (10C)
FPP (15C)
GGPP (20C)
Tryptophan
Primary metabolism
Secondary metabolism
Paspaline
Paxilline
Lolitrem B

separates this domain from the carboxy-terminal domain, which contains the hydrophilic catalytic site which projects into the cytoplasm (Basson *et al.* 1988; Liscum *et al.* 1985). Mammalian Hmg is encoded by a single gene, resulting in a single glycoprotein localised to the ER. In contrast, plants possess two or more isozymes of Hmg and these are encoded by small gene families. The activity of these enzymes is associated with plastid, mitochondrial and ER membranes.

The amino-terminal domain shows a low level of sequence similarity between phylogenetic kingdoms. This membrane situated domain is complex and appears to consist of 7 or 8 transmembrane domains in yeast, animals and fungi (Basson *et al.* 1988; Croxen *et al.* 1994; Liscum *et al.* 1985; Roitelman *et al.* 1992) in comparison with two domains in plants (Denbow *et al.* 1996; Enjuto *et al.* 1994). Despite the lack of sequence conservation between organisms, this domain mediates important regulatory controls such as the stability of the protein (Gil *et al.* 1985; Kumagai *et al.* 1995).

In comparison, the amino acid sequence of the catalytic domain is conserved across eukaryotes, archaebacteria and eubacteria, suggesting it is under strong selective pressure (Lum *et al.* 1996). This domain catalyses the conversion of HMG CoA to mevalonate, the first committed step of the isoprenoid biosynthetic pathway. The crystal structure of Hmg from *Pseudomonas mevalonii* has been determined, showing Hmg to be a tightly bound dimer that brings together at the subunit interface the conserved residues implicated in substrate binding and catalysis. Regions have been identified that might be involved in the binding of Hmg-CoA as well as the reduced and oxidised forms of NAD(P) (Lawrence *et al.* 1995). A number of specific amino acid residues within the conserved catalytic domain are known to be required for catalysis and production of mevalonate; these include histidine, aspartate and glutamate residues (Bischoff and Rodwell 1997; Darnay *et al.* 1992; Frimpong and Rodwell 1994). Mevalonate provides a five-carbon building block for the synthesis of a wide and diverse range of products produced via this pathway (Goldstein and Brown 1990). Many essential primary metabolites are produced as well as secondary metabolites which are not essential for the functioning of the organism.

1.2.2 Fungal Hmg

The complete *hmg* gene has been isolated and characterised from the fungi *Gibberella fujikuroi* (Woitek *et al.* 1997) and *Ustilago maydis* (Croxen *et al.* 1994). *G. fujikuroi* is a rice pathogen which produces high levels of gibberellins, a family of diterpenoid plant hormones, while *U. maydis* is a pathogenic basidiomycete which induces tumours in

maize. The gene has also been studied in the yeasts *Schizosaccharomyces pombe* (Lum *et al.* 1996) and *Saccharomyces cerevisiae* (Basson *et al.* 1986).

G. fujikuroi contains only one *hmg* gene, encoding a predicted protein of 976 amino acids. This is somewhat smaller than other fungal Hmg polypeptides and Woitek *et al.* (1997) determined this to be due to a shortened N-terminal domain. This truncated domain has been determined to still contain seven putative transmembrane regions by the Kyte and Doolittle (1982) method. The *G. fujikuroi hmg* gene consists of a 2928 bp open reading frame which is interrupted by a putative 47 bp long intron. *U. maydis* appears to contain two structural genes encoding Hmg. Only one of these has been isolated and sequenced, and this was found to encode a 1165 amino acid polypeptide (Senstag *et al.* 1990). The N-terminal domain in *U. maydis* has been shown to contain eight putative hydrophobic domains, as determined by the Kyte and Doolittle (1982) method. Although, of these eight, only seven of these have been recognised as possible transmembrane regions. The *U. maydis* sequence does not possess any intronic regions and the position and presence of introns does not appear to be conserved throughout eukaryotes. The catalytic domain of Hmg from another plant pathogenic fungus, *Sphaceloma manihoticola* has also been isolated and sequenced. It has been determined that this fungus also contains only one *hmg* gene (Woitek *et al.* 1997).

S. pombe hmg encodes a predicted protein of 1053 amino acids and is considered to be intron free, as no intron consensus sequences have been located within the open reading frame. It is also predicted to contain eight transmembrane domains. *S. cerevisiae* has been found to possess two Hmg isozymes, each encoded by genes with single continuous open reading frames, 1054 amino acids for HMG1 and 1045 amino acids for HMG2.

Each of the Hmgs contain the recognised catalytically important histidine, aspartate and glutamate amino acid residues (Bischoff and Rodwell 1997; Darnay *et al.* 1992; Frimpong and Rodwell 1994) as well as the residues implicated in substrate binding (Lawrence *et al.* 1995). Of importance with respect to the regulation of Hmg activity is a serine located 6 residues downstream from the catalytically important histidine. This residue is present in each of the fungal Hmgs cloned to date and is known to be phosphorylated by an AMP-activated protein kinase, resulting in a loss of catalytic activity. The yeast Hmg peptides lack this regulatory serine residue and are therefore not regulated by phosphorylation at this position.

1.2.3 Regulation

The compounds and enzymes of primary metabolism have essential roles in the survival of an organism. Intermediates of primary metabolic pathways often supply precursors of secondary metabolism, which give rise to compounds whose function is not essential. Therefore factors that influence primary metabolism will have some effect on secondary metabolism (Drew and Demain 1977). Primary metabolism is controlled by a number of mechanisms, including substrate induction, feedback repression and inhibition, catabolite repression and inhibition, and ATP regulation. In each case, regulation is a dynamic process dependent on the local concentration of effector molecules (Drew and Demain 1977).

1.2.3.1 Regulation of Hmg

In higher eukaryotes Hmg is considered to be the key regulatory enzyme of the isoprenoid biosynthetic pathway and as it is a typical housekeeping gene it is usually transcriptionally active. In mammals, Hmg activity can be regulated by feedback inhibition of steady state levels of Hmg mRNA, as well as at the molecular level to control the activity of the enzyme. This occurs by a coordinate modulation of synthesis of the protein when the pathway flux is low and degradation when the level of the pathway products are high. Regulation is therefore dependent on the concentration of an end-product. For regulation at both levels, two products of the terpenoid pathway are necessary: a sterol (low density lipoprotein) and a non-sterol product whose structural identity is yet to be determined. The amino-terminal anchor has been found to be necessary and sufficient to mediate the regulated degradation of Hmg in the endoplasmic reticulum (Hampton *et al.* 1996; Stermer *et al.* 1994).

Hmg activity is also regulated by processes independent of the mevalonate pathway such as reversible phosphorylation. Enzyme activity is down-modulated due to phosphorylation by an AMP-activated protein kinase, which has no effect on the stability of the enzyme but instead alters the kinetics of the enzyme (Hampton *et al.* 1996; Stermer *et al.* 1994). Phosphorylation occurs at a serine located 6 residues from the catalytic histidine, a spacing that is conserved in the *hmg* of most eukaryotes (Friesen and Rodwell 1997).

In plants, the number of genes encoding Hmg varies, depending on the species. Differential expression of the multiple *hmg* genes could play an important role in the regulation of enzyme activity. These *hmg* genes are also under a number of developmental and environmental regulatory mechanisms unique to plants, such as

light, phytohormones, pathogen attack, wounding, feedback mechanisms and endogenous protein factors. Much of the isoprenoid biosynthesis in plants is localised in specific compartments within a cell, which allows for the independent regulation of parallel pathways that produce different end products (Stermer *et al.* 1994).

Regulation of Hmg activity in turn determines mevalonate production and thus the amount of end-products produced.

1.3 AIMS AND BACKGROUND OF THIS STUDY

Although filamentous fungal endophytes confer many beneficial properties onto their host, perennial ryegrass, the production of animal toxins is agriculturally undesirable for livestock. Genetic strategies are therefore being undertaken to manipulate fungal endophytes and their hosts for agricultural benefit. Mevalonic acid is a precursor of both paxilline and ergovaline and synthesis of this compound by Hmg is considered an important regulatory step. Therefore, disruption of the paxilline pathway could increase the pool of mevalonic acid available for ergovaline synthesis, due to continued expression of *hmg*. It is also possible that in the symbiotic association, the endophyte draws on mevalonic acid produced by the plant, and so does not use its own enzyme at all. This may still lead to an increased pool of mevalonic acid within the endophyte even if Hmg is tightly regulated. It is therefore important to understand how this gene is regulated *in planta*. The aim of this project was to clone the *hmg* gene from the haploid *Neotyphodium* isolate Lp19 (LpTG-1) (Christensen *et al.* 1993) and to identify the upstream regulatory region of this gene. The availability of these sequences will allow *hmg* expression to be determined *in planta* using a *gus* reporter gene construct.

Chapter 2

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 1.

2.2 GROWTH OF CULTURES

Neotyphodium cultures were grown at 22°C on potato dextrose agar (Section 2.3.2). These plate cultures were sealed with parafilm and stored at 4°C.

For the growth of *Neotyphodium* cultures from which nucleic acid was to be extracted, 250 µl of a solution containing mycelia, homogenised in potato dextrose (PD) broth (Section 2.3.2), was used to inoculate 125 ml flasks containing 20 ml of PD broth. The cultures were incubated at 22°C with shaking at 200 rpm for 7-10 days. The mycelia were harvested by centrifugation at 2,500 g for 10 minutes at 4°C and the pellet washed twice in MilliQ water. The mycelia were frozen in liquid nitrogen and freeze dried.

Escherichia coli cultures were grown at 37°C in Luria broth (Section 2.3.1) with shaking at 250 rpm. *E. coli* cultures were maintained on LB agar plates sealed with parafilm and stored at 4°C. Antibiotics were supplemented as required at the following concentrations:

Antibiotic	Stock Concentration (mg/ml)	Required at (µg/ml)
Ampicillin	100	100
Tetracycline	10	10

Table 1 Strains, λ clones and plasmids.

Strain, λ clone or plasmid	Relevant characteristics	Source or reference
Fungal strains		
Neotyphodium sp. (=LpTG-2)		
Lp1	*E. typhina-N. lolii* hybrid from *Lolium perenne*	Christensen *et al.* 1993
Neotyphodium sp. (=LpTG-1)		
Lp19	*N. lolii* from *L. perenne*	Christensen *et al.* 1993
Epichloë typhina (=MP-I)		
E8	*E. typhina* from *L. perenne*	Byrd *et al.* 1990
Penicillium paxilli		
PN2013	Single spore purified strain of PN2012 (Pax^+ Brs^+ Hyg^S)	Itoh *et al.* 1994
Saccharomyces cerevisiae		
Bacterial strains		
Escherichia coli		
KW251	F^-, *sup*E44, *gal*T22, *met*B1, *hsd*R2, *mcr*A [*arg*A81: Tn10], *rec*D1014, Tet^R	Promega
XL-1-Blue	*sup*E44 *hsd*R17 *rec*A1 *end*A1 *gyr*A46 *thi rel*A1 lac-F' [*pro*AB^+ *lac*I^q *lacZ*Δ M15 Tn 10 (Tet^r)]	Bullock *et al.* 1987
λ Clones		
λGEM-12		Promega
λJD1	λGEM-12 clone containing *hmg* from Lp19	This study
λJD7	λGEM-12 clone containing *hmg* from Lp19	This study
λJD12	λGEM-12 clone containing *hmg* from Lp19	This study

Table 1 continued.

λJD14	λGEM-12 clone containing *hmg* from Lp19	This study
Plasmids		
pUC118	3.2 kb Amp^r	
pUC18	2.69 kb Amp^r	
pJD1	pUC118 containing a 5.2 kb *Sst*I fragment from λJD12	This study
pJD4	Identical to pJD11 except (-) orientation	This study
pJD6	Identical to pJD12 except (-) orientation	This study
pJD11	pUC118 containing a 0.6 kb *Sst*I fragment from λJD12	This study
pJD12	pUC118 containing a 0.8 kb *Sst*I fragment from λJD12	This study
pJD16	pUC118 containing a 1.8 kb *Sst*I fragment from λJD12	This study
pJD20	Identical to pJD1 but (-) orientation	This study
pJD21	Identical to pJD48 but (-) orientation	This study
pJD35	Identical to pJD16 but (-) orientation	This study
pJD48	pUC118 containing a 4.9 kb *Sst*I fragment from λJD12	This study
pJD124	pUC18 containing a 6.5 kb *Bam*HI fragment from λJD12	This study

Saccharomyces cerevisiae cultures were grown overnight at 30°C on YCM agar plates (Section 2.3.5), sealed with parafilm and stored at 4°C.

2.3 MEDIA

All media was prepared with MilliQ water and sterilised at 121°C for 15 minutes prior to use. Liquid media was cooled to room temperature before addition of antibiotic(s), supplements and inoculation. Solid media was cooled to approximately 50°C before antibiotic addition and pouring. Uninoculated plates were stored at 4°C.

2.3.1 Luria Broth (LB) Medium

LB media (Miller 1972) contained (g/l): sodium chloride, 5; tryptone, 10 and yeast extract, 5. The pH was adjusted to 7-7.5 with 10 M NaOH prior to autoclaving. For solid media, agar (Davis) or agarose 15 (BDH) was added to 15 g/l. Where required media was supplemented with $MgSO_4.7H_2O$ (10 mM), maltose (0.2%), isopropyl-β-D-galactoside (IPTG) (40 μg/ml) and 5-bromo-4-chloro-3-indoyl-β-D-galactoside in dimethylformamide (X-gal) (40 μg/ml).

2.3.2 Potato Dextrose Broth (PD) Medium

PD media contained 24 g of Potato Dextrose Broth (Difco) rehydrated in 1 litre of MilliQ water. The pH was adjusted to 6.5 before autoclaving. Solid media was prepared by adding 15 g/l agar (Davis).

2.3.3 SOC Medium

SOC medium (Dower *et al.* 1988) contained (g/l): Tryptone, 20; yeast extract, 5; NaCl, 0.6; KCl, 0.2; $MgCl_2$, 0.95; $MgSO_4.7H_2O$, 2.5; and glucose, 3.6.

2.3.4 TB Top Agarose

Top agarose contained (g/l): tryptone, 10; NaCl, 5 and agarose 15 (BDH), 8.

2.3.5 YCM Medium

YCM media contained (g/l): Yeast extract, 3; malt extract, 3; peptone, 5 and glucose, 10. For solid media agar (Davis) was added to 15 g/l.

2.4 SOLUTIONS

2.4.1 Electrophoresis

2.4.1.1 1 x TBE Electrophoresis Buffer contained 89 mM Tris-HCl, 2.5 mM Na_2EDTA and 89 mM boric acid, pH 8.3.

2.4.1.2 10 x TBE Sequencing Buffer contained Tris-HCl, 1.3 M; Na_2EDTA, 0.25 M and boric acid, 4.4 M. For running sequencing gels this buffer was diluted to 1 x with MilliQ water.

2.4.1.3 SDS Loading Buffer contained 1% (w/v) sodium dodecyl sulphate (SDS), 0.02% (w/v) bromophenol blue, 20% (w/v) sucrose and 5 mM Na_2EDTA (pH 8.0).

2.4.1.4 Acrylamide Mix contained (g/l): urea, 480; acrylamide 57 and bis-acrylamide, 3. This mix was made up to 800 ml and deionised with Amberlite MB-3 (Sigma), and filtered through a sintered glass funnel (porosity 1). To this 100 ml of 10 x sequencing TBE buffer (Section 2.4.1.2) was added and the volume made up to 1 litre with MilliQ water.

2.4.2 Southern Blotting and Hybridisation

2.4.2.1 Southern Blotting Solution 1 contained 0.25 M HCl.

2.4.2.2 Southern Blotting Solution 2 contained 0.5 M NaOH and 0.5 M NaCl.

2.4.2.3 Southern Blotting Solution 3 contained 0.5 M Tris-HCl (pH 7.4) and 2 M NaCl.

2.4.2.4 20 x SSC contained 3 M NaCl and 0.3 M sodium citrate.

2.4.2.5 2 x SSC was prepared by a 10-fold dilution of 20 x SSC.

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

2.4.2.7 10 x Denhardts (Southern 1975) contained (per litre): 50 ml 1 M HEPES (BRL), pH 7.0; 150 ml 20 x SSC (Section 2.4.2.3); 6 ml phenol extracted herring sperm DNA (3 mg/ml, Sigma); 5 ml of 20% (w/v) SDS; 2 g Ficoll (Sigma); 2 ml *Escherichia coli* transfer RNA (10 mg/ml Sigma); 2 g bovine serum albumin (Sigma), 2 g polyvinylpyrrolidone (PVP-10, Sigma).

2.4.3 DNA Isolation

2.4.3.1 SM Buffer contained NaCl, 100 mM; $MgSO_4.7H_2O$, 8 mM and 50 ml of 1 M Tris-HCl, pH 7.5.

2.4.3.2 Tris-Phenol (Amersham) was prepared by adding the antioxidant 8-hydroxyquinone.

2.4.3.3 TE (10/0.1) contained 10 mM Tris-HCl and 0.1 mM Na_2EDTA, pH 8.0.

2.4.3.4 DNase free RNase was prepared from RNaseH (Sigma) at 10 mg/ml in 10 mM Tris-HCl (pH 7.5) and 15 mM NaCl, heated to 100°C for 15 minutes. The solution was allowed to cool to room temperature then stored in aliquots at -20°C.

2.4.4 DIG Solutions

2.4.4.1 10% Blocking Reagent contained 10% (w/v) blocking reagent (Boehringer Mannheim) in DIG Buffer 1.

2.4.4.2 DIG Standard Hybridisation Solution contained 5 x SSC, 0.1% (w/v) sodium-lauroylsarcosine, 0.02% (w/v) SDS and 1% (v/v) blocking reagent.

2.4.4.3 DIG Buffer 1 contained 150 mM NaCl and 100 mM maleic acid (pH 7.5).

2.4.4.4 DIG Buffer 2 contained 1% (v/v) 10% blocking reagent in DIG Buffer 1.

2.4.4.5 Anti-digoxigenin-AP Fab Fragments were diluted 1:10,000 in DIG Buffer 2.

2.4.4.6 DIG Buffer 3 contained 100 mM Tris-HCl (pH 9.5) and 100 mM NaCl.

2.5 DNA ISOLATION

2.5.1 Total DNA Isolation from Fungal Cultures by the Yoder (1988) Method.

In a pre-cooled mortar and pestle approximately 300 mg of freeze dried mycelium was ground to a fine powder under liquid nitrogen then resuspended in 6 ml of phenol (Section 2.4.3.2) and 6 ml of extraction buffer (100 mM LiCl, 10 mM Na_2EDTA, 10 mM Tris-HCl [pH 7.4], 0.5% [w/v] SDS). The solution was centrifuged at 20,000 g for 15 minutes at 4°C. The aqueous phase was phenol/chloroform extracted (Section 2.6) and the resulting aqueous phase centrifuged at 24,000 g for 20 minutes. DNA was then ethanol precipitated (Section 2.7), harvested by spooling, washed with 70% ethanol, dried and resuspended in 1 ml sterile MilliQ water or TE (Section 2.4.3.3).

2.5.2 Total DNA Isolation from Fungal Cultures by the Byrd *et al.* (1990) Method.

In a pre-cooled mortar and pestle approximately 0.5 g of freeze dried mycelia was ground to a fine powder under liquid nitrogen and resuspended in 10 ml of extraction buffer (150 mM Na_2EDTA, 50 mM Tris-HCl [pH 8.0], 1% [w/v] sodium-lauroylsarcosine and 2 mg/ml proteinase K). The solution was centrifuged at 2,000 g for 10 minutes (all centrifugations were carried out at 4°C). The supernatant was incubated at 37°C for 20 minutes then phenol/chloroform extracted (Section 2.6). The aqueous phase was centrifuged at 24,000 g for 20 minutes, the DNA precipitated with ice-cold isopropanol (Section 2.7) and resuspended in 1 ml of MilliQ water.

2.5.3 Purification of DNA from Phage λ

Phage plated to confluent lysis on LB agarose (Section 2.3.1) were overlayed with 5 ml of SM buffer (Section 2.4.3.1) and left overnight at 4°C. The resulting lysate was collected and centrifuged at 4,200 g for 10 minutes (all centrifugations were performed

at 4°C). DNase and RNase were added to the supernatant at a concentration of 1 µg/ml, and incubated at 37°C for 30 minutes. After this step, 5 ml of 20% (w/v) PEG 8000 in 2 M NaCl was added and the mixture incubated on ice for 1 hour. The phage were pelleted by centrifugation at 4,200 g for 30 minutes and the pellet resuspended in 0.5 ml of SM buffer containing 5 µl of 10% (w/v) SDS and 10 µl of 250 mM Na_2EDTA (pH 8.0). The mixture was incubated at 68°C for 15 minutes, then vortexed. An equal volume of phenol (Section 2.4.3.2) was added, the solution vortexed for 10 seconds, left for 5 minutes then vortexed for another 10 seconds. The aqueous phase was removed from the phenolic phase and phenol/chloroform extracted (Section 2.6). The DNA was resuspended in 50 µl MilliQ water containing 4 µg DNase free RNase (Section 2.4.3.4) and quantitated (Section 2.8.2).

2.5.4 Plasmid Isolation by the Alkaline Lysis Method

The cells from a 3 ml overnight *E. coli* culture (Section 2.2) were pelleted by a 3 minute centrifugation in a microcentrifuge. The supernatant was drained and the pellet resuspended in 100 µl of a solution containing 50 mM glucose, 25 mM Tris-HCl (pH 8.0) and 10 mM Na_2EDTA (pH 8.0). To this, 200 µl of a solution containing 0.2 M sodium hydroxide and 1% (w/v) SDS was added and mixed by rapid inversion several times. The mixture was stored on ice and 150 µl of potassium acetate solution (29.44 g potassium acetate and 11.5 ml glacial acetic acid per 100 ml) was added, mixed briefly by vortexing and incubated on ice for a further 5 minutes. The resulting precipitate was pelleted by centrifugation in a microcentrifuge for 5 minutes then the supernatant transferred to a fresh tube and phenol/chloroform extracted (Section 2.6). The DNA was ethanol precipitated (Section 2.7), resuspended in 50 µl of MilliQ water and quantitated (Section 2.8.1, 2.8.2). This method is based on that of Sambrook *et al.* (1989).

2.5.5 Plasmid Isolation by the Rapid Boil Method

Initially plasmids were isolated by the method of Holmes and Quigley (1981), however once the alkaline lysis method (Section 2.5.4) was found to give satisfactory yields, this was used exclusively.

The cells from a 3 ml overnight *E. coli* culture (Section 2.2) were pelleted by centrifugation for 3 minutes in a microcentrifuge. The supernatant was drained and the pellet resuspended in 350 µl of STET buffer (8% [w/v] sucrose, 5% [v/v] Triton X-100, 50 mM Na_2EDTA [pH 8.0] and 50 mM Tris-HCl [pH 8.0]). Lysozyme (25 µl of a 10 mg/ml solution) was added and the tube placed in a boiling water bath for 40 seconds.

The solution was centrifuged for 10 minutes in a microcentrifuge and the resulting gelatinous pellet removed with a sterile toothpick. The DNA was precipitated by the addition of an equal volume of isopropanol, mixed by inversion and incubated at -20°C for 10-20 minutes. Plasmid DNA was pelleted by centrifugation for 5 minutes, washed in 70% ethanol and dried at 37°C. The DNA was resuspended in 50 µl MilliQ water containing 2 µg of DNase free RNase (Section 2.4.3.4) and quantitated (Section 2.8.2).

2.6 EXTRACTION OF DNA BY PHENOL/CHLOROFORM

DNA samples isolated by methods described in Section 2.5 were extracted once with two volumes of Tris-equilibrated phenol (Section 2.4.3.2) and centrifuged at 12,000 g. The aqueous phase was extracted with an equal volume of Tris-equilibrated phenol and chloroform and re-centrifuged. The resulting aqueous phase was then re-extracted with two volumes of chloroform and the DNA precipitated with either ethanol or isopropanol as described in Section 2.7. This method is based on that of Sambrook *et al.* (1989).

2.7 PRECIPITATION OF DNA WITH ETHANOL OR ISOPROPANOL

One tenth volume of 3 M sodium acetate and either 2.5 volumes of 95% ethanol or 0.6 volumes of isopropanol, were added to the tube containing DNA to be precipitated. The solution was mixed by inversion and allowed to stand on ice for at least 30 minutes after which the DNA was pelleted by centrifugation at 12,000 g for 5 minutes. The pellet was washed once with 70% ethanol and dried before resuspension in MilliQ water or TE (Section 2.4.3.3) containing 2 µg of DNase free RNase (Section 2.4.3.4). This method is based on that of Sambrook *et al.* (1989).

2.8 DNA QUANTIFICATION

DNA was quantified by two methods, fluorometrically (Section 2.8.1) or by intensity of ethidium bromide fluorescence (Section 2.8.2).

2.8.1 Fluorometric Quantitation of DNA

DNA was quantitated on a Hoefer TKO 100 Fluorometer. This requires only 2 µl of a DNA sample for quantification. DNA was quantitated in a dye solution containing 1 x TNE (10 mM Tris-HCl, 1 mM Na_2EDTA and 100 mM NaCl, pH 7.4) and 0.1 µg/ml Hoechst 33258. The scale of the fluorometer was zeroed with dye solution until a steady state was obtained then set to 100 using 2 µl of 100 µg/ml calf thymus DNA added to 2 ml of the dye solution. Once the scale was set, 2 µl of sample DNA was added to 2 ml of dye solution giving the concentration of the sample DNA solution in ng/µl.

2.8.2 Minigel Method for Determination of DNA Concentration

A sample from the DNA solution of interest was examined by electrophoresis through an agarose gel (Section 2.10) with a series of standard DNA solutions of known concentration. After the bromophenol blue dye front had migrated at least 2/3 of the way into the gel the DNA was stained with ethidium bromide and photographed (Section 2.10). The sample DNA concentration was determined by comparison of its intensity of fluorescence to that of the known DNA standards.

2.9 RESTRICTION ENDONUCLEASE DIGESTION OF DNA

Restriction endonuclease digests were carried out with the manufacturers buffer supplied with the enzyme. DNA to be digested was quantitated (Section 2.8) and following the manufacturers guidelines an excess of enzyme was added to digest the DNA.

Digestion of plasmid and phage λ DNA was performed at 37°C for 1-2 hours and stored on ice, or at -20°C while an aliquot was checked on an agarose gel (Section 2.10) to ensure that the digestion had gone to completion. In the event that digestion was incomplete, fresh enzyme was added and the mixture incubated a further hour. If the DNA digest had still not reached completion, the stock was further purified by phenol/chloroform extraction (Section 2.6) and the digest repeated. Once digestion was complete, the reaction was stopped by the addition of 1/5 volume of SDS loading buffer (Section 2.4.1.3). Digestion of fungal DNA was the same as that for plasmid and phage DNA except digestion times were increased to a minimum of 3 hours and a maximum of overnight incubation.

2.10 AGAROSE GEL ELECTROPHORESIS OF DNA

Horizontal agarose gel electrophoresis was performed in a Mini-gel apparatus for 1-2 hours or in a Bio-Rad DNA Sub Cell (150 x 200 mm gel bed) overnight. Agarose concentrations varied from 0.4% to 1% (w/v) in 1 x TBE (Section 2.4.1.1), except for NuSieve agarose gels that were used at concentrations of 2% and 3% (w/v) in 1 x TBE. DNA samples (with addition of 1/5 volume of SDS loading buffer, Section 2.4.1.3) were loaded on the gel and the fragments separated by electrophoresis at 80 V to 120 V (Mini-gels) or 30 V (Bio-Rad Sub Cell). After electrophoresis, gels were stained in ethidium bromide (5 μg/μl) for 5-15 minutes, destained 5-15 minutes in water, visualised on a UV transilluminator and photographed on Polaroid 667 film or by the gel documentation system (Alpha Innotech).

DNA fragment sizes were determined, after gel electrophoresis, by measuring the distance a fragment had migrated from the well. The molecular weight was then calculated by interpolation from a plot of the distance migrated in the same gel by suitable size markers, against the logarithm of the molecular weight (kb or bp) of the size markers, or by the use of the fragment size determination program of the gel documentation system (Alpha Innotech).

2.11 RECOVERY OF DNA FROM AGAROSE GELS

DNA was recovered from NuSieve agarose gels (2% [w/v] in 1 x TBE electrophoresis buffer, Section 2.4.1.1) by the Thuring *et al.* (1975) phenol-freeze extraction method.

After gel electrophoresis the DNA fragments of interest were visualised under long-wave UV, excised from the gel with the minimum amount of excess agarose and placed in microcentrifuge tubes. The agarose was melted at 65°C, covered with Tris-equilibrated phenol (Section 2.4.3.2), mixed by vortexing and frozen at -20°C for at least 2 hours. The mixture was microcentrifuged for 10 minutes and the aqueous phase recovered, phenol/chloroform extracted (Section 2.6) and precipitated with ethanol (Section 2.7).

2.12 AMPLIFICATION OF DNA BY THE POLYMERASE CHAIN REACTION (PCR)

Each PCR reaction in a 0.2 ml strip tube contained: 1 U *Taq* DNA polymerase (Boehringer Mannheim), 1.25 mM of each dNTP, 10 pmol of "forward" primer and 10 pmol "reverse" primer, appropriate DNA at an optimal concentration, 1 x *Taq* polymerase buffer (either with Mg^{2+} [1.5 mM, Boehringer Mannheim] or Mg^{2+} free [Promega] plus $MgCl_2$) and MilliQ water to 25 µl. The reactions were placed in the Corbett thermal cycler, a hot top which does not require oil. The sequences of the primers used were dependant on the sequence being amplified and were 17-23 bases in length (Table 2). A negative control containing water was included in each PCR run and this was prepared as the last reaction in each set. The reaction vessels were placed in the thermal cycler preheated to 94°C and after an initial 2 minute melt of 94°C were subjected to 30 cycles of 94°C for 45 seconds, with varied temperatures for 45 seconds to allow annealing and 72°C for 90 seconds. After these cycles were complete the reactions were incubated at 72°C for 5 minutes, checked on an agarose mini-gel (Section 2.10) and stored at -20°C. This method is based on those outlined by (Innis and Gelfand 1990). The annealing temperature used was dependent on the melting temperature (Tm) of the primers. The Tm of each primer was determined using the calculation Tm = 2 x [A+T] + 4 x [G+C] (Itakura *et al.* 1984)

2.13 DNA SEQUENCING

2.13.1 Amplicycle Sequencing

The protocol used for sequencing was that from Perkin Elmer's AmpliCycle sequencing kit. This method is based on the dideoxy-mediated chain termination method of Sanger *et al.* (1977).

For each set of reactions four 0.2 ml strip tubes were labelled A, G, C and T and 2 µl of the appropriate ddNTP mix added to each tube and left on ice. A cocktail was prepared as follows: template DNA (0.5-1 µg), 1 pmol of the appropriate primer, 1 µl of [α-^{33}P]dATP, 4 µl of 10 x cycling mix and sterile MilliQ water to 30 µl. To each of the four ddNTP tubes 6 µl of cocktail was added and gently mixed. The tubes were placed in the Perkin Elmer thermal cycler pre-heated to 95°C and after an initial 1 minute melt

Table 2 PCR and Sequencing Primers

Primer	Size	Tm°C	Sequence (5'-3')	Source
hmg 1	17 mer	44	ATGGGIATGAAIATGAT	this study
hmg 3	23 mer	66	ACITTTTGIGCIGGITCTTGICC	this study
hmg 4	21 mer	66	GTTGAGCATGCGCTCAACGTC	this study
hmg 5	21 mer	66	TTGAAGCCACCGATCGAACCG	this study
hmg 7	21 mer	64	TCGCTCGTCTACAGCACATGA	this study
hmg 8	21 mer	64	TTGATCGCTTTGCAGCCACGA	this study
hmg 9	21 mer	60	TGACCATTGCTCAAGACAAGT	this study
hmg 10	20 mer	56	GGTACTCTAGTTCGTATTTG	this study
hmg 11	20 mer	56	CAAATACGAACTAGAGTACC	this study
hmg 12	21 mer	60	ACTTGTCTTGAGCAATGGTCA	this study
hmg 13	20 mer	62	CTGCGCCCAACGATAGAATC	this study
hmg 14	20 mer	62	GATTCTATCGTTGGGCGCAG	this study
hmg 18	20 mer	60	CAAGCACCATGATGACGATG	this study
hmg 19	20 mer	60	CATCGTCATCATGGTGCTTG	this study
hmg 20	20 mer	62	GTTACCATCGCCGTTCTAGC	this study
hmg 21	20 mer	62	GAGATGACCAGTCAGCCTTG	this study
hmg 22	20 mer	66	CTCTTGCTTTACACGGGACG	this study
hmg 23	21 mer	62	GATTATCGTCATGTGGCTGAC	this study
hmg 26	21 mer	66	CGGAGTAGCAGCGTTCCAAAG	this study
hmg 27	21 mer	66	CTTGCGGTCAATTCCGTGGTC	this study
hmg 28	21 mer	66	CACCATCTTGGAGCCTCAGAG	this study
hmg 29	21 mer	64	GACGTGGTCAAATCCGTGTTG	this study
hmg 30	20 mer	62	GTTGGATGTGATCCTCGCAC	this study
NS7	21 mer	68	TGCGGGTGCGCTATCGAGATG	White *et al.* 1990
NS8	21 mer	62	GCAAATCACAGTCACCAGCGG	White *et al.* 1990

Table 2 continued.

H-2	21 mer	62	ATAACGAGGATCCACGCAAGA	Hotter 1997
H-3	21 mer	62	GGCTCTTCTGACATTCGTCTT	Hotter 1997
H-4	21 mer	62	AGGTCAACTCTCGAAGTTCCA	Hotter 1997
pUC/M13 Forward	22 mer	70	GCCAGGGTTTTCCCAGTCACGA	Perkin Elmer
Poly(T)-G anchored	16 mer	38	AAGCTTTTTTTTTTG	Collett 1995
Poly(T)-A anchored	16 mer	36	AAGCTTTTTTTTTTA	Collett 1995
Poly(T)-C anchored	16 mer	38	AAGCTTTTTTTTTTC	Collett 1995

at 95°C, subjected to 25 cycles of 95°C for 30 seconds, standard 50°C (45°C for primers hmg 1 and hmg 3) for 30 seconds and 72°C for 60 seconds. After the cycles were complete 4 µl of stop solution was added to the reactions which were then stored at -20°C.

2.13.2 PAGE Electrophoresis of Sequencing Reactions

Sequencing reaction extension products (Section 2.13.1) were separated on the basis of length by polyacrylamide gel electrophoresis (PAGE). The sequencing plates were assembled to the manufacturers specifications (BRL) and sequencing gels were poured with 60 ml of acrylamide mix (Section 2.4.1.4) containing 36 µl of TEMED and 360 µl of 10% (w/v) ammonium persulphate. Once the gels were polymerised, they were pre-run for at least 15 minutes with constant power (65 W) in 1 x TBE sequencing buffer (Section 2.4.1.2). The sequencing reactions were denatured at 95°C for 2 minutes and 3-5 µl loaded onto the sequencing gel which was run for varying lengths of time. A long run went for 3 Bromophenol Blue dye fronts, taking approximately 6 hours, a medium run was for 2 dye fronts (approximately 4 hours) and short runs were for 1 dye front (approximately 2 hours). The gel was then disassembled, dried for 1 hour under vacuum at 80°C, exposed to Fuji X-ray film at room temperature (overnight for short and medium runs and for 2-3 days for long runs), the film developed (Section 2.13.4) and the sequence analysed.

2.13.3 Automatic Sequencing

Automatic sequencing of DNA samples was carried out using the ABI Prism 377 DNA Sequencer. Double stranded DNA was used at a concentration of 200 ng/µl and PCR products at approximately 25 ng/µl, both in a volume of 12 µl. Primers were used at a concentration of 0.8 pmol/µl in a volume of 5 µl. The DNA samples and primers were dried and sent to the Waikato DNA Sequencing Facility.

2.13.4 Developing

Once autoradiographs had been exposed for an appropriate time, they were developed in Kodak developer D-19 developing solution for 5 minutes, rinsed with water and fixed in Kodak Rapid Fixer solution A for 5 minutes. The autoradiographs were rinsed in water and dried before analysis.

2.14 SOUTHERN BLOTTING AND HYBRIDISATION

2.14.1 Southern Blotting

DNA to be transferred to the Nylon membrane (Amersham Hybond) was separated by overnight electrophoresis, stained, visualised and photographed as described in Section 2.10 and the gel dimension measured after removal of the gel portion above the wells.

The gel was then gently agitated for 15 minutes in Southern Blotting Solution 1 (Section 2.4.2.1) before immersion in Southern Blotting Solution 2 (Section 2.4.2.2) for 30 minutes with shaking. The gel was removed and placed in Southern Blotting Solution 3 (Section 2.4.2.3) for 1 hour with gentle agitation. Finally the gel was washed for 2 minutes in 2 x SSC (Section 2.4.2.5).

While the gel was being treated the blotting apparatus was assembled. Two sheets of Whatman 3MM chromatography paper (pre-soaked in 20 x SSC [Section 2.4.2.4]) were placed on a plastic trough with the wells at each end filled with 20 x SSC so the paper ends went into the wells and contacted the solution. A sheet of Gladwrap was placed over the trough and pressed flat. A grid 2 mm less than the gel size was marked out on the Gladwrap and removed. The treated gel was then placed over the grid so that the edges of the gel overlapped the edges of the Gladwrap. A piece of nylon membrane (Hybond-N), cut 2 mm larger than the gel size was pre-soaked in 2 x SSC (Section 2.4.2.5) and placed over the gel ensuring that no air bubbles were present. Two pieces of Whatman 3MM paper, cut 2 mm less than the gel size (pre-soaked in 2 x SSC) were placed over the membrane followed by two pieces of dry Whatman 3MM paper. A stack of paper towels were placed upon the chromatography paper, followed sufficient weight to keep the stack flat. After overnight transfer, the apparatus was disassembled and the membrane washed for 5 minutes in 2 x SSC, dried between blotting paper and baked under vacuum at 80°C for 2 hours. This method is based on the method of Southern (1975).

2.14.2 Preparation of [α-^{32}P]-Labelled Probe DNA with the High Prime Labelling Kit

DNA to be labelled (25 ng) in a volume of 11 µl was denatured by heating in a boiling water bath for 10 minutes and immediately placed on ice. To this 4 µl of High Prime solution (Boehringer Mannheim) and 5 µl (50 µCi) [α-^{32}P]dCTP was added. The reaction was mixed, spun briefly in a microcentrifuge and incubated at 37°C for 1 hour.

The reaction was stopped by the addition of 2 µl of 0.2 M Na_2EDTA (pH 8) and 28 µl TES buffer (Section 2.4.2.6), the tube was then mixed by vortexing and spun briefly. Unincorporated nucleotides were separated from labelled DNA on a minispin Sephadex G-50 column (Amersham) equilibrated with TES (10/1/100) (Section 2.4.2.6). The column was vortexed, the end broken off and the column inserted into a 1.5 ml microcentrifuge tube. These were centrifuged for 1 minute, radiolabelled DNA was added and the column centrifuged for a further 2 minutes. The probe was collected and transferred to a fresh tube to be stored at -20°C.

2.14.3 Hybridisation of [α-^{32}P]-Labelled Probe DNA to Southern Blots

The Southern blot (Section 2.14.1) to be probed was prehybridised in a sealed glass tube with 10 x Denhardts solution (Section 2.4.2.7) for at least 2 hours at 65°C. After pre-hybridisation, boiled [α-^{32}P]dCTP-labelled DNA was added (Section 2.14.2) to the tube containing approximately 5 ml of the hybridisation buffer.

After overnight hybridisation at 65°C, the filter was removed and washed three times, for 20 minutes each, in 2 x SSC (Section 2.4.2.5) at room temperature. The washed filter was wrapped in Gladwrap and exposed, in the presence on a Cronex intensifying screen, to a sheet of Fuji X-ray film. After exposure for a suitable period of time at -70°C, the film was developed (Section 2.13.4).

2.14.4 Preparation of DIG-Labelled Probe DNA

Digoxigenin-11-dUTP (DIG) labelled DNA was prepared by the random priming method using the PCR DIG DNA labelling Kit. The method is based on that in the Boehringer Mannheim DIG User's Guide for Filter Hybridisation (1995). DNA to be labelled (25 ng in a 10 µl volume) was denatured in a boiling water bath for 5 minutes then placed immediately on ice. To this, the following reagents were added: 1 x hexonucleotide mix, 1 x dNTP labelling mix, 2 U Klenow enzyme and sterile MilliQ water to 20 µl. The reagents were gently mixed and incubated at 37°C for 3 hours and 1.6 µl of 250 mM Na_2EDTA (pH 8) was added to terminate the reaction. The DNA was precipitated with 0.1 volumes of 4 M LiCl and 2.5 volumes of 95% ethanol at -20°C overnight. The probe was microcentrifuged for 15 minutes at 4°C, the ethanol decanted off and the pellet washed in 70% ethanol and dried. The probe was resuspended in 50 µl MilliQ water.

To estimate the probe yield stepwise dilutions, 1 ng/µl-0.01 pg/µl, of the DIG-labelled DNA were prepared in sterile MilliQ water. Each dilution (1 µl) was spotted onto a positively charged nylon membrane (Boehringer Mannheim) along with control DNA of the same concentrations. The membrane was baked in a vacuum oven at 80°C for 2 hours and the probe detected using chemiluminescence (Section 2.14.6).

2.14.5 Hybridisation of DIG-Labelled DNA Probe to Southern Blot

The Southern blot (Section 2.14.1) to be probed was pre-hybridised in a sealed glass tube with DIG Standard Hybridisation Buffer (Section 2.4.4.2), for 1 hour at 65°C. To this, 3.5-7 µl of DIG-labelled DNA probe, boiled in 2 ml of Standard Hybridisation Buffer for 10 minutes was added. After overnight hybridisation at 65°C the filter was removed and washed twice in 2 x SSC (Section 2.4.2.5) containing 0.1% (w/v) SDS for 5 minutes at room temperature, with gentle agitation. All subsequent washes were carried out at room temperature. The membrane was then washed twice for 15 minutes in 0.1 x SSC containing 0.1% (w/v) SDS prior to detection (Section 2.14.6).

2.14.6 Detection of Hybridised DNA Using Chemiluminescence

The hybridised filters (Section 2.14.5) were equilibrated in DIG Buffer 1 (Section 2.4.4.3), for 1 minute, gently agitated in DIG Buffer 2 (Section 2.4.4.4) for 1 hour before incubation in Anti-digoxigenin-AP Fab fragments (Section 2.4.4.5) for 30 minutes. The filters were washed twice for 15 minutes in DIG Buffer 1 containing 0.3% (v/v) Tween 20, then equilibrated in DIG Buffer 3 (Section 2.4.4.6) for 5 minutes. The filter was placed on an acetate sheet and 200-500 µl of CSPD (Boehringer Mannheim) diluted 1:100 in DIG Buffer 3 (Section 2.4.4.6) was pipetted onto it. Another acetate sheet was placed over the membrane and the bubbles smoothed out. The filters were incubated at room temperature between the plastic sheets for 5 minutes, then removed and placed between fresh plastic sheets and incubated at 37°C for 15 minutes. The filters were exposed to X-ray film for 15-30 minutes at 37°C before developing (Section 2.13.4).

2.14.7 Stripping Hybridised DNA off Southern Blots

The filter to be stripped was immersed in a solution of boiling 0.1% (w/v) SDS and gently shaken overnight while the solution cooled to room temperature. This was performed twice before the filter was checked by autoradiography (Section 2.13.4 for [α-^{32}P]dCTP labelled DNA, Section 2.14.3) or chemiluminescence (Section 2.14.6 for

DIG labelled DNA, Section 2.15.7) to ensure that stripping of the filter had occurred. If stripping was incomplete the process was repeated.

2.15 LIBRARY CONSTRUCTION

2.15.1 Digestion of Genomic DNA

2.15.1.1 Trial Partial Digestion of Genomic DNA

Total genomic DNA was isolated from freeze-dried mycelia of Lp19 by the Byrd *et al.* (1990) method (Section 2.5.2). Genomic DNA was partially digested (Section 2.9) with *Mbo*I to generate the maximum yield of fragments in the size range of 9-23 kb. A reaction mix (300 µl) containing 10 µg genomic DNA and 30 µl REact Buffer 2 (BRL) was prepared with 60 µl aliquoted into tube 1 and 30 µl into tubes 2-8. To tube 1, 10 U of *Mbo*I was added and mixed well before transferring 30 µl to tube 2. Serial dilution of *Mbo*I was carried through to the eighth tube, leaving the remaining tube as a control. All of the reaction tubes were incubated at 37°C for 1 hour before addition of 7 µl of SDS loading buffer (Section 2.4.1.3) to terminate the reaction. The samples were checked on a 0.4% (w/v) agarose overnight gel (Section 2.10), supported with approximately 90 ml of 1% (w/v) agarose.

2.15.1.2 Large-Scale Preparation of Partially Digested Genomic DNA

Using the optimised conditions determined in Section 2.15.1 a large scale reaction was carried out with 20 µg genomic DNA. Once the reaction had reached completion the mix was inactivated by incubation at 65°C for 15 minutes, analysed on a 0.4% (w/v) agarose gel (Section 2.10) and stored at 4°C. DNase free RNase (Section 2.4.3.4) was added to a final concentration of 10 µg/ml and incubated at 37°C for 30 minutes before extraction with phenol/chloroform (Section 2.6) and the ethanol precipitated DNA resuspended in 50 µl MilliQ water.

2.15.2 Partial Fill-In of Genomic DNA

The cloning strategy used with the λGEM-12 *Xho*I Half-Site Arms (Promega) relies on the high specificity with which partially filled-in Arms can be combined with partially filled-in genomic DNA digested with *Mbo*I. The digested DNA (7 µg) was combined with: 5 µl REact buffer 4 (BRL), 1 mM dATP, 1 mM dGTP, 8 U Klenow enzyme

(Boehringer Mannheim) and MilliQ water to 50 µl. The reaction was gently mixed and incubated at 37°C for 30 minutes. The DNA was phenol/chloroform extracted (Section 2.6), the precipitated DNA resuspended in 15 µl MilliQ water and quantitated (Section 2.8.2).

2.15.3 Optimising Ligation Conditions for Genomic Inserts

To determine the optimal conditions, small-scale ligations were carried out. Ligation reactions contained: λ arms (1 µg) and insert DNA at varying ratios, 1 x ligation buffer (New England Biolabs, NEB), 1 µl T4-DNA ligase (NEB) and MilliQ water to a final volume of 10 µl. With each set of ligation reactions Positive Control Insert DNA (Promega) was ligated and a negative control reaction (containing no insert DNA) was used to determine background levels of religated arms. The ligation reactions were gently mixed and incubated at 4°C overnight before packaging (Section 2.15.5) and titration (Section 2.15.6) of the phage. The titres from the various ratios were compared to determine the optimal concentrations of vector arms and insert DNA. Large scale packaging based on the optimal ratios was carried out.

2.15.4 Packaging of Ligated DNA

Small scale packaging reactions requiring 25 µl of Packagene extract (Promega) and 2.5 µl of ligation mix were carried out to check ligations before large scale packaging reactions, containing 50 µl of Packagene extract and 5 µl of ligation mix, were performed. The Packagene Extract (Promega) was thawed on ice before addition of ligation mix (Section 2.15.3) and incubated at 22°C for 3 hours. SM Buffer (445 µl, Section 2.4.3.1) and chloroform (25 µl) were added and the packaged phage placed at 4°C for short-term storage or at -20°C in 7% DMSO for long-term storage. These volumes were halved for small scale packaging reactions.

2.15.5 The Titre of the Packaged Phage

Serial dilutions of the packaging reactions were carried out in SM buffer (Section 2.4.3.1) and 100 µl of the diluted phage mixed with 100 µl of an overnight culture of KW251 grown in LB medium supplemented with: $MgSO_4.7H_2O$, maltose (Section 2.3.1) and tetracycline (Section 2.2). The phage/KW251 mixture was incubated at 37°C for 30 minutes, then added to 3 ml of top agarose (Section 2.3.4) at 50°C, and poured onto pre-warmed LB agar plates (Section 2.3.1). The top agarose was allowed to set, then the plates were incubated at 37°C overnight. The plaques were counted and the library titre (plaque forming units (PFU) per ml) determined.

2.15.6 Amplification of a Genomic Library

Plates from Section 2.15.5 were overlayed with 5 ml of SM buffer (Section 2.4.3.1) and left at 4°C overnight. The phage lysate was harvested and centrifuged 4,200 g for 10 minutes (all centrifugations were carried out at 4°C). The supernatant was recovered and 400 µl of chloroform added before incubation at room temperature for 15 minutes with occasional shaking. The phage were re-centrifuged at 4,000 g for 15 minutes, the supernatant recovered and plated out for confluent lysis. The phage from the confluent plates were retitred and stored in 7% (v/v) DMSO at -70°C.

2.16 LIBRARY SCREENING

Recombinant phage were screened by plating (Section 2.15.5) at approximately $2x10^4$ pfu/ml. The phage were incubated at 37°C for approximately 8 hours (until small plaques were visible) then left at 4°C overnight. Filter lifts of the resulting plates (Section 2.16.1) were then hybridised to an appropriate probe (Section 2.16.2). Plugs containing positive plaques were stored overnight in SM buffer (Section 2.4.3.1) containing a drop of chloroform. The phage from these plaques were purified twice more by screening as outlined above except 30-300 phage were plated. DNA was extracted from the positive plaques (Section 2.5.3) and analysed by restriction enzyme digestion (Section 2.9). This method is based on that of Sambrook *et al.* (1989).

2.16.1 Filter Lifts

Filters (Hybond-N+) marked asymmetrically, were placed on the KW251/phage lawn and these marks transferred to the plate. After 60 seconds the filter was removed and placed DNA side up on 3 layers of 3MM paper soaked in Southern Blotting Solution 2 (Section 2.4.2.2) for 2 minutes, Southern Blotting Solution 3 (Section 2.4.2.3) for 5 minutes and finally in 2 x SSC (Section 2.4.2.5) for 2 minutes. The filters were then air dried and baked in a vacuum oven at 80°C for 2 hours. When duplicate lifts were performed, the second filter was left on the plate for 90 seconds.

2.16.2 Plaque DNA Hybridisations

Labelling of the DNA to be used as a probe, hybridisation of labelled DNA to filters, detection and autoradiography was performed as detailed in sections 2.14.2-2.14.6, except hybridisations were carried out in plastic pots approximately 10 cm in diameter.

Positive plaques were identified by positive signals on the autoradiographs (Section 2.16.3).

2.16.3 Isolating Positive λ Clones

Positive plaques were identified as signals on the autoradiographs in identical positions on duplicate filters. The positions of the positive plaques on the KW251 lawn were established by alignment of the asymmetrical markings on the plates and filters with the autoradiograph. Plaques corresponding to signals were picked with the pipette tip of a 1 ml Gilson autopipettor from which the end (5-10 mm) had been removed and were stored at 4°C in 500 µl SM buffer (Section 2.4.3.1) containing a drop of chloroform.

2.17 SUBCLONING

2.17.1 DNA Ligations

2.17.1.1 CAP-Treatment of Vector DNA

Approximately 5 µg of vector DNA was digested to completion with the appropriate restriction endonuclease (Section 2.9). Calf alkaline phosphatase (0.5 U, CAP, Boehringer Mannheim) was added and the mixture incubated for 30 minutes at 37°C before addition of Na_2EDTA (5 mM) and SDS (0.5% [w/v]). Proteinase K was added (to a final concentration of 50 µg/ml) and the mixture incubated at 56°C for 30 minutes, phenol/chloroform extracted (Section 2.6) and the precipitated DNA resuspended in MilliQ water. This method is based on that of Sambrook *et al.* (1989).

2.17.1.2 Ligation

Ligation mixtures containing 1 x Ligation buffer (NEB), 20 ng of insert DNA, 20 ng of CAP treated vector DNA, 1 µl of 1/10 T4-DNA ligase (NEB) and MilliQ water to 10 µl were prepared. With each set of ligation reactions a control of CAP vector only was also prepared. The reactions were mixed well, briefly centrifuged and incubated overnight at 4°C.

To check that ligation was efficient, 2 µl of the ligation mix was removed prior to the addition of T4-DNA ligase and another 2 µl after ligation was complete. The samples

were examined on an agarose mini-gel (Section 2.10). This method is based on that of Dugaiczyk *et al.* (1975).

2.17.2 Transformation of *E. coli* by Electroporation

2.17.2.1 Preparation of Electrocompetent Cells

One litre of LB broth (Section 2.3.1) was inoculated (1/100) with the desired *E. coli* strain and grown at 37°C with vigorous shaking to mid-log phase (OD_{600} 0.5-1.0, approximately 3 hours). The cells were chilled on ice for 20 minutes then harvested by centrifugation at 4,000 g for 10 minutes (all centrifugations were at 4°C). The cells were washed in 1 L of ice cold water, centrifuged at 4,000 g, re-washed in 500 ml of ice cold water and re-centrifuged. The pellet was finally washed in 20 ml of ice-cold 10% glycerol, centrifuged at 4,000 g and resuspended in 4 ml of ice-cold 10% glycerol. Cell suspensions were stored at -70°C in 200 µl aliquots.

2.17.3.2 Electroporation

Electrocompetent *E. coli* cells were thawed gently at room temperature. Cells (40 µl) and ligated DNA (2 µl, Section 2.17.1.2) were mixed well and left on ice for 1 minute. The Gene Pulser apparatus (Bio-Rad) was set to 25 µF and 2.5 kV and the pulse controller to 200 W, in parallel with the sample chamber. The mixture of DNA and cells was transferred to a cold 0.2 cm cuvette, shaken to the bottom, then pulsed at the above settings resulting in a time constant of 4-5 msec. The cells were immediately resuspended in 1 ml of SOC medium (Section 2.3.3) and incubated at 37°C for 1 hour. The controls (cells and vector only) were also included. Cells were plated at suitable dilutions onto LB plates (Section 2.3.1) containing Ampicillin (Section 2.2), X-gal and IPTG (Section 2.3.1) and recombinants selected using α-complementation. White colonies were screened by gel electrophoresis (Section 2.10) of diagnostic restriction digests (Section 2.9) of isolated plasmid DNA (Section 2.5.3). This method is based on that of Dower *et al.* (1988).

2.18 TECHNIQUE FOR WORKING WITH RNA

All RNA work was performed in either acid washed glassware which had been baked in a dry air oven at 180°C for at least 2 hours or with previously untouched disposable plastic labware. All equipment was handled with clean disposable gloves and all

experimental work was performed wearing these. Plastic labware was stored in acid washed, baked glass beakers and sterilised by autoclaving at 121°C for 15 minutes, separate from other reagents and labware. One set of Gilson pipettors was used exclusively for RNA work and solutions were prepared from previously unopened stocks of reagents. All reagents (except those containing Tris-HCl, Na_2EDTA or 3-[N-morpholino] propane-sulfonic acid [MOPS]) were treated with diethyl pyrocarbonate (DEPC), incubated overnight at 37°C then the DEPC removed by autoclaving twice. Reagents and labware for RNA work was stored separately from that used for other purposes.

2.19 ISOLATION OF RNA FROM FUNGAL CULTURES

The precautions outlined in Section 2.18 were used throughout. Mycelia was grown (Section 2.2), harvested, snap frozen in liquid nitrogen and stored at -70°C. Approximately 5 g of frozen mycelia was ground to a fine powder under liquid nitrogen, resuspended in 5 ml TRIZOL (BRL), homogenised until smooth and incubated at room temperature in a securely covered Corex tube for 5 minutes. All subsequent incubations are carried out at room temperature unless otherwise stated. The homogenate was centrifuged for 10 minutes at 12,000 g (all centrifugations were carried out at 4°C). To the cleared homogenate, 1 ml of chloroform was added, the tube sealed and the mixture shaken vigorously for 15-20 seconds. After incubation for 2-3 minutes, the mixture was centrifuged at 12,000 g for 15 minutes. The aqueous phase was transferred to a tube containing 2.5 ml of isopropanol, mixed, incubated for 10 minutes then centrifuged at 12,000 g for 10 minutes. The supernatant was drained and the RNA precipitate washed with 5 ml of 70% ethanol, the tube sealed with parafilm and left on ice for at least 20 minutes. The RNA was re-centrifuged at 7,500 g for 5 minutes, the ethanol drained and the pellet briefly dried before resuspension in 200 µl water. The RNA was DNase treated (Section 2.19.1) before quantitation (Section 2.19.2).

2.19.1 DNase Treatment of RNA

The isolated RNA (Section 2.19) was treated with RNase free DNase I to remove any contaminating DNA. All the precautions outlined in Section 2.18 were used throughout. The RNA (40 µg) was mixed with: 20 U RNase free DNase I (Boehringer Mannheim); 50 U cloned RNase inhibitor (BRL); 0.5 mM 1,4-dithiothreitol (DTT) and DNase I buffer (100 mM sodium acetate and 5 mM $MgSO_4$ [pH 5.0]) to a final volume of 800 µl. After incubation at 37°C for 30 minutes, a further 20 U of RNase free DNase

I was added and incubated for 30 minutes at 37°C. The reaction was phenol/chloroform extracted (Section 2.6) and the purified RNA precipitated with 0.1 volumes of 3 M sodium acetate and 2.5 volumes of 95% ethanol. The tubes were incubated on ice for 30 minutes before centrifugation in a microcentrifuge for 15 minutes at 4°C. The supernatant was drained, the RNA precipitate washed with 1 ml of 75% ethanol and dried at room temperature for 15 minutes, followed by 5 minutes at 37°C . The RNA was resuspended in 20 µl of MilliQ water, quantitated (Section 2.20.1) and stored at -70°C.

2.20 QUANTIFICATION OF RNA

2.20.1 Spectrophotometric Determination of the Amount of RNA

For quantitation of RNA, readings were taken at wavelengths of 260 nm and 280 nm. The reading at 260 nm allowed calculation of the concentration of nucleic acid in the sample, as an Absorbance of 1 corresponds to approximately 40 µg/ml for RNA. The ratio between the readings at 260 nm and 280 nm provides an estimate of the purity of the nucleic acid, as pure preparations of RNA have a ratio of 2. Concentrated RNA solutions were diluted appropriately and the absorbance determined at both 260 nm and 280 nm.

2.21 NORTHERN BLOTTING OF RNA AND HYBRIDISATION

2.21.1 Glyoxylation of RNA

From the DNased RNA stock 20 µl was removed for glyoxylation. To this 1/10 volume 3 M sodium acetate and 0.55 volumes isopropanol were added, incubated on ice for 15 minutes and centrifuged in a microcentrifuge for 15 minutes at 4°C. The pellet was washed with 70 µl of 75% ethanol, dried at room temperature for 15 minutes followed by 5 minutes at 37°C, before resuspension in 4 µl water. A mixture of deionised Glyoxal (Sigma), 20 x MOPS buffer (400 mM MOPS, 60 mM sodium acetate and 20 mM Na_2EDTA, pH 7.0) and dimethylsulfoxide (DMSO) in a ratio of 50:8:100 was prepared and 16 µl of this was added to the RNA (Section 2.19.1). After mixing thoroughly the RNA was incubated for 1 hour at 50°C under a layer of light mineral oil and stored at -70°C (1-2 days), until run on a Glyoxal gel (Section 2.21.2).

2.21.2 Glyoxal Gel

All precautions outlined in Section 2.18 were used throughout. The electrophoresis equipment was soaked for 30 minutes in a 0.1% (w/v) SDS solution, the plate and comb were soaked in a 0.5% (w/v) sodium hydroxide solution then rinsed with MilliQ water and wiped with 75% ethanol. A surface tension 1% (w/v) agarose gel in 1 x MOPS buffer (Section 2.21.1) was poured onto a 75 x 50 mm glass plate. The gel was placed in the electrophoresis tank and pre-run at 70 V for 10-15 minutes. The glyoxylated RNA (10 µl, Section 2.21.1) and tracker dye (included in a size marker) were electrophoresed with a voltage of 70 at 4°C, until the marker dye had migrated 4/5 the length of the gel. The gel was stained with ethidium bromide (5 µg/µl), destained in water, visualised on a UV transilluminator and photographed using the gel documentation system (Alpha Innotech).

2.21.3 Northern Blotting and Hybridisation

RNA from the Glyoxal gel (Section 2.21.2) was transferred to a Nylon membrane (Amersham) by capillary blotting as in Section 2.14.1 except the gel was not pre-treated in any way and the transfer was performed overnight at 4°C. The unwashed membrane was baked at 80°C in a vacuum oven for 2 hours to reverse the glyoxylation and fix the RNA. The membrane was washed in 6 x SSC for 2 minutes, then in hybridisation solution [6 x SSC; 2 x Denhardts Reagent (dilution of 10 x Denhardts in Section 2.4.2.6) and 0.1% SDS (w/v)], for 2 hours before addition of a [α-^{32}P]dCTP probe prepared as in Section 2.14.2 and overnight hybridisation in 0.2 ml of hybridisation buffer per square cm of filter, at 65°C. The membrane was washed at room temperature in 1 x SSC containing 0.1% (w/v) SDS for 20 minutes and then three times in 0.2 x SSC, 0.1% (w/v) SDS for 20 minutes each. The blot was then exposed, in the presence of a Cronex intensifying screen, to Fuji X-ray film for a suitable period of time at -70°C then developed (Section 2.13.4). If additional washes were required these were carried out for twenty minutes in 0.1 x SSC containing 0.1% (w/v) SDS. This method is based on that of Sambrook *et al.* (1989).

2.22 RT-PCR ANALYSIS OF RNA

This method is based on that of Kawasaki (1990). The precautions outlined in Section 2.18 were used throughout.

2.22.1 Reverse Transcriptase Synthesis of cDNA

Before cDNA synthesis and PCR amplification of cDNA it was necessary to remove all traces of contaminating DNA from the RNA preparation. RNA was therefore treated with DNase I as in Section 2.19.1. In a microcentrifuge tube 50 pmols of random primers (BRL) and approximately 1 µg of DNase I treated RNA were mixed and placed at 95°C for 5 minutes then chilled immediately on ice. To this 1 x Expand Reverse Transcriptase Buffer (Boehringer Mannheim), 1 mM of each dNTP, 20 U cloned RNase inhibitor (BRL), 10 mM DTT, 50 units Expand Reverse Transcriptase (Boehringer Mannheim) and water to 20 µl were added, mixed and incubated at room temperature for 10 minutes, then at 42°C for 45 minutes. The reaction was then stored at -20°C.

2.22.2 Amplification of cDNA by PCR

To 10 µl of the cDNA from Section 2.22.1, 10 pmol of each of the two primers for amplification of the target sequence of interest, 1 x PCR buffer, 2.5 U *Taq* DNA Polymerase (Boehringer Mannheim), 1.25 mM of each dNTP and MilliQ water to 50 µl were added. The reaction was then placed in a Corbett thermocycler preheated to 94°C and after an initial 2 minute melt at 94°C, subjected to 35 cycles of 94°C for 45 seconds, 50°C for 45 seconds and 72°C for 60 seconds. After the 35 cycles were complete the reactions were incubated at 72°C for 5 minutes, then stored at -20°C. The products were separated by electrophoresis as in Section 2.10.

Chapter 3

RESULTS

3.1 MOLECULAR CLONING OF HMG CoA REDUCTASE FROM *NEOTYPHODIUM LOLII*

3.1.1 Primer Design

A data-base search and alignment of previously sequenced *hmg* genes revealed a high degree of sequence conservation in the 3' region of the gene, corresponding to the catalytic domain of the enzyme. This analysis indicated that it would be possible to clone *hmg* from *N. lolii* by PCR using primers designed to this region. The polypeptide sequences of *S. cerevisiae* HMG1 (m22002), *S. cerevisiae* HMG2 (m22255), and *Ustilago maydis* (z30085, translation of nucleotide sequence) from the GenBank data-base, and the human sequence (p04035) from SWISS-PROT, were aligned using the PILEUP function of the GCG package. The conserved region was identified and sequences were selected for primer design (Figure 2A). The sequences identified and the primers designed on the basis of the degenerate code are shown in Figure 2B. Instead of making degenerate primers, inosine was incorporated at all degenerate positions (Sambrook *et al.* 1989).

3.1.2 Isolation of *hmg* from Lp19

3.1.2.1 Establishing Optimal PCR Conditions

A number of parameters including DNA concentration, annealing temperature and magnesium concentration play an important role in determining optimal PCR amplification of a target sequence (Innis and Gelfand 1990). Therefore a series of trial reactions were carried out to test these parameters and optimise the yield of PCR product.

Figure 2A Alignment of partial Hmg polypeptide sequences for primer design.

Alignment of partial Hmg sequences of: *S. cerevisiae* HMG1 (Basson *et al.* 1986), *S. cerevisiae* HMG2 (Basson *et al.* 1986), *U. maydis* (Croxen *et al.* 1994) and human (Luskey and Stevens 1985). The conserved sequences identified for designing primers are shown in blue.

Figure 2B Design of primers to clone *hmg* from *N. lolii*.

Two peptide sequences were identified and selected for primer design. Inosine (blue) was incorporated at all degenerate positions in the nucleotide sequence of primers hmg 1 and hmg 3. These primers should amplify a 359 bp *hmg* fragment from *S. cerevisiae*.

Figure 2A

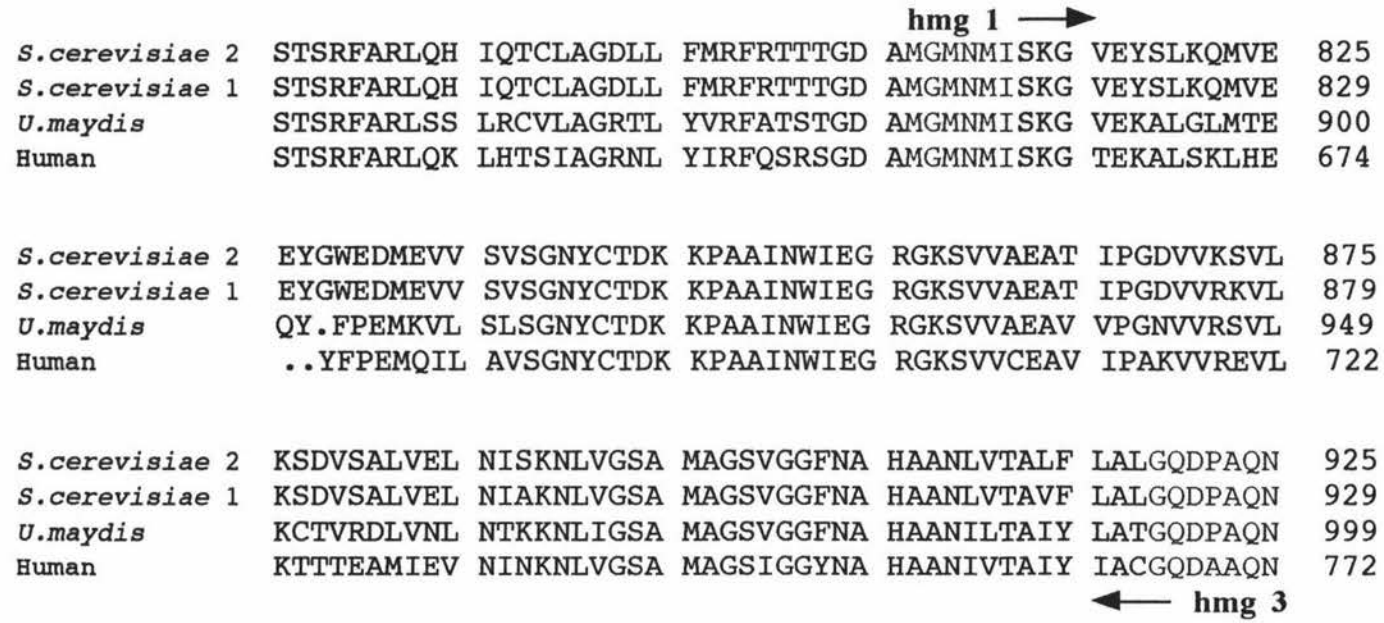
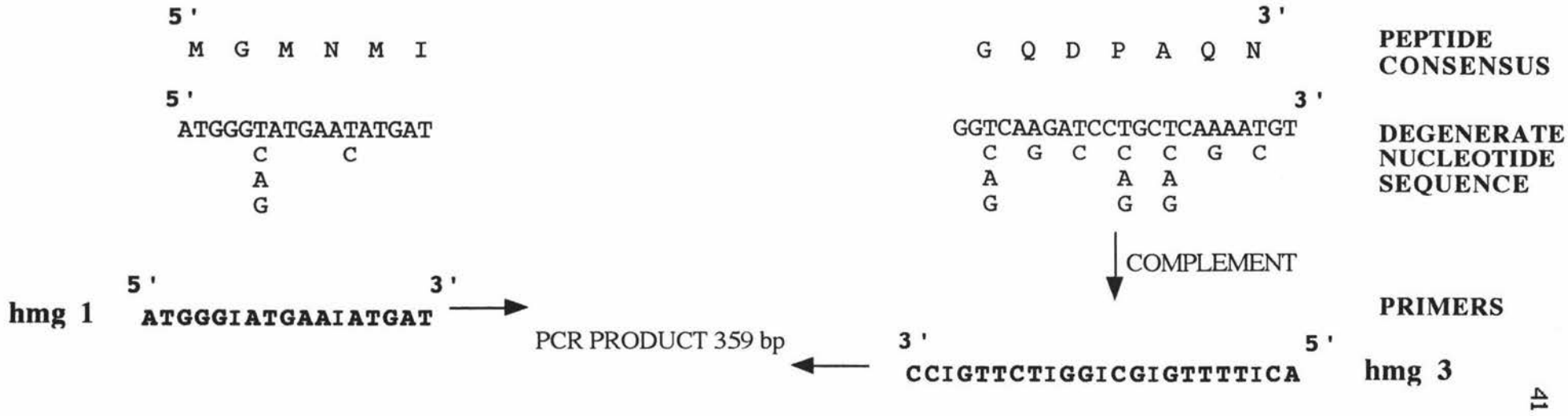


Figure 2B



In addition to *N. lolii*, in some instances, *S. cerevisiae* and *Penicillium paxilli* DNA (25 ng) were also amplified. The primers NS7 and NS8 (Table 2), which amplify a 377 bp fragment of the conserved 18S rRNA gene (White *et al.* 1990), were included as an internal control as they were known to amplify over a range of annealing temperatures.

The first parameter to be tested was the DNA concentration required for optimal amplification. Lp1 and Lp19 DNA was extracted and purified (Section 2.5.1) then used as a template for amplification with primers NS7 and NS8 at an annealing temperature of 55°C in standard Boehringer Mannheim buffer (Section 2.12). These primers amplified a product of the expected size (White *et al.* 1990). The optimal DNA concentration for both Lp1 and Lp19 DNA was 25 ng (Figure 3A).

The next parameter tested was the annealing temperature of the PCR reaction which is determined by the melting temperatures (Tm) of the primers. The Tms for hmg 1 and hmg 3, were 44°C and 66°C respectively (Section 2.12), so a range of annealing temperatures, 42°C-55°C, were trialed using 25 ng of DNA in standard Boehringer Mannheim buffer (Section 2.12). At 48°C and above no products were obtained and below this temperature a range of different sized products were produced (Figure 3B). The greatest yield of the expected 359 bp product was obtained at an annealing temperature of 47°C although other products were still present (Figure 3D). All subsequent PCR reactions with hmg 1 and hmg 3 were carried out at this temperature.

The effect of different magnesium concentrations in the PCR reaction was also assayed. PCR amplifications of 25 ng of DNA were carried out in Promega magnesium free buffer (Section 2.12) with a range of $MgCl_2$ concentrations (0-3 mM), at an annealing temperature of 47°C. Optimal yield of the 359 bp fragment was achieved with 1.5 mM Mg^{2+} (Figure 3C), the same concentration as used previously. The Promega buffer appeared to give a better yield and so was subsequently used for further amplifications with hmg 1 and hmg 3.

3.1.2.2 Isolation of Lp19 *hmg*

Having optimised conditions for amplifying PCR products from Lp19, several reactions were carried out to obtain sufficient product for purification. PCR reactions (Section 2.12) containing 25 ng of Lp19 DNA and 1.5 mM Mg^{2+} were performed with primers hmg 1 and hmg 3 (Section 3.1.1) at an annealing temperature of 47°C. The amplified DNA was separated by electrophoresis on a 2% NuSieve gel (Section 2.10, Figure 3D). A range of different sized products were amplified which may be due to the low annealing temperature used and the degeneracy of the primers. This may have created a

Figure 3 Optimisation of PCR amplification conditions.

Figure 3A Effect of DNA concentration on yield of PCR products.

PCR was carried out on Lp1 and Lp19 DNA at a range of concentrations with the primers NS7 and NS8. Lanes 3-7 contain Lp1 DNA at: 10 pg, 5 ng, 10 ng, 25 ng and 300 ng and lanes 8-12 contain Lp19 DNA at: 10 pg, 5 ng, 10 ng, 25 ng and 500 ng. No DNA was present in lane 2 so the resulting fragment may be due to cross contamination. Lane 1 contains the BRL 1 kb ladder.

Figure 3B Amplification of DNA at 46°C.

Lanes 2-6 contain amplification products produced with primers NS7 and NS8. The reactions contained: No DNA, lane 2; Lp1, lane 3; Lp19, lane 4; *P. paxilli*, lane 5 and *S. cerevisiae*, lane 6. Lanes 7-11 contained the amplification reactions using hmg 1 and hmg 3: No DNA, lane 7; Lp1, lane 8; Lp19, lane 9; *P. paxilli*, lane 10 and *S. cerevisiae*, lane 11. Lane 1 contained the BRL 1 kb ladder.

Figure 3C Effect of magnesium concentration on amplification of PCR products.

PCR amplification of Lp19 DNA was carried out in Promega *Taq* buffer with a range of Mg^{2+} concentrations: 0.5 mM, lane 4; 1 mM, lane 5; 1.5 mM, lane 6; 2 mM, lane 7; 2.5 mM, lane 8 and 3 mM, lane 9. Lane 1 contained the BRL 1 kb ladder; lane 2, no DNA in Boehringer Mannheim buffer; lane 3, no Mg^{2+} in Promega buffer and lane 10, *S. cerevisiae* DNA in Boehringer Mannheim buffer.

Figure 3D PCR amplification of fungal genomic DNA.

PCR was carried out on Lp19 DNA (25 ng) at 1.5 mM Mg^{2+} (Promega) with the primers hmg 1 and hmg 3 and an annealing temperature of 47°C (lane 2). Lane 1 contains the BRL 1 kb ladder.

Figure 3A

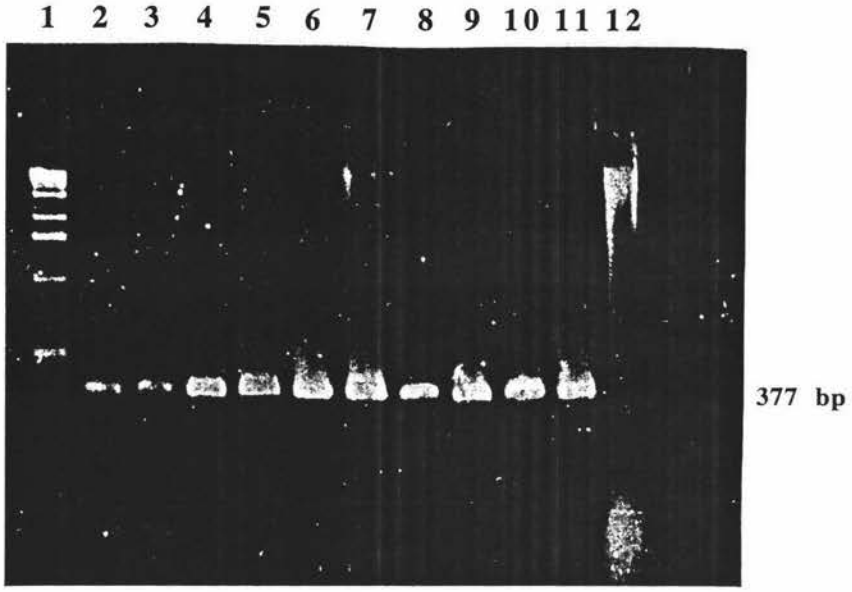


Figure 3B

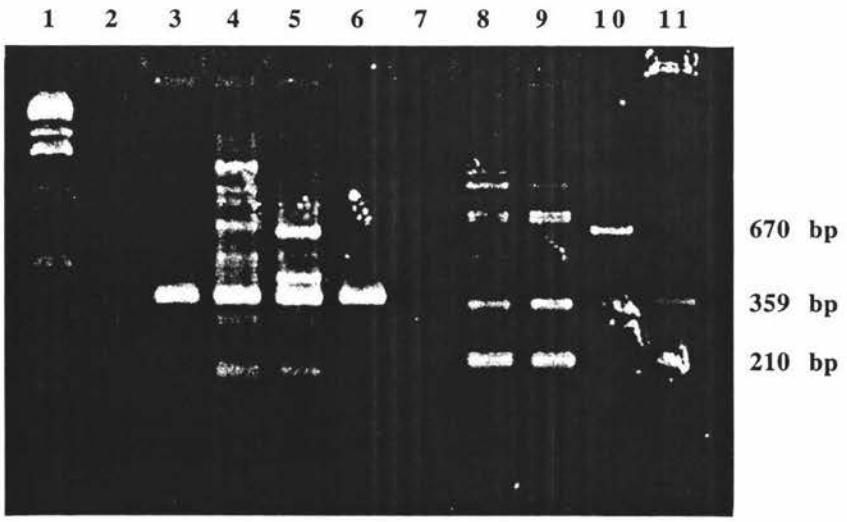


Figure 3C

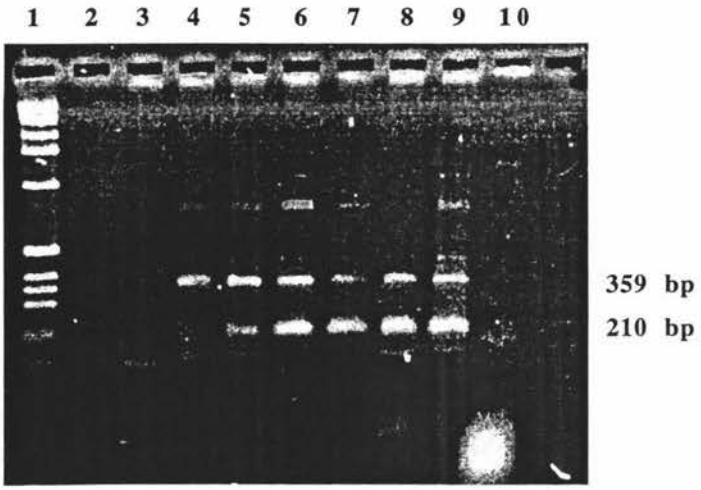
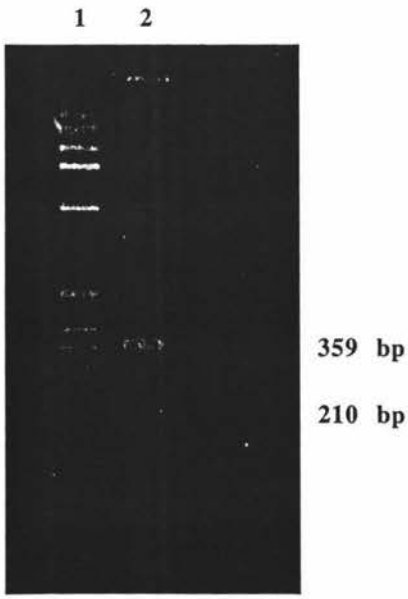


Figure 3D



RAPD-like reaction amplifying repetitive sequences or as the primers were created in the Hmg catalytic domain, other enzymes with similar catalytic functions may have been amplified. The most abundant products were a 210 bp fragment and the expected 359 bp fragment (Figure 3D), both of which were purified from a 2% NuSieve gel (Section 2.11).

3.1.3 Confirmation that Isolated Fragments were from the Lp19 Genome

3.1.3.1 Analysis by Southern Hybridisation

To confirm that the PCR amplified fragments isolated were part of the *N. lolii* genome both were labelled with [α-^{32}P]dCTP (Section 2.14.2) and hybridised to Southern blots (Section 2.14) containing Lp19 and Lp1 genomic DNA (both 2 µg) digested with *Eco*RI and *Hin*dIII as well as 10 µl of 1:100 dilution of the total PCR reaction (Figure 4A). Both labelled PCR fragments hybridised to sequences within the Lp1 and Lp19 genomes (Figures 4B and 4C). The 359 bp fragment hybridised (Figure 4B) to a DNA fragment greater than 23 kb in the *Eco*RI digest (lane 4) and one of approximately 18 kb in the *Hin*dIII digest (lane 5) of Lp19, whereas it hybridised to two fragments in the Lp1 genomic digest. In the *Eco*RI digest (lane 2), a fragment of approximately 4 kb and one greater than 23 kb hybridised and in the *Hin*dIII digest fragments of approximately 8 kb and 18 kb hybridised (lane 3). These results are consistent with Lp19 being haploid and Lp1 an interspecific hybrid (diploid) (Schardl *et al.* 1994). As expected the PCR product hybridised to the corresponding sized fragment in the PCR reaction (lane 10). The presence of an extra band may be due to cross-contamination during probe isolation.

The 210 bp fragment hybridised to a number of different sized fragments in the Lp1 and Lp19 genomic digests (Figure 4C). Of the multiple bands in the Lp1 digests two bands hybridised more strongly. In the *Eco*RI digest (lane 2) fragments of approximately 12 kb and one greater than 23 kb hybridised and fragments of approximately 20 kb and 6 kb in the *Hin*dIII digest (lane 3). From the multiple hybridising fragments in the Lp19 digests, one band in each digest hybridised more strongly. A fragment of approximately 12 kb in the *Eco*RI digest (lane 4) and one of approximately 5 kb in the *Hin*dIII digest (lane 5). Due to the degeneracy of the primers, the low PCR annealing temperature and the fact that the primers were designed within the Hmg catalytic domain, the multiple bands may relate to an enzyme family with similar catalytic functions or the 210 bp fragment may contain repetitive sequences. The PCR product

Figure 4 Southern analysis of the PCR isolated fragments.

Figure 4A Southern blot of *Eco*RI and *Hin*dIII digested Lp1 and Lp19 genomic DNA.

This gel contained: 2 μg of Lp1 DNA digested with *Eco*RI (lane 2) and *Hin*dIII (lane 3); 2 μg of Lp19 DNA digested with *Eco*RI (lane 4) and *Hin*dIII (lane 5); and 10 μl of a 1:100 dilution of the PCR amplified products shown in Figure 3D (lane 10). Lane 1 contains *Hin*dIII digest λ DNA and lane 6 contains the BRL 1 kb ladder.

Figure 4B Confirmation that the 359 nt fragment is part of the *N. lolii* genome.

An autoradiograph of a Southern blot of Figure 4A hybridised with the 359 bp fragment.

Figure 4C Confirmation that the 210 nt fragment is part of the *N. lolii* genome.

An autoradiograph of a Southern blot of Figure 4A hybridised with the 210 bp fragment.

Figure 4A

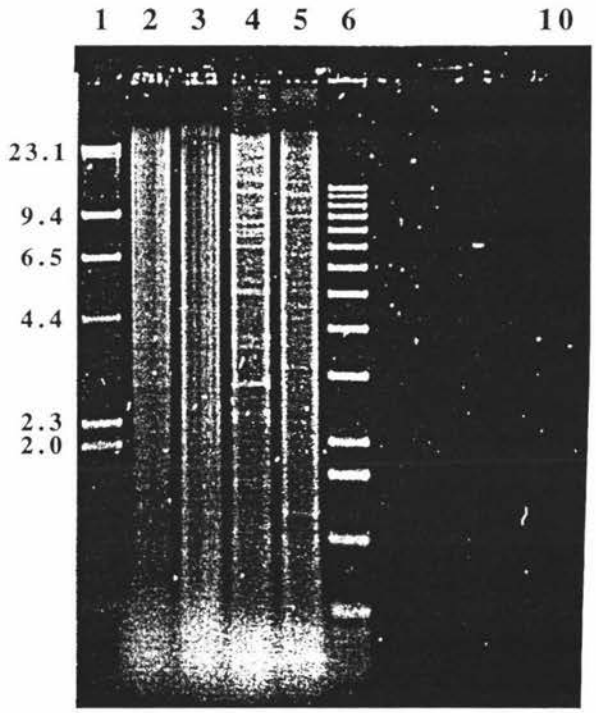


Figure 4B

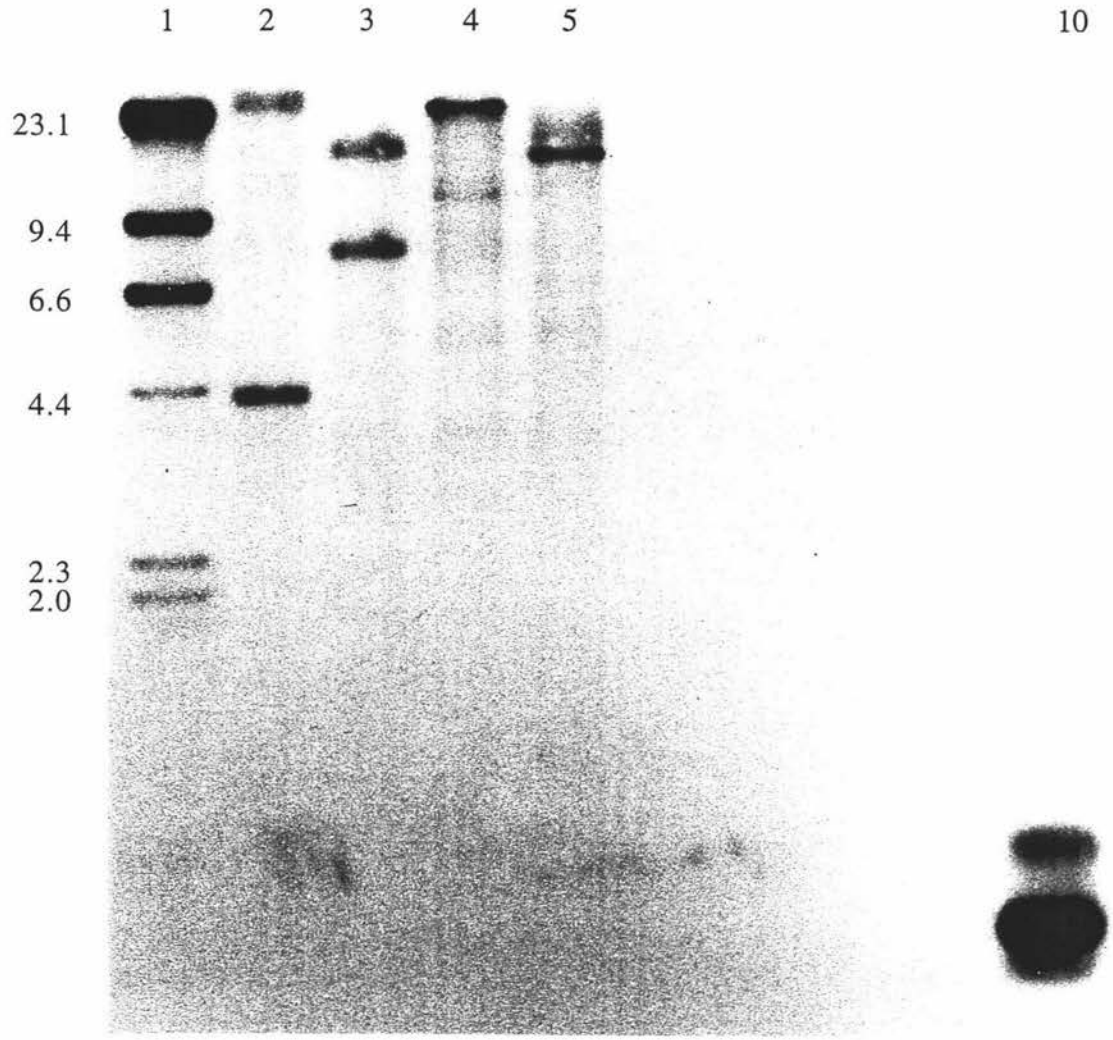
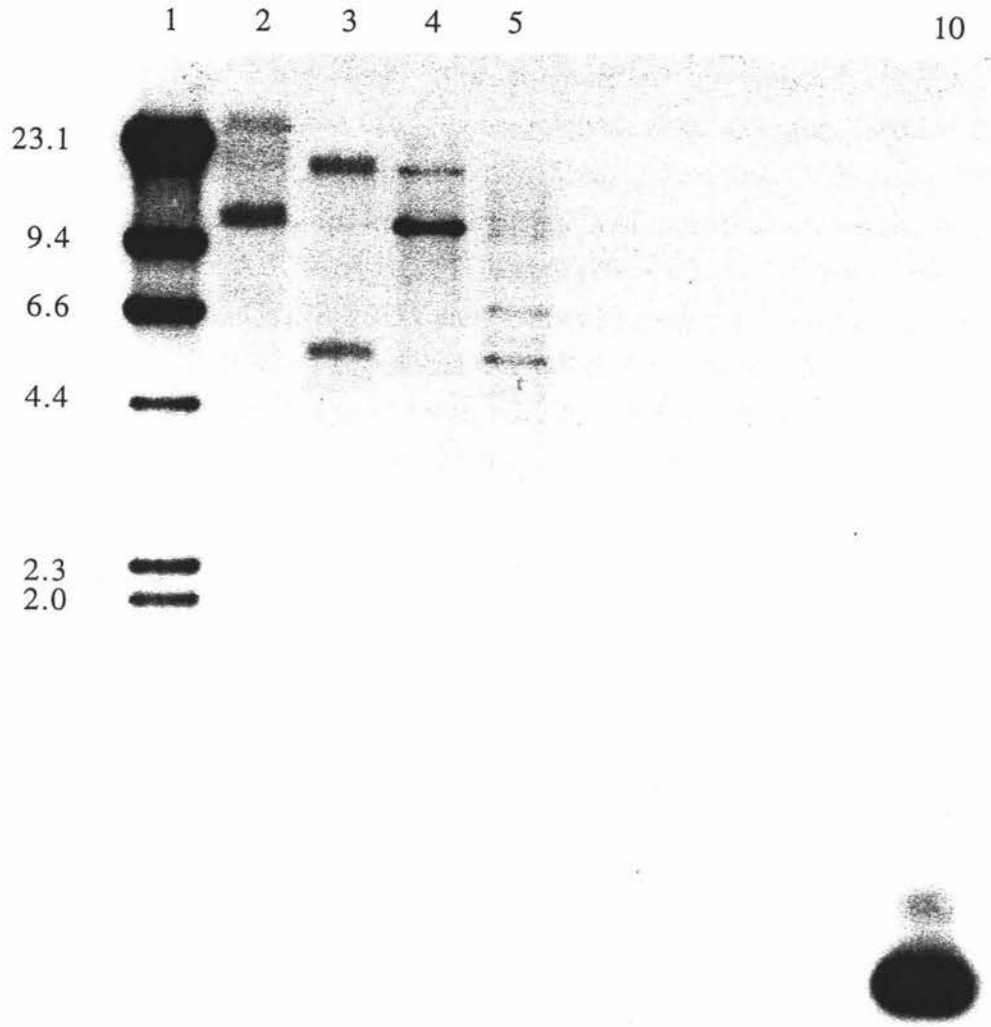


Figure 4C



hybridised to the expected sized fragment in the PCR reaction and the appearance of an extra band may again be due to cross-contamination during probe isolation (lane 10).

3.1.3.2 Sequence Analysis of the PCR Isolated 359 nt Fragment

Sequence analysis of the 359 bp PCR amplified fragment was carried out to confirm that it was part of Lp19 *hmg*. The gel purified fragment (Section 3.1.2.2) was sequenced in both directions using the Amplicycle Sequencing kit (Section 2.13.1) and the same primers used for amplification, hmg 1 and hmg 3 (Section 3.1.1). The nucleotide sequence obtained and the translation of that sequence are shown in Figure 5 (see also Appendix A 1.1). A BLAST search (Altschul *et al.* 1990) of the protein databases (Brookhaven, SWISS-PROT and GenBank) confirmed that the sequence obtained was similar to other Hmg sequences, including those from: *S. cerevisiae* HMG2, *S. cerevisiae* HMG1, tomato and wheat, with BLAST scores of 447, 444, 359 and 355 respectively. Comparison of this translation with other Hmg polypeptide sequences confirmed that the fragment sequenced was a unique *hmg* (Figure 6). This sequence information was used to design primers, hmg 4 and hmg 5 (Figure 5, Table 2), that were identical to the Lp19 *hmg*, so that a single fragment could be obtained for library screening.

3.1.4 Amplification of Lp19 *hmg* Gene Fragment

PCR amplification (Section 2.12) of 25 ng of Lp19 genomic DNA was carried out with primers hmg 4 and hmg 5 (Section 3.1.4.2, Table 2) at an annealing temperature of 65°C using standard Boehringer Mannheim buffer (Section 2.12). The amplified DNA was separated on a 3% NuSieve gel (Section 2.10) and shown to be a unique product of 265 bp (Figure 5, data not shown). This 265 bp fragment was sequenced in both directions using the Amplicycle kit (Section 2.13.1) with primers hmg 4 and hmg 5 and was shown to have 100% identity to the sequence of the 359 bp fragment. The 265 bp *hmg* fragment was then used to screen an Lp19 genomic library.

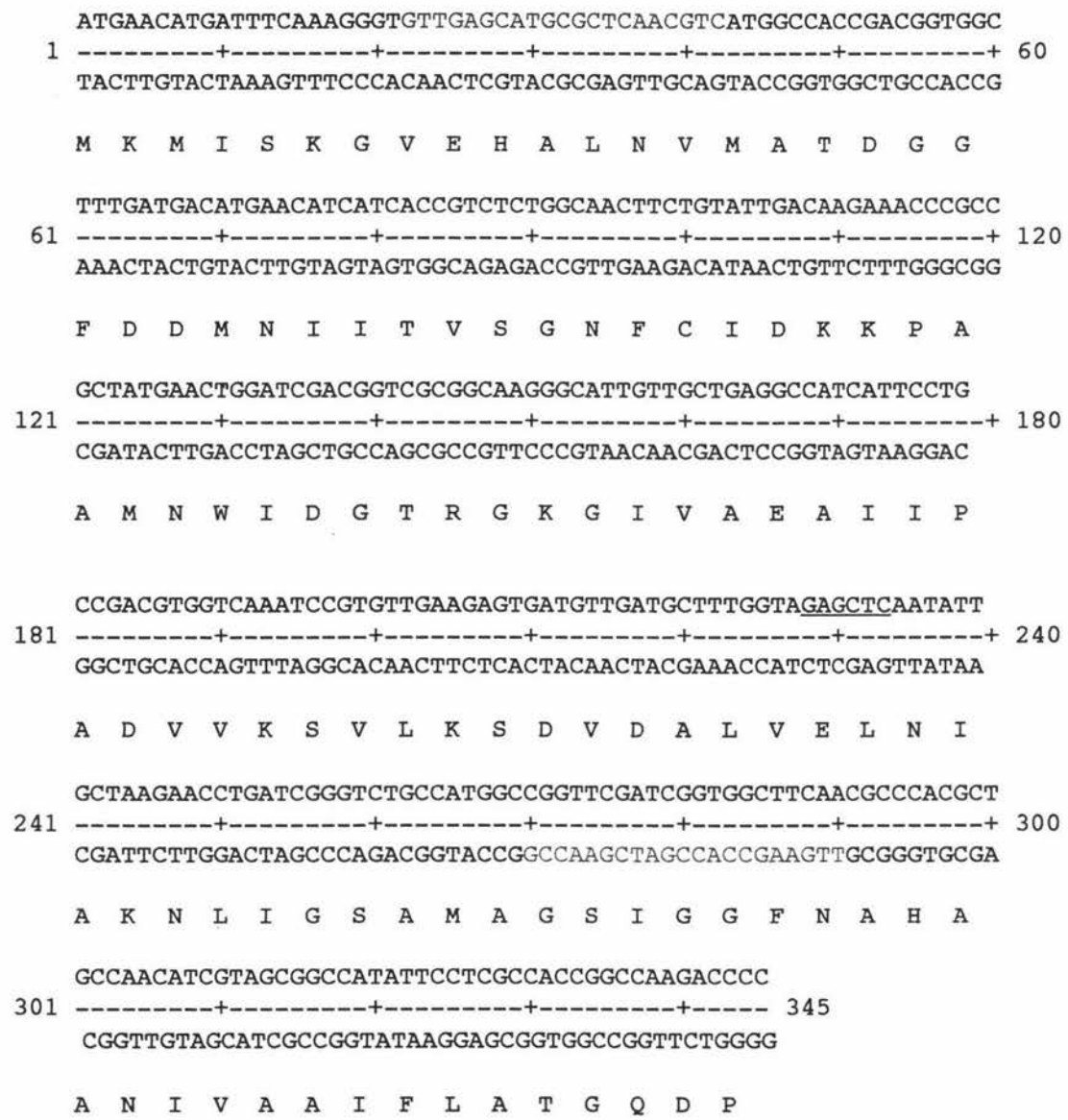


Figure 5 Partial DNA sequence and deduced amino acid sequence of the 359 bp *hmg* fragment.

The partial DNA sequence (345 bp) and deduced amino acid sequence of *hmg*, obtained by PCR amplification of Lp19 DNA with primers hmg 1 and hmg 3. The sequence was used to design primers unique to Lp19 *hmg*, and are shown in red (hmg 4) and blue (hmg 5). An *Sst*I restriction site was located within this fragment (underlined) which was important for later analysis.

```
Lp19                                                MKMISKG VEHALNVMAT
S.cere 2   STSRFARLQH IQTCLAGDLL FMRFRTTTGD AMGMNMISKG VEYSLKQMVE  825
S.cere 1   STSRFARLQH IQTCLAGDLL FMRFRTTTGD AMGMNMISKG VEYSLKQMVE  829
U.maydis   STSRFARLSS LRCVLAGRTL YVRFATSTGD AMGMNMISKG VEKALGLMTE  900
Human      STSRFARLQK LHTSIAGRNL YIRFQSRSGD AMGMNMISKG TEKALSKLHE  674

Lp19       DGGFDDMNII TVSGNFCIDK KPAAMNWIDG RGKGIVAEAI IPADVVKSVL
S.cere 2   EYGWEDMEVV SVSGNYCTDK KPAAINWIEG RGKSVVAEAT IPGDVVKSVL  875
S.cere 1   EYGWEDMEVV SVSGNYCTDK KPAAINWIEG RGKSVVAEAT IPGDVVRKVL  879
U.maydis   QY.FPEMKVL SLSGNYCTDK KPAAINWIEG RGKSVVAEAV VPGNVVRSVL  949
Human      ..YFPEMQIL AVSGNYCTDK KPAAINWIEG RGKSVVCEAV IPAKVVREVL  722

Lp19       KSDVDALVEL NIAKNLIGSA MAGSIGGFNA HAANIVXXIF LATGQDPAKT
S.cere 2   KSDVSALVEL NISKNLVGSA MAGSVGGFNA HAANLVTALF LALGQDPAQN  925
S.cere 1   KSDVSALVEL NIAKNLVGSA MAGSVGGFNA HAANLVTAVF LALGQDPAQN  929
U.maydis   KCTVRDLVNL NTKKNLIGSA MAGSVGGFNA HAANILTAIY LATGQDPAQN  999
Human      KTTTEAMIEV NINKNLVGSA MAGSIGGYNA HAANIVTAIY IACGQDAAQN  772
```

Figure 6 Pile up of partial Hmg sequences showing similarity to Lp19 Hmg.

Alignment of the deduced polypeptide sequence of *hmg* from *N. lolii* Lp19 with Hmg sequences from: *S. cerevisiae* HMG1 (*S. cere* 1; Basson *et al.* 1986), *S. cerevisiae* HMG2 (*S. cere* 2; Basson *et al.* 1986), *U. maydis* (Croxen *et al.* 1994) and human (Luskey and Stevens 1985). *S. cerevisiae* has the greatest sequence similarity to Lp19 Hmg, 71%. Sequence identity is shown in blue.

3.2 LIBRARY CONSTRUCTION

3.2.1 Optimising Conditions for Enzymatic Digestion of DNA

N. lolii Lp19 DNA (200 ng/µl) was used to construct a λGEM-12 genomic library. Trial digests (Section 2.9) of 1 µg of DNA with the isochizomers *Mbo*I and *Sau*3A were performed and *Mbo*I was found to digest Lp19 DNA efficiently (data not shown). *Mbo*I trial digests were undertaken to determine the enzyme concentration required to produce the maximum yield of 9-23 kb fragments of Lp19 genomic DNA (Section 2.15.1). Enzyme concentrations ranging from 0.008-1.0 U were trialed and the digested DNA separated by electrophoresis on a 0.4% agarose gel (Section 2.10, Figure 7). The greatest mass of DNA fragments in the size range of 9-23 kb were obtained with 0.03 U (lane 7) of *Mbo*I. The best yield of fragments of the desired size range in a large scale partial digest usually occurs when one half the determined optimal enzyme concentration is used (Sambrook *et al.* 1989), in this case 0.015 U (lane 8). A large scale partial digest (Section 2.15.2) was carried out using this enzyme concentration as well as additional reactions with 0.06 U and 0.03 U to ensure the optimal digest was obtained. The samples were checked on a 0.4% agarose overnight gel (Section 2.10) and the digests using *Mbo*I concentrations of 0.03 U and 0.015 U gave the optimal concentration of fragments in the 9-23 kb size range so these were pooled for ligation into λ arms and the third reaction mix discarded.

3.2.2 Determination of Optimal Ligation Conditions

Prior to ligation of insert DNA into λ arms the insert ends were half-site filled with adenine and guanine using Boehringer Mannheim's Klenow enzyme, to make them complementary to the *Xho*I half-site arms of λGEM-12 (Section 2.15.3). The DNA was phenol/chloroform extracted (Section 2.6), ethanol precipitated and resuspended in water to a final concentration of approximately 200 ng/µl.

Trial ligations (Section 2.15.4) with varying vector:insert DNA ratios were performed to determine the optimal ligation conditions to give sufficient plaque forming units (PFU) for a representative Lp19 genomic library (Table 3). These ligations (2.5 µl) were packaged into phage in small scale reactions (250 µl, Section 2.15.5) and titred (Section 2.15.6) to analyse the efficiency of the ligation and packaging reactions. The minimum requirement for a genomic library of 9-23 kb inserts representing 99% of a 25 Mbp genome is approximately 7.2×10^3 PFUs (Clarke and Carbon 1976), but a 10-fold over-representation increases the probability of obtaining the gene of interest. Ligation ratios

Figure 7 Establishing conditions for partial digestion of high-molecular weight Lp19 DNA.

Lp19 genomic DNA (1 μg) was digested with varying amounts of *Mbo*I to determine conditions that gave the greatest mass of fragments in the 9-23 kb size range. Lane 2 contained 1 U; lane 3, 0.5 U; lane 4, 0.25 U; lane 5, 0.125 U; lane 6, 0.6 U; lane 7, 0.03 U; lane 8, 0.015 U and lane 9, 0.008 U. Lane 10 contained 1 μg undigested Lp19 DNA and lane 1 contained the BRL high molecular weight markers.

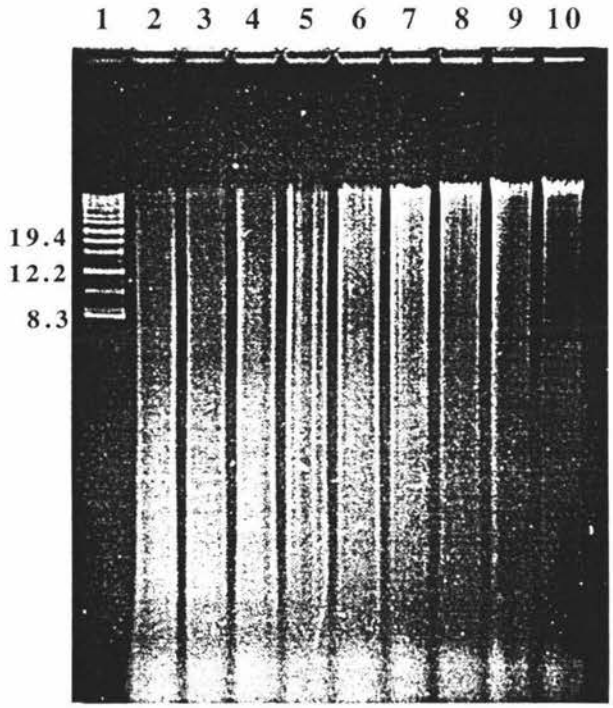


Table 3 Establishing optimal ligation conditions for library construction.

	arms:insert ratio				
	1:1	1:2	1:3	4:1	control
[1]Trial ligations					
µg	1:1	1:2	1:3	1:0.25	[2]1:1
PFU/ml	$1.79x10^5$	$2.44x10^5$	$1.2x10^4$	$1.45x10^4$	$2.8x10^6$
Total PFUs in 250 µl reaction	$4.48x10^4$	$6.1x10^4$	$3.0x10^3$	$3.6x10^3$	$7.0x10^5$
[3]Large scale ligation					
PFU/ml	$2.9x10^5$	$2.5x10^5$			
PFU/100µl	$2.9x10^4$	$2.5x10^4$			
Total number of λ plaques	$1.2x10^5$	$1.0x10^5$			

[1] Trial ligation ratios 1:1 and 1:2 were shown to give sufficient plaques for a representative library when packaged (250 µl reaction) and titred. These ligations were then re-packaged in a large scale packaging reaction (500 µl).

[2] Positive control insert DNA (pTI11, pre-digested with *Bam*HI and half-site filled) was ligated into λ arms.

[3] Three 500 µl and one 250 µl packaging reactions of each ligation ratio (1:1 and 1:2) were carried out. The reaction titres were determined and the reactions pooled having sufficient PFUs for a representative library. The Lp19 library was amplified before storage.

of 1:1, 1:2, 1:3 and 4:1 (λ arms:insert DNA) were trialed (Section 2.15.4). Ligated DNA was packaged into phage in small scale reactions (Section 2.15.5) and these were titred (Section 2.15.6) to determine which ratio gave sufficient plaques for a representative library of the Lp19 genome. From these trial reactions the 1:1 and 1:2 ratios produced the highest number of plaques (Table 3) so these were selected for large scale packaging to produce the library. The 1:1 and 1:2 ligations (5 µl) were repackaged into λ arms in a large scale reaction (500 µl), pooled, titred and estimated to have $2.2x10^5$ PFU, sufficient for a representative library of the *N. lolii* Lp19 genome. The Lp19 genomic library was amplified (Section 2.15.7) to $1.5x10^9$ PFU/ml as the original phage titre was too low for long-term storage in 7% DMSO at -20°C.

3.3 LIBRARY SCREENING AND MAPPING OF *HMG* POSITIVE CLONES

3.3.1 Library Screening

Filters containing approximately 23,000 PFU from the λGEM-12 genomic library prepared from Lp19 DNA (Section 2.15), were screened by plaque hybridisation (Section 2.16.2) with the 265 bp *hmg* fragment (Section 3.2.5). This probe was [α-^{32}P]dCTP-labelled (Section 2.14.2-2.14.3) for round one of screening and DIG-labelled (Section 2.14.5) for subsequent rounds. In the first round of hybridisation 16 positive plaques were isolated (Section 2.16.3), 5 of which were selected for second round screening. Four of these were confirmed as positive (λJD1, λJD7, λJD12 and λJD14). After a third round of plaque purification, DNA was isolated from 3 (λJD1, λJD7 and λJD12) of the 4 positive clones (Section 2.5.2).

3.3.2 Physical Mapping of λ Clones

3.3.2.1 Restriction Digestion of λ Clones and Hybridisation to Lp19 *hmg*

DNA prepared from λJD1, λJD7 and λJD12 was digested (Section 2.9) with *Sst*I, *Hin*dIII and *Sst*I-*Hin*dII. *Sst*I was selected for analysis as it cleaves within the multiple cloning site, releasing the entire insert from the λ arms, whereas *Hin*dIII does not (see Appendix 2.0). Digested λ DNA was separated by electrophoresis on a 0.7% agarose overnight gel (Section 2.10, Figure 8A), alongside digested Lp19 genomic DNA. The sizes of the fragments produced are summarised in Table 4. All three clones shared common insert bands suggesting they were from a common region of the genome. A

Southern blot (Section 2.14.1) of this gel was probed with the DIG-labelled (Section 2.14.5-2.14.7) 265 bp *hmg* fragment of Lp19 (Figure 8B). The autoradiograph of this blot revealed that all three λ clones contained fragments hybridising to the *hmg* probe (fragments hybridising are identified by an asterisk in Table 4). The fragments hybridising in each *Hin*dIII digest contain a segment of the right λGEM-12 arm as there is no *Hin*dIII site within the multiple cloning cassette (see Appendix 2.0). The presence of two hybridising *Sst*I fragments is in agreement with the presence of an *Sst*I site within the probe (Figure 5). This also shows there to be a *Hin*dIII site within the 0.8 kb *Sst*I fragment of each clone decreasing it to a 0.7 kb fragment. The absence of hybridisation to the Lp19 genomic digest may be a result of the DNA concentration being too low (see repeat experiment, Figure 11B). Clone λJD12 was selected for further analysis.

3.3.2.2 Mapping of λJD12

Further restriction analysis (Section 2.9) of clone λJD12 was carried out using *Bam*HI, *Eco*RI, *Hin*dIII and *Sst*I to generate a physical map. With the exception of *Hin*dIII these restriction enzymes all cleave at the multiple cloning site in λGEM-12 thereby completely releasing insert DNA (see Appendix 2.0). Digests with each enzyme alone and in combination with *Sst*I, were carried out and the DNA separated by electrophoresis on a 0.7% agarose overnight gel (Section 2.10, Figure 9A). The size of the fragments obtained is summarised in Table 5. A Southern blot (Section 2.14.1) of this gel was probed with the DIG-labelled (Section 2.14.5-2.14.7) 265 bp *hmg* fragment (Section 3.2.5, Figure 9B). Hybridising fragments are identified by an asterisk in Table 5.

Fragments obtained from restriction analysis of λJD12 (Figure 9A) were used to construct a physical map (Figure 10). The map was assembled by comparing the fragment sizes obtained from single enzyme digests to those obtained in the double digests with *Sst*I. To aid mapping, each of the five λJD12 *Sst*I fragments (5.2 kb, 4.9 kb. 1.8 kb, 0.8 kb and 0.6 kb, Figure 10) were subcloned (Section 2.17) into pUC118 (see Appendix A 3.1). Initially, 48 positive transformants were isolated and DNA was obtained from these using alkaline lysis (Section 2.5.3). The transformants were screened by restriction digestion with *Sst*I (Section 2.9) and analysed gel electrophoresis (Section 2.10, data not shown). The plasmids were then digested with the restriction enzymes *Bam*HI, *Eco*RI, *Hin*dIII, and *Pst*I to determine the orientation of the insert within the vector. Plasmids containing each *Sst*I fragment of lJD12 in both orientations were isolated and recorded as: pJD4(-) and pJD11(+), 0.6 kb; pJD6(-) and pJD12(+),

Figure 8A Restriction digestion profiles of λ clones digested with *Sst*I, *Hin*dIII and *Sst*I-*Hin*dIII.

λJD1, λJD7, λJD12 (approximately 300 ng) and 1 μg of Lp19 total DNA was digested with: *Hin*dIII, lanes 2-5; *Sst*I, lanes 6-9 and *Hin*dIII-*Sst*I lanes 10-13. Lanes 1 and 14 contain *Hin*dIII digested λ DNA and the BRL 1 kb ladder, respectively.

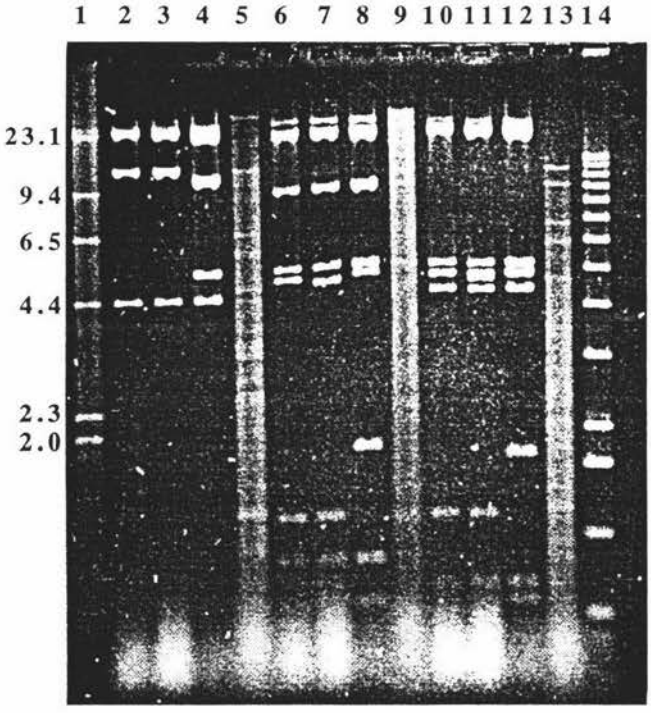


Figure 8B Mapping the position of *hmg* on λJD1, λJD7 and λJD12

An autoradiograph of a Southern blot of Figure 8A hybridised with the DIG-labelled 265 bp *hmg* fragment.

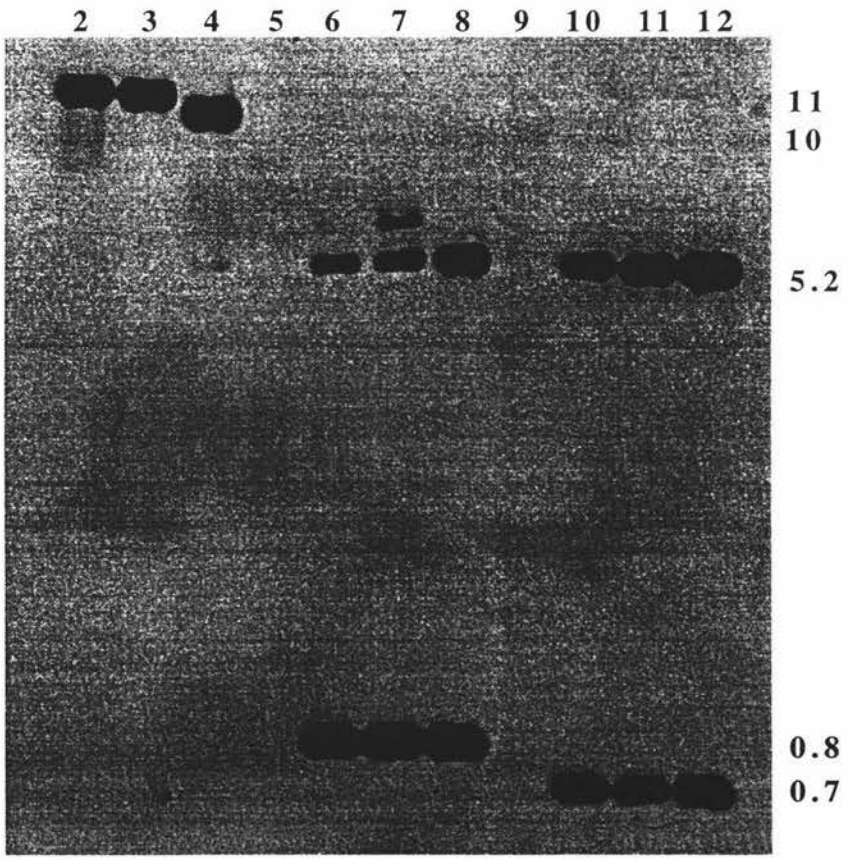


Table 4 Summary of restriction fragment sizes of clones λJD1, λJD7 and λJD12.

*Hin*dIII			*Sst*I			*Sst*I-*Hin*dIII		
λJD1	λJD7	λJD12	λJD1	λJD7	λJD12	λJD1	λJD7	λJD12
11*	11*							
		10*						
			9.2	9.2	9.2			
					5.3	5.3*	5.3*	5.3*
				5.1				
			5.0		5.0	5.0	5.0	5.0
		4.9					4.9	
			4.8	4.8				
						4.7	4.7	4.7
4.4	4.4	4.4						
					1.9			
								1.8
			1.2	1.2		1.2		
			0.8*	0.8*	0.8*			
						0.7*	0.7*	0.7*
					0.6			0.6

* Indicates fragments in Figure 8B hybridising to the 265 bp DIG-labelled *hmg* fragment.

The left λGEM-12 arm (20 kb) is not included in this Table.

Fragment sizes were determined using the Gel Documentation System (Alpha Innotech).

Figure 9A Restriction digestion profile of λJD12.

λJD12 (approximately 300 ng) digested with: *Sst*I (lane 3), *Bam*HI (lane 4), *Bam*HI-*Sst*I (lane 5), *Eco*RI (lane 6), *Eco*RI-*Sst*I (lane 7), *Hin*dIII (lane 8) and *Hin*dIII-*Sst*I (lane 9). Lane 1 and 2 contain the BRL 1 kb ladder and *Hin*dIII digested λ DNA, respectively.

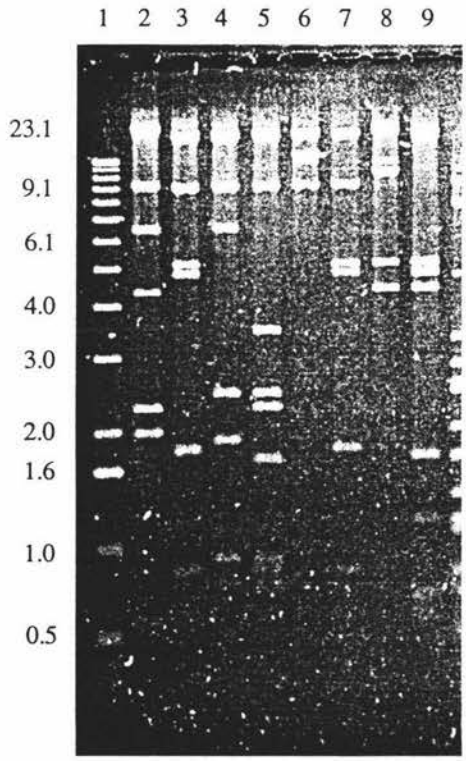


Figure 9B Mapping position of *hmg* within λJD12.

Autoradiograph of the Southern blot of Figure 9A hybridised with the DIG-labelled 265 bp *hmg* fragment of Lp19. The background in the autoradiograph may be due to too much DIG-labelled probe being added during hybridisation.

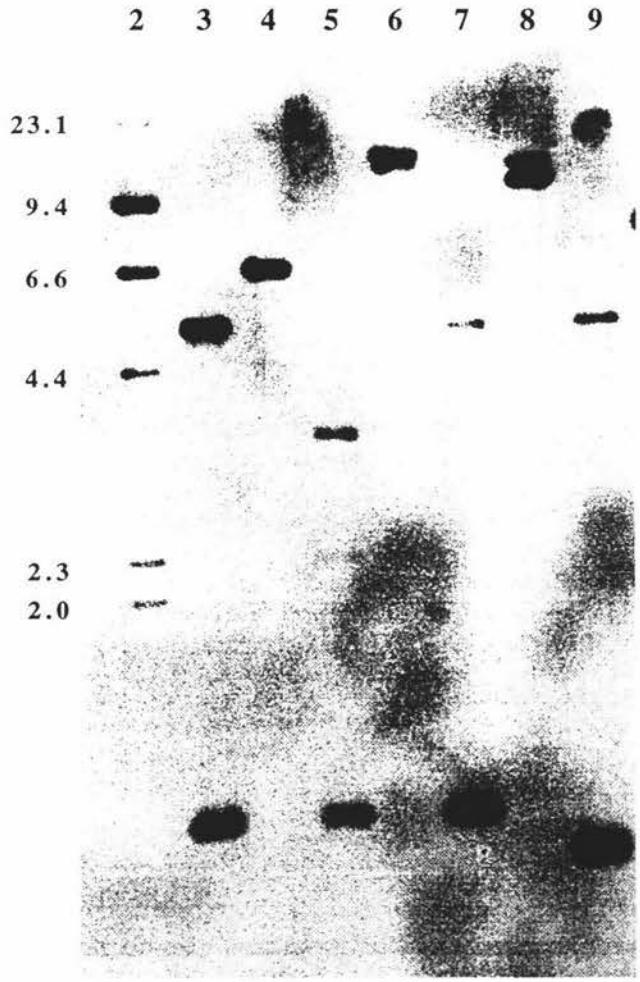


Table 5 Summary of restriction fragment sizes generated by digests of λJD12.

SstI	*Bam*HI	B-S	*Eco*RI	E-S	*Hin*dIII	H-S
			13.2*			
						10*
9.4	9.4	9.4	9.4	9.4		
	6.5*					
5.2*				5.2*	5.2	5.2*
4.9				4.9		
						4.8
					4.4	4.4
		3.6*				
				2.6	2.6	
		2.3				
	2.0					
	1.9				1.9	
			1.8		1.8	
						1.1
	0.9	0.9				
0.8*		0.8*		0.8		
						0.7*
0.6						0.6

B-S = *Bam*HI-*Sst*I double digest.

E-S = *Eco*RI-*Sst*I double digest.

H-S = *Hin*dIII-*Sst*I double digest.

* Indicates fragments in Figure 9B hybridising to the 265 bp DIG-labbeled *hmg* fragment of Lp19.

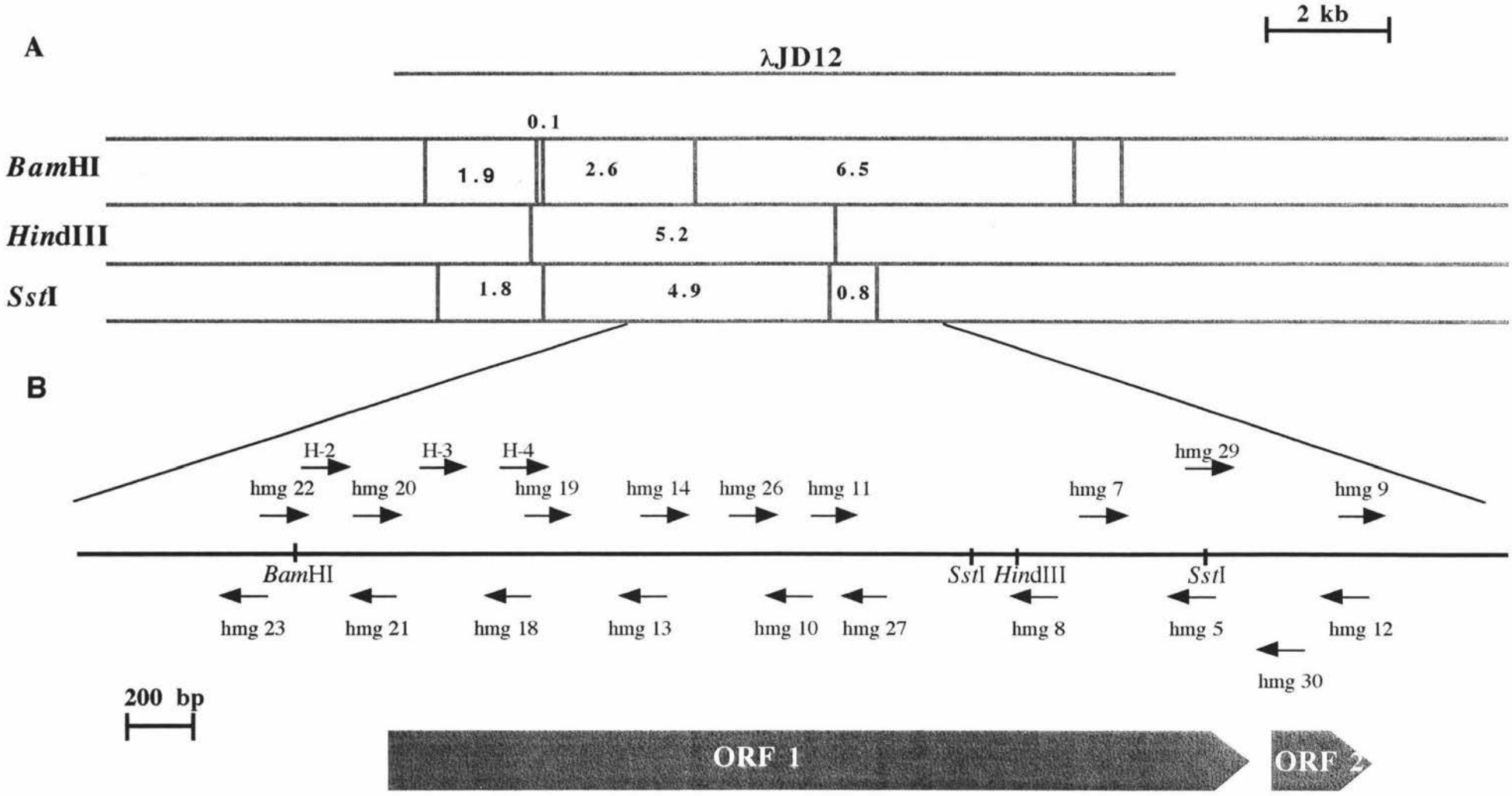
The left λGEM-12 arm is not included on this Table.

Fragment sizes were determined using the Alpha Innotech Gel Documentation System.

Figure 10 Physical map of the *hmg* region of the Lp19 genome.

A. Restriction map of the λJD12 clone for the enzymes *Bam*HI, *Hin*dIII and *Sst*I.

B. The region sequenced, the *hmg* ORFs and the primers used for PCR and sequencing, are shown.


2 kb
A
λJD12
0.1
BamHI
1.9
2.6
6.5
HindIII
5.2
SstI
1.8
4.9
0.8
B
H-2
H-3
H-4
hmg 29
hmg 22
hmg 20
hmg 19
hmg 14
hmg 26
hmg 11
hmg 7
hmg 9
BamHI
SstI
HindIII
SstI
hmg 23
hmg 21
hmg 18
hmg 13
hmg 10
hmg 27
hmg 8
hmg 5
hmg 12
hmg 30
200 bp
ORF 1
ORF 2

0.8 kb; pJD35(-) and pJD16(+), 1.8 kb; pJD21(-) and pJD48(+), 4.9 kb and pJD20(-) and pJD1(+), 5.2 kb (the plasmids are designated + or - depending on their orientation with respect to the map shown in Figure 10). Single and double digests with the aforementioned enzymes confirmed map orientation and enabled small fragments of the λ clone to be identified and mapped, completing the physical map of λJD12 for the restriction enzymes *Bam*HI, *Hin*dIII and *Sst*I (Figure 10).

3.3.3 Identification of *Hmg* in Lp19 Genomic DNA

To demonstrate that regions hybridising to Lp19 *hmg* identified in the λ clone were present in Lp19 genomic DNA, a Southern blot (Section 2.14.1) containing total DNA from Lp19, Lp1, E8 and λJD12 digested (Figure 11A, Section 2.9) with the restriction enzymes *Sst*I and *Hin*dIII, was probed with the [α-^{32}P]dCTP-labelled (Section 2.14.2-2.14.3) λJD12 0.8 kb *Sst*I fragment (Figure 11B). The 0.8 kb λJD12 fragment hybridised to the corresponding 0.8 kb fragment in each of the *Sst*I digests. The fragments hybridising within the λJD12 digests aligned with the those of Lp19, demonstrating that Lp19 possesses genomic fragments corresponding to those within the λ clone. The Lp1 *Hin*dIII genomic digest contained hybridising fragments of approximately 5.2 kb, 5.7 kb, 7.9 kb and one greater than 10 kb. The fragments hybridising within the *Hin*dIII digests of E8 (approximately 7.9 and 5.7 kb) and the Lp19 (approximately 5.2 kb and one greater than 10 kb) aligned with the hybridising fragments of the Lp1 digest. Lp1 is considered to be an interspecific hybrid of the *N. lolii* and *E. typhina* genomes (Schardl *et al.* 1994) and this blot provides evidence to support that conclusion.

3.4 SEQUENCE ANALYSIS OF Lp19 *HMG*

3.4.1 Sequencing of Lp19 *hmg*

To define the boundaries of *hmg* within the λ clone the end of each cloned *Sst*I fragment (Section 3.3.3.2) was sequenced using the Amplicycle Sequencing kit (Section 2.13.1) with the pUC/M13 "universal" forward primer (Perkin Elmer, Table 2). Sequence similarity searches were performed using the Molecular Biology Computational Resource (MBCR) database, and it was deduced that Lp19 *hmg* spanned the 4.9 kb, 0.8 kb and 5.2 kb *Sst*I fragments of λJD12 (Figure 10). The strategy employed to sequence the Lp19 *hmg* was that of primer walking. This was used to extend the sequence from the ends of pJD6, pJD12, pJD20 and pJD48. Approximately 600 bp of sequence was

Figure 11 Demonstration that insert DNA present in λ clones corresponds to regions of the Lp19 genome.

Figure 11A Southern blot of *Sst*I and *Hin*dIII digested Lp1, E8, Lp19 and λJD12 DNA.

Lp1, E8 and Lp19 genomic DNA (2 μg) and λJD12 (approximately 10 ng), were digested with *Sst*I and *Hin*dIII. Lanes 2-9 contain *Sst*I and *Hin*dIII digests of Lp1 (lanes 2 and 3), E8 (lanes 4 and 5), Lp19 (lanes 6 and 7) and λJD12 (lanes 8 and 9). Lane 1 and 10 contain the BRL 1 kb ladder and *Hin*dIII digested λ DNA respectively.

Figure 11B Demonstration that λJD12 contains the *hmg* gene of Lp19.

Autoradiograph of a Southern blot of Figure 11A probed with the 0.8 kb *Sst*I fragment of λJD12.

Figure 11A

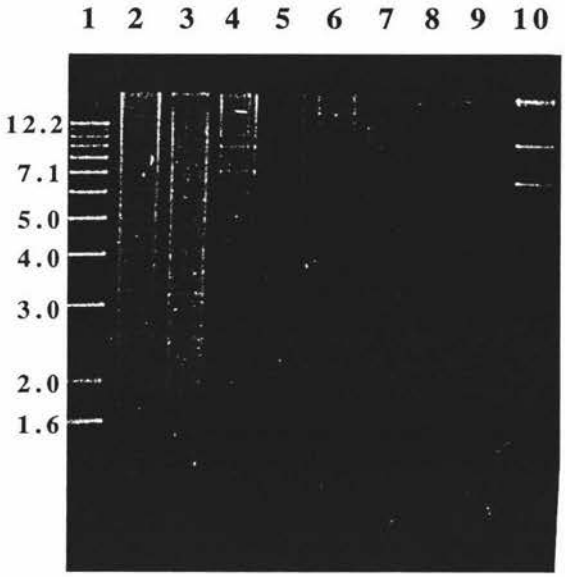


Figure 11B

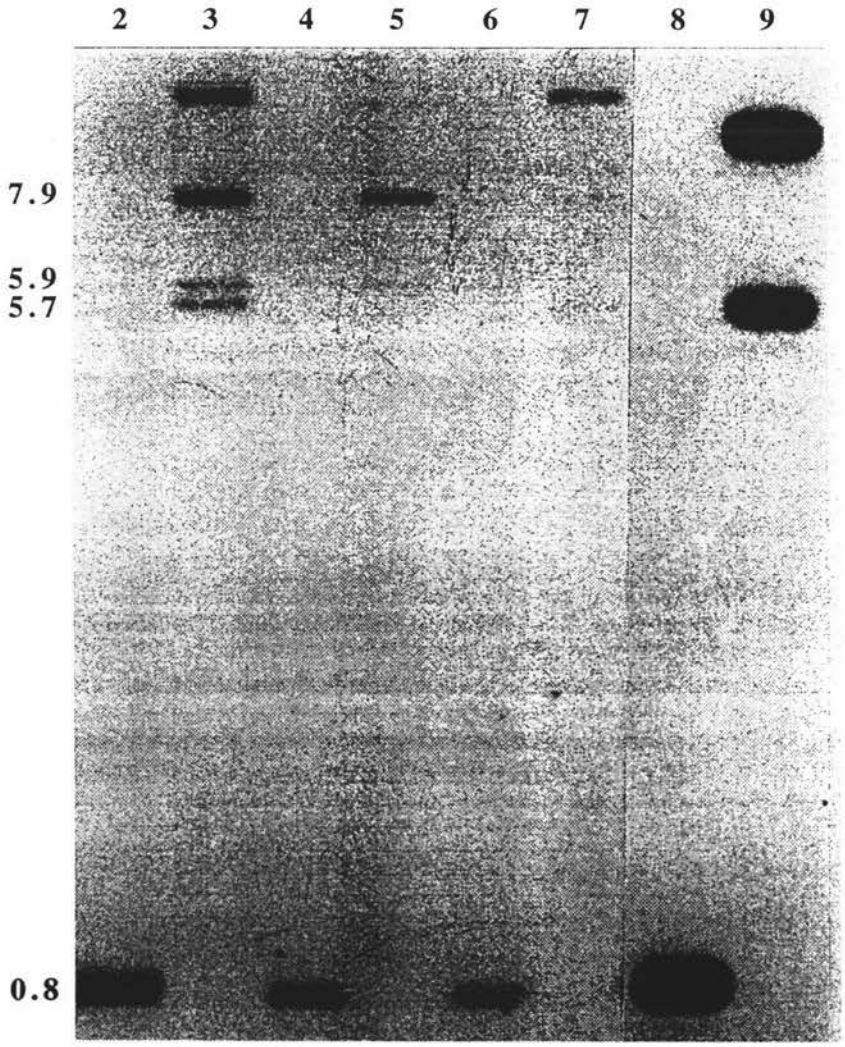
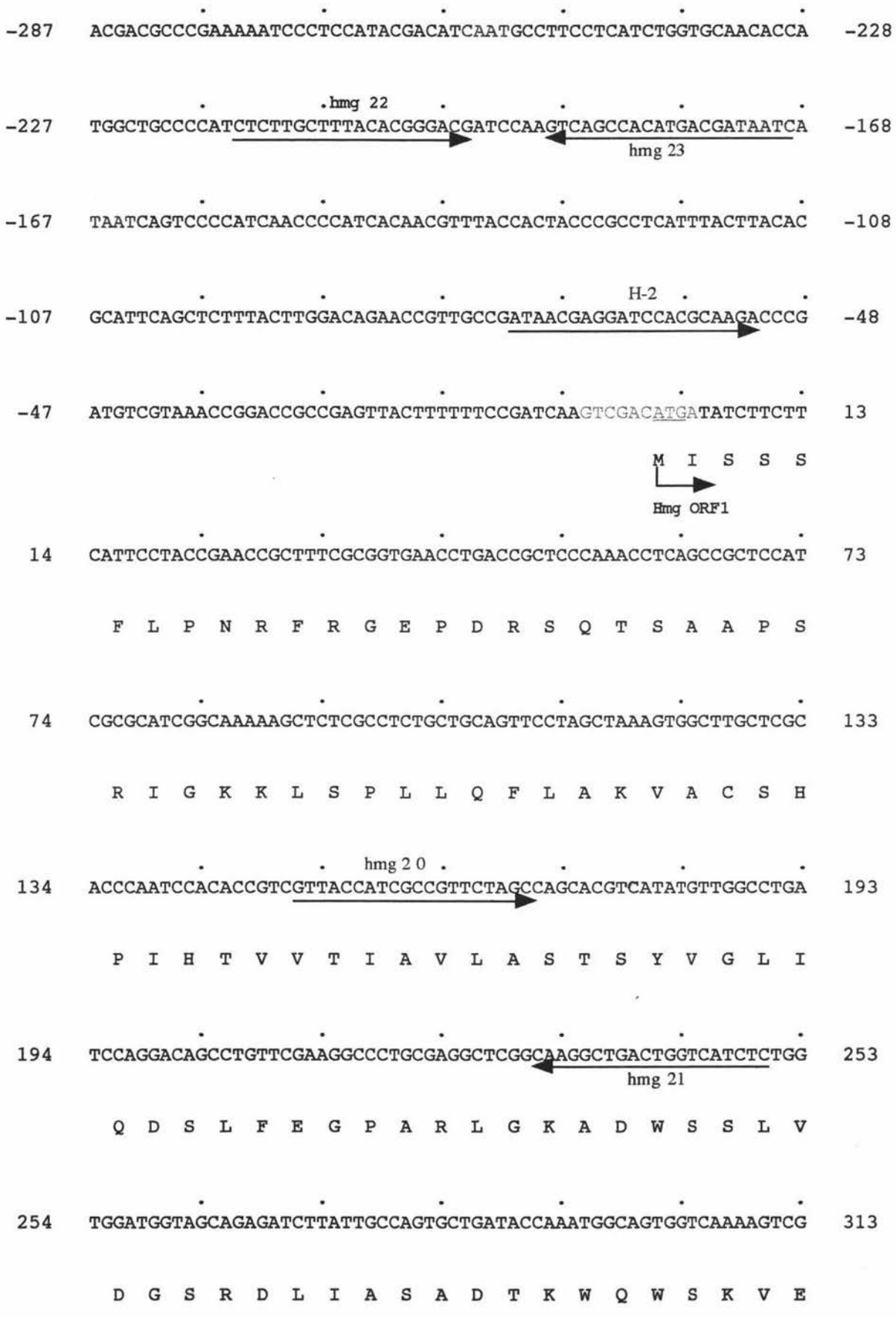
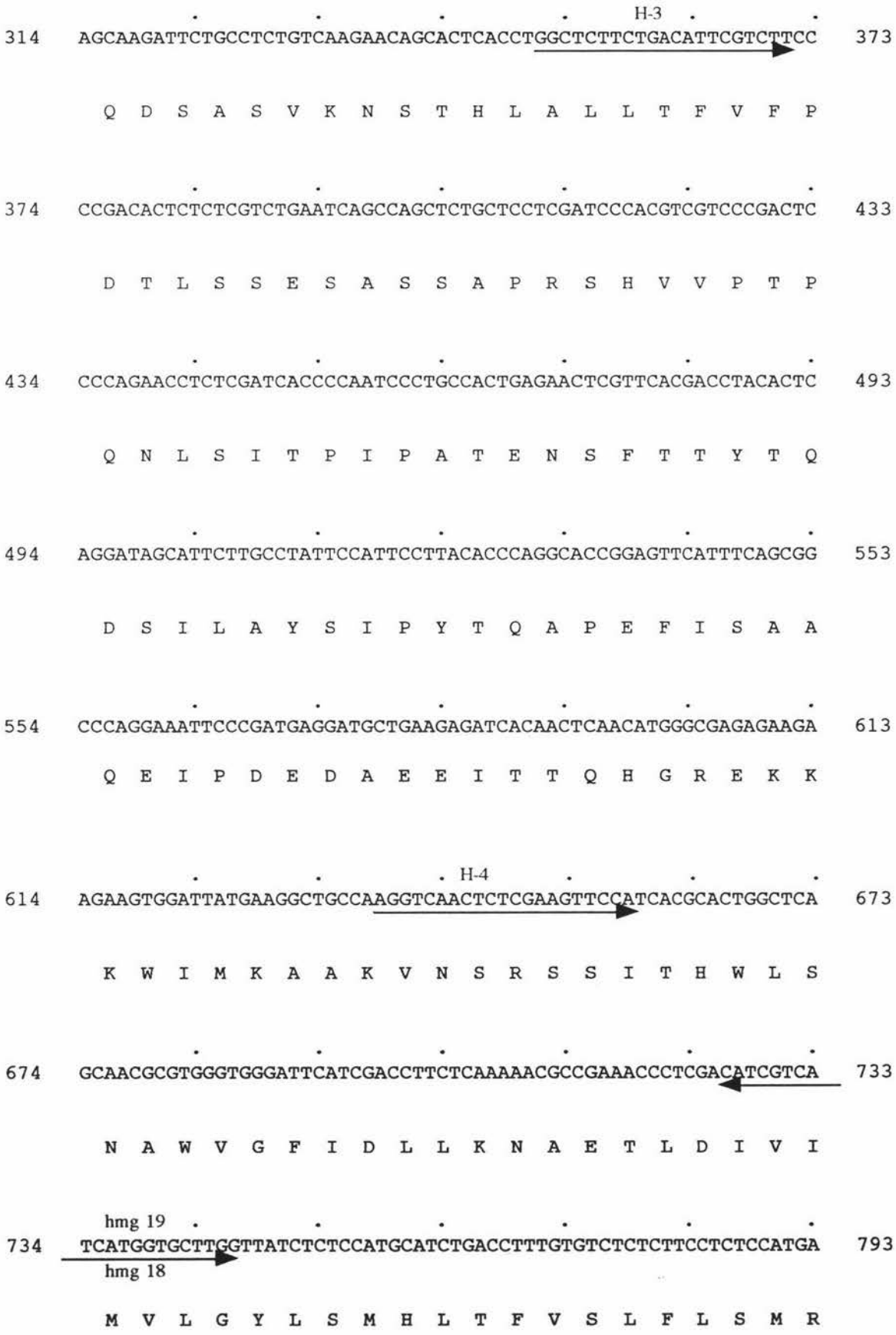


Figure 12 DNA sequence and deduced amino acid sequence of *hmg* from Lp19.

The 4,464 bp Lp19 *hmg* sequence shown contains a portion of the 5' non-coding region (⁻287-⁻1), the *hmg* coding sequence (1-3637) and 3' non-coding sequence (3638-4177). The coding sequence is interrupted by an intron (3186-3258) and the intron consensus sequences are shown (boxed, Section 3.4.2). The coding sequence begins at an ATG (purple and underlined) and ends at TAA (pink and underlined). A poly(A) termination sequence, AATAAA (pink and double underlined) occurs 42 bases downstream of the STOP codon (pink). The transcription start site is thought to occur between primers hmg 22 and H-2 and this is supported by the presence of putative CAAT (red) and TATA (green) consensus sequences. A microsatellite $(TGTC)_{18}$ occurs within the 3' non-coding region (blue). Primers used for PCR and sequencing are shown, as are the *Sst*I restriction sites (double underlined).


-287 ACGACGCCCGAAAAATCCCTCCATACGACATCAATGCCTTCCTCATCTGGTGCAACACCA -228
hmg 22
-227 TGGCTGCCCCATCTCTTGCTTTACACGGGACGATCCAAGTCAGCCACATGACGATAATCA -168
hmg 23
-167 TAATCAGTCCCCATCAACCCCATCACAACGTTTACCACTACCCGCCTCATTTACTTACAC -108
H-2
-107 GCATTCAGCTCTTTACTTGGACAGAACCGTTGCCGATAACGAGGATCCACGCAAGACCCG -48
-47 ATGTCGTAAACCGGACCGCCGAGTTACTTTTTTCCGATCAAGTCGACATGATATCTTCTT 13
M I S S S
Hmg ORF1
14 CATTCCTACCGAACCGCTTTCGCGGTGAACCTGACCGCTCCCAAACCTCAGCCGCTCCAT 73
F L P N R F R G E P D R S Q T S A A P S
74 CGCGCATCGGCAAAAAGCTCTCGCCTCTGCTGCAGTTCCTAGCTAAAGTGGCTTGCTCGC 133
R I G K K L S P L L Q F L A K V A C S H
hmg 2 0
134 ACCCAATCCACACCGTCGTTACCATCGCCGTTCTAGCCAGCACGTCATATGTTGGCCTGA 193
P I H T V V T I A V L A S T S Y V G L I
194 TCCAGGACAGCCTGTTCGAAGGCCCTGCGAGGCTCGGCAAGGCTGACTGGTCATCTCTGG 253
hmg 21
Q D S L F E G P A R L G K A D W S S L V
254 TGGATGGTAGCAGAGATCTTATTGCCAGTGCTGATACCAAATGGCAGTGGTCAAAAGTCG 313
D G S R D L I A S A D T K W Q W S K V E


H-3
314 AGCAAGATTCTGCCTCTGTCAAGAACAGCACTCACCTGGCTCTTCTGACATTCGTCTTCC 373
Q D S A S V K N S T H L A L L T F V F P
374 CCGACACTCTCTCGTCTGAATCAGCCAGCTCTGCTCCTCGATCCCACGTCGTCCCGACTC 433
D T L S S E S A S S A P R S H V V P T P
434 CCCAGAACCTCTCGATCACCCCAATCCCTGCCACTGAGAACTCGTTCACGACCTACACTC 493
Q N L S I T P I P A T E N S F T T Y T Q
494 AGGATAGCATTCTTGCCTATTCCATTCCTTACACCCAGGCACCGGAGTTCATTTCAGCGG 553
D S I L A Y S I P Y T Q A P E F I S A A
554 CCCAGGAAATTCCCGATGAGGATGCTGAAGAGATCACAACTCAACATGGGCGAGAGAAGA 613
Q E I P D E D A E E I T T Q H G R E K K
H-4
614 AGAAGTGGATTATGAAGGCTGCCAAGGTCAACTCTCGAAGTTCCATCACGCACTGGCTCA 673
K W I M K A A K V N S R S S I T H W L S
674 GCAACGCGTGGGTGGGATTCATCGACCTTCTCAAAAACGCCGAAACCCTCGACATCGTCA 733
N A W V G F I D L L K N A E T L D I V I
hmg 19
734 TCATGGTGCTTGGTTATCTCTCCATGCATCTGACCTTTGTGTCTCTCTTCCTCTCCATGA 793
hmg 18
M V L G Y L S M H L T F V S L F L S M R

```
                .         .         .         .         .         .
794   GACGAATGGGCTCCAACTTCTGGCTAGGCACGAGCACCCTTTTTTCGTCTGTCTTTGCCT   853

        R   M   G   S   N   F   W   L   G   T   S   T   L   F   S   S   V   F   A   F

                .         .         .         .         .         .
854   TTCTTTTCGGTCTAGCCGTGACCACTAAGCTCGGCGTACCGATCAGTGTTATTCTCTTGT   913

        L   F   G   L   A   V   T   T   K   L   G   V   P   I   S   V   I   L   L   S

                .         .         .         .         .         .
914   CCGAGGGCCTGCCGTTCCTTGTTGTAACAATTGGTTTTGAGAAGAACATTGTCCTCACCA   973

        E   G   L   P   F   L   V   V   T   I   G   F   E   K   N   I   V   L   T   R

                .         .         .         .         .         .
974   GAGCTGTCCTCAGCCATGCCGTTGAGCATCGTCGAACGCAAGGAGGCCGTGAGGTCCAGC   1033

        A   V   L   S   H   A   V   E   H   R   R   T   Q   G   G   R   E   V   Q   P

                .         .         .         .         .         .
1034  CCGGCAACAAGTCCGGTGGGGACAAATCACAAAACATCATTTCCTACGCCATTCAAGCTG   1093

        G   N   K   S   G   G   D   K   S   Q   N   I   I   S   Y   A   I   Q   A   A

                .         .         .         .         .         .
1094  CCATCAAGGACAAGGGTTATGAAATTCTGCGTGACTATGCCATCGAAATCTTGATTCTAT   1153
                                                                  <----------

        I   K   D   K   G   Y   E   I   L   R   D   Y   A   I   E   I   L   I   L   S

          hmg 14  .         .         .         .         .         .
1154  CGTTGGGCGCAGCCTCGGGTGTTCAGGGAGGTCTCCAGCAGTTCTGCTTCTTGGCTGCGT   1213
      ---------->
          hmg 13

        L   G   A   A   S   G   V   Q   G   G   L   Q   Q   F   C   F   L   A   A   W

                .         .         .         .         .         .
1214  GGATCTTGTTCTTTGACTGTATTCTCTTGTTTACCTTCTACACTGCAATCCTCAGTATCA   1273

        I   L   F   F   D   C   I   L   L   F   T   F   Y   T   A   I   L   S   I   K
```

1274 AACTCGAAATCAACCGCATCAAGCGACACTACGAGATGCGCATGGCCCTCGAAGCAGATG 1333

L E I N R I K R H Y E M R M A L E A D G

1334 GTGTTAGTCGGCGCGTAGCCGAGAAGGTGGCCAAGAGCAACGATGACTGGACCCAGTCCA 1393

V S R R V A E K V A K S N D D W T Q S S

hmg 26

1394 GTGGATCTGAATCCAAGAACACAACCTTGTTTGGTCGCATGCGGAGTAGCAGCGTTCCAA 1453

G S E S K N T T L F G R M R S S S V P K

1454 AGTTCAAGGTGCTTATGATTTCTGGGTTCGTTCTCATCAACGTAATCAACATTTGCACGA 1513

F K V L M I S G F V L I N V I N I C T I

1514 TCCCTTTCCGCAGCGCAAGCTCTCTATCAACCCTTCGATCGTGGGCTGGTGGTCTGGGCG 1573

P F R S A S S L S T L R S W A G G L G G

1574 GTGTTGTGTCTGCTCCTCCGGTTGACCCTGTCAAGGTTGCTGGCAAGGGTCTTGACGCCA 1633

V V S A P P V D P V K V A G K G L D A I

1634 TTCTGACCGCCGCCAAAGCAAGTGGGAAGGCAACCCTGGTCACTGTCCTTACACCCATCA 1693

L T A A K A S G K A T L V T V L T P I K

hmg 11

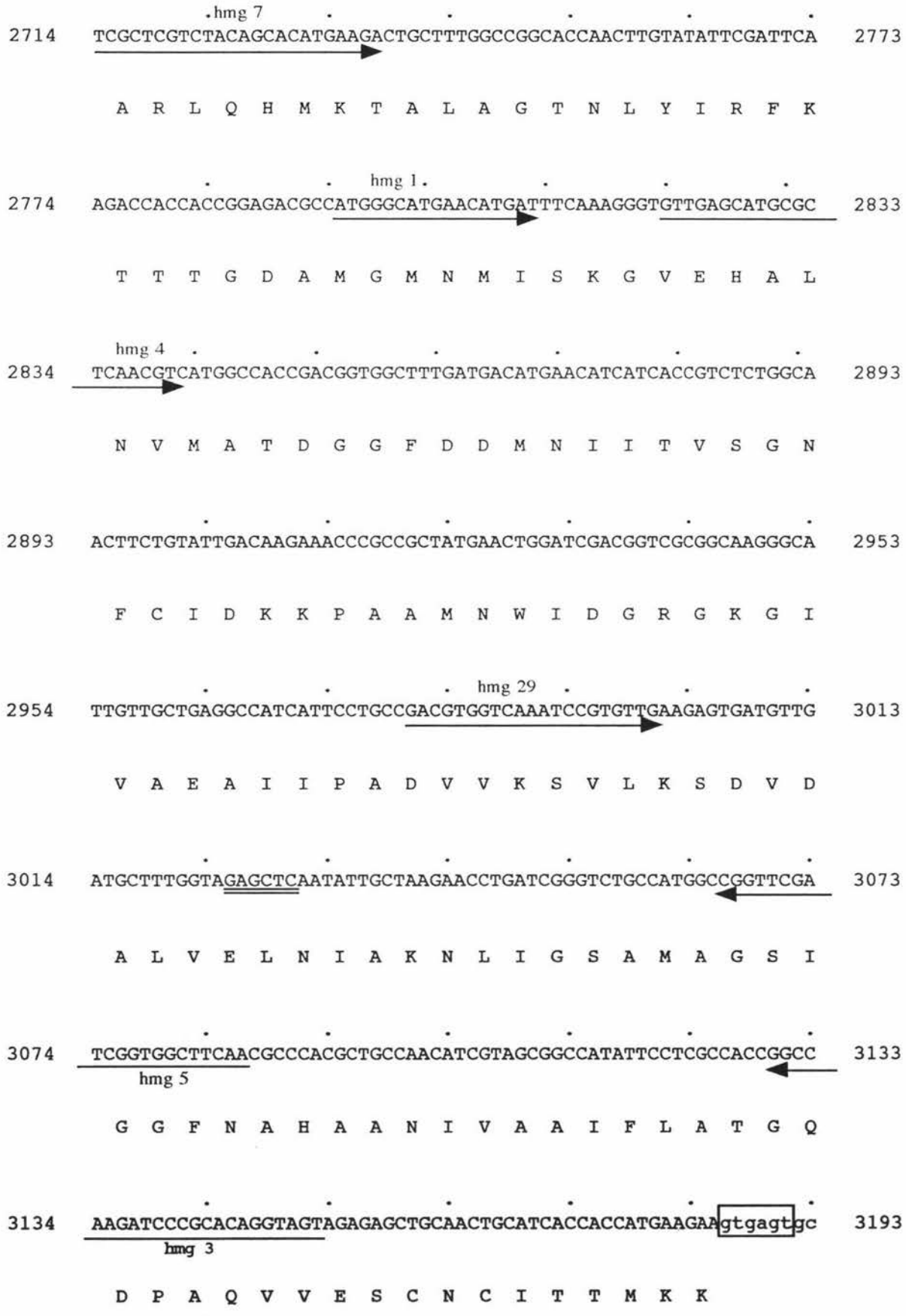
1694 AATACGAACTAGAGTACCCCTCTGTTCACTACGCGCTGTCATCCTCGTTGAGAGACGGCG 1753

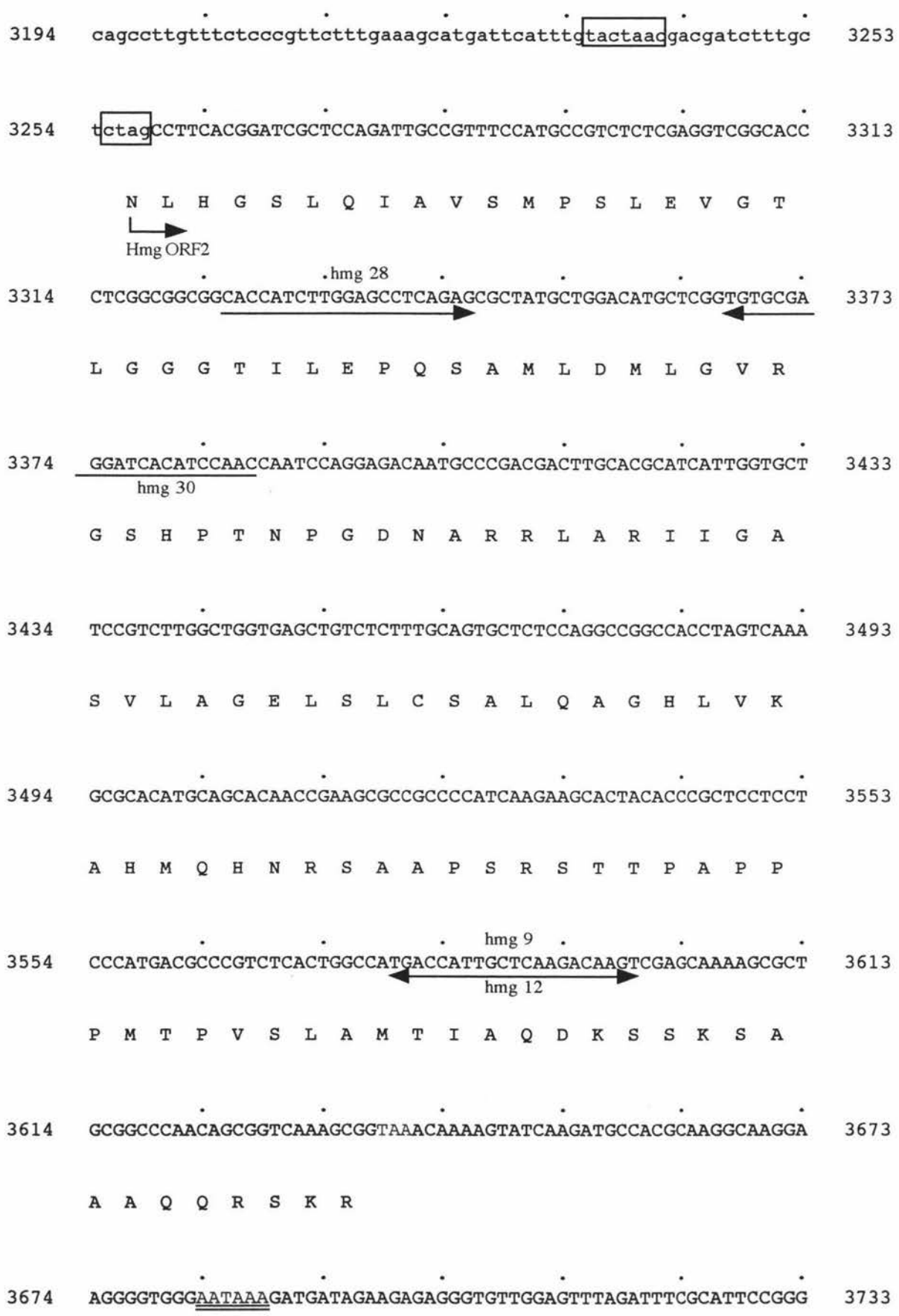
hmg 10

Y E L E Y P S V H Y A L S S S L R D G A

```
1754  CAGCCGGTGCCGTAAGTTCTGCCGTTCAGTTTGACAATTATGGGGTTGGTGGCCGAATGG  1813

        A  G  A  V  S  S  A  V  Q  F  D  N  Y  G  V  G  G  R  M  V

1814  TTGGGAGCCTCTTGAAGAGCCTCGAGGACCCCGTACTTTCAAAATGGATAGTCATCGCCC  1873

        G  S  L  L  K  S  L  E  D  P  V  L  S  K  W  I  V  I  A  L

1874  TGGCCCTAAGTGTTGGCCTCAATGGATACCTCTTCAATGTCGCTAGATGGAGCATCAAAG  1933

        A  L  S  V  G  L  N  G  Y  L  F  N  V  A  R  W  S  I  K  D

1934  ACCCTAACGTTCCGGACCACGGAATTGACCGCAAGGAGCTGGCCCGGGCTCAACGCTTCA  1993
                    <-------------------
                          hmg 27

        P  N  V  P  D  H  G  I  D  R  K  E  L  A  R  A  Q  R  F  N

1994  ACGAGACAGAATCCGCTACCCTACCCCTCGGCGAATATGTTCCCCCAACACCCAGTTGTA  2053

        E  T  E  S  A  T  L  P  L  G  E  Y  V  P  P  T  P  S  C  T

2054  CCGAGCCTGCTACGCCTGCATTGACTGATGACGAAGGTGATGGTCTGTCCATGTCAAAGC  2113

        E  P  A  T  P  A  L  T  D  D  E  G  D  G  L  S  M  S  K  L

2114  TCAGGTCCTCGCAATCTCGGTCTCAGTTTGAGCATCGATCAATTGAGGAACTGGAGAAGT  2173

        R  S  S  Q  S  R  S  Q  F  E  H  R  S  I  E  E  L  E  K  L

2174  TGATCGTGGAGAAGCGAACCCATGAGCTCAATGACGAGGAAGTTGTAACCATGTCTATGC  2233

        I  V  E  K  R  T  H  E  L  N  D  E  E  V  V  T  M  S  M  R
```

```
                .         .         .         .         .         .
2234  GTGGAAAGGTTCCCGGTTATGCACTCGAAAAGGCTCTGAAGGACTTTACTCGGGCCGTCA  2293

        G   K   V   P   G   Y   A   L   E   K   A   L   K   D   F   T   R   A   V   K

                .         .         .         .         .         .
2294  AGATCCGTCGTACTATCATTTCTCGCACCAAAGCCACTTCCGAAATCACCAATGGACTTG  2353

        I   R   R   T   I   I   S   R   T   K   A   T   S   E   I   T   N   G   L   D

                .         .         .         .         .         .
2354  ACCGATCCAAGCTTCCCTTCGAAAACTACAACTGGGAAAGGGTGTTTGGTGCCTGCTGCG  2413

        R   S   K   L   P   F   E   N   Y   N   W   E   R   V   F   G   A   C   C   E

                .         .         .         .         .         .
2414  AGAATGTCATCGGCTACCTGCCCCTTCCCGTTGGCGTCGCCGGCCCGCTCGTTATTGATG  2473

        N   V   I   G   Y   L   P   L   P   V   G   V   A   G   P   L   V   I   D   G

                .         .         .         .         .         .
2474  GGCAGAGCTATTTCATTCCTATGGCCACGACGGAAGGTGTCTTGGTTGCCAGCACTAGTC  2533

        Q   S   Y   F   I   P   M   A   T   T   E   G   V   L   V   A   S   T   S   R

                .         .         .         .         .         .
2534  GTGGCTGCAAAGCGATCAACTCTGGTGGCGGTGCCATTACTGTCTTAACAAGCGATGGTA  2593
      ------------------
          hmg 8

        G   C   K   A   I   N   S   G   G   G   A   I   T   V   L   T   S   D   G   M

                .         .         .         .         .         .
2594  TGACCCGTGGACCTTGCGTTAGTTTTGAGACTTTGGAGCGTGCCGGTGCCGCCAAGCTCT  2653

        T   R   G   P   C   V   S   F   E   T   L   E   R   A   G   A   A   K   L   W

                .         .         .         .         .         .
2654  GGCTGGATTCTGAGGCAGGCCAGAACACCATGAAGAAGGCTTTCAACTCGACCAGTCGGT  2713

        L   D   S   E   A   G   Q   N   T   M   K   K   A   F   N   S   T   S   R   F
```


hmg 7
2714 TCGCTCGTCTACAGCACATGAAGACTGCTTTGGCCGGCACCAACTTGTATATTCGATTCA 2773
A R L Q H M K T A L A G T N L Y I R F K
hmg 1
2774 AGACCACCACCGGAGACGCCATGGGCATGAACATGATTTCAAAGGGTGTTGAGCATGCGC 2833
T T T G D A M G M N M I S K G V E H A L
hmg 4
2834 TCAACGTCATGGCCACCGACGGTGGCTTTGATGACATGAACATCATCACCGTCTCTGGCA 2893
N V M A T D G G F D D M N I I T V S G N
2893 ACTTCTGTATTGACAAGAAACCCGCCGCTATGAACTGGATCGACGGTCGCGGCAAGGGCA 2953
F C I D K K P A A M N W I D G R G K G I
hmg 29
2954 TTGTTGCTGAGGCCATCATTCCTGCCGACGTGGTCAAATCCGTGTTGAAGAGTGATGTTG 3013
V A E A I I P A D V V K S V L K S D V D
3014 ATGCTTTGGTAGAGCTCAATATTGCTAAGAACCTGATCGGGTCTGCCATGGCCGGTTCGA 3073
A L V E L N I A K N L I G S A M A G S I
3074 TCGGTGGCTTCAACGCCCACGCTGCCAACATCGTAGCGGCCATATTCCTCGCCACCGGCC 3133
hmg 5
G G F N A H A A N I V A A I F L A T G Q
3134 AAGATCCCGCACAGGTAGTAGAGAGCTGCAACTGCATCACCACCATGAAGAAgtgagtgc 3193
hmg 3
D P A Q V V E S C N C I T T M K K


3194 cagccttgtttctcccgttctttgaaagcatgattcatttgtactaacgacgatctttgc 3253
3254 tctagCCTTCACGGATCGCTCCAGATTGCCGTTTCCATGCCGTCTCTCGAGGTCGGCACC 3313
N L H G S L Q I A V S M P S L E V G T
Hmg ORF2
hmg 28
3314 CTCGGCGGCGGCACCATCTTGGAGCCTCAGAGCGCTATGCTGGACATGCTCGGTGTGCGA 3373
L G G G T I L E P Q S A M L D M L G V R
3374 GGATCACATCCAACCAATCCAGGAGACAATGCCCGACGACTTGCACGCATCATTGGTGCT 3433
hmg 30
G S H P T N P G D N A R R L A R I I G A
3434 TCCGTCTTGGCTGGTGAGCTGTCTCTTTGCAGTGCTCTCCAGGCCGGCCACCTAGTCAAA 3493
S V L A G E L S L C S A L Q A G H L V K
3494 GCGCACATGCAGCACAACCGAAGCGCCGCCCCATCAAGAAGCACTACACCCGCTCCTCCT 3553
A H M Q H N R S A A P S R S T T P A P P
hmg 9
3554 CCCATGACGCCCGTCTCACTGGCCATGACCATTGCTCAAGACAAGTCGAGCAAAAGCGCT 3613
hmg 12
P M T P V S L A M T I A Q D K S S K S A
3614 GCGGCCCAACAGCGGTCAAAGCGGTAAACAAAAGTATCAAGATGCCACGCAAGGCAAGGA 3673
A A Q Q R S K R
3674 AGGGGTGGGAATAAAGATGATAGAAGAGAGGGTGTTGGAGTTTAGATTTCGCATTCCGGG 3733

```
3734  AGTTCATTATTGACTACGCGCAGGGCCTGACATACAATAGCACCTACCTATATAGACGAT  3793

3794  ATACTGCTACATATAAGGCCTGAGGGTGGGTTTCCAGAGATCTTAGACCTCCACGGTCAT  3853

3854  TATAATCGTTACTTTGTTTCACTGCTACAATTCTGTCCAGCTTCGATATCTCCTGTACAT  3913

3914  GTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT  3973

3974  GTCTGTCTGTCCATCTGTAACTGTCCCTCACACCAGCGCCTTGACTCGTGCAATCTCTTG  4033

4034  TCCATCCATGGAAATCATTTTGACACCCTTGTTCATAATCTCTCGTCTGCATTCATCCCA  4093

4094  CTGCTCCGGCATGACAGGCTTTGTAGCCTCCTCGACAATGTACGACCTAAAACCTTGCTC  4153

4154  GATGGCATGCTCCGCCGTAGCTTT  4177
```

obtained from each sequencing run (Section 2.13.2) and from this, primers were designed to extend the contig (Figure 10, Table 2). The sequence was assembled using the GELASSEMBLE program of the GCG package (Figure 12) and a consensus generated (see Appendix A 1.3 for PRETTYOUT).

To confirm that the 4.9 kb *Sst*I fragment was immediately adjacent to the 0.8 kb *Sst*I fragment the 6.5 kb *Bam*HI fragment of λJD12 was subcloned (Section 2.17) into pUC18 (see Appendix A 3.2), to generate pJD124. Plasmid DNA was isolated (Section 2.5.3) and sequenced (Section 2.13.1) with hmg 8 (Figure 10, Table 2) providing sequence information across the 4.9 kb and 0.8 kb fragment junction. The 0.8 kb and 5.2 kb *Sst*I junction occurred within the isolated PCR products (Section 3.1.3.2) and so had previously been sequenced (Figure 5).

Regions where ambiguities occurred within the consensus sequence of *hmg* were resequenced. A segment of Lp19 DNA containing sequencing errors was PCR amplified using primers hmg 26 and hmg 27 (Figure 10, Table 2) in standard Boehringer Mannheim buffer at an annealing temperature of 65°C. A product of 532 bp was amplified (Figure 14, lane 6) and sequenced (Section 2.13.2) using primer hmg 26 to remove any ambiguities A 6 frame translation of the nucleotide sequence was performed using the MAP program of the GCG package, identifying a number of open reading frames (ORF's, Figure 13). Two of these ORFs, ORF1 and ORF2, showed sequence similarity to the *Gibberella fujikuroi* Hmg ((Woitek *et al.* 1997), accession number X94307). The frame shift between these two ORFs indicated the presence of an intron.

3.4.2 Identification of Intronic region by RT-PCR Analysis

The putative intron was located by discontinuity of ORFs and by the presence of intron consensus sequences associated with the splicing reaction. The 5'-donor site (A-GTGAGT); the splice branch site (TACTAAC) and the 3' acceptor site (CTAG-C) (Figure 12) show similarities to consensus sequences determined for these sites in *Neurospora crassa* genes (Bruchez *et al.* 1993): 5' donor site (G-GTRMGY); the splice branch site (RCTRAC) and the 3' acceptor site (GWYAGG-G), where M=C or A, R=A or G, W=A or T and Y=C or T. The presence of stop codons in each of the three forward frames provided further evidence for this to be a non-coding region.

To confirm the presence of an intron between positions 3893 and 3965 in the nucleotide sequence (Figure 12), total RNA was isolated from Lp19 (Section 2.19) and analysed by RT-PCR (Section 2.22), using primers hmg 29 and hmg 30 (Table 2, Figure 10). These

Figure 13 A 6 frame translation of Lp19 *hmg* showing all open reading frames.

The FRAMES program of the GCG package plots the open reading frames of a nucleic acid sequence as boxes bordered by potential start and stop codons. Potential start codons are shown as short lines that extend above the box and potential stop codons are shown as short lines that extend below the box. By default, only the start and stop codons at the ends of open reading frames are shown in the frame display; if a stop codon has been passed, no stops are shown again until a start codon is passed; if a start codon is passed, no start codons are shown again until a stop codon is passed.

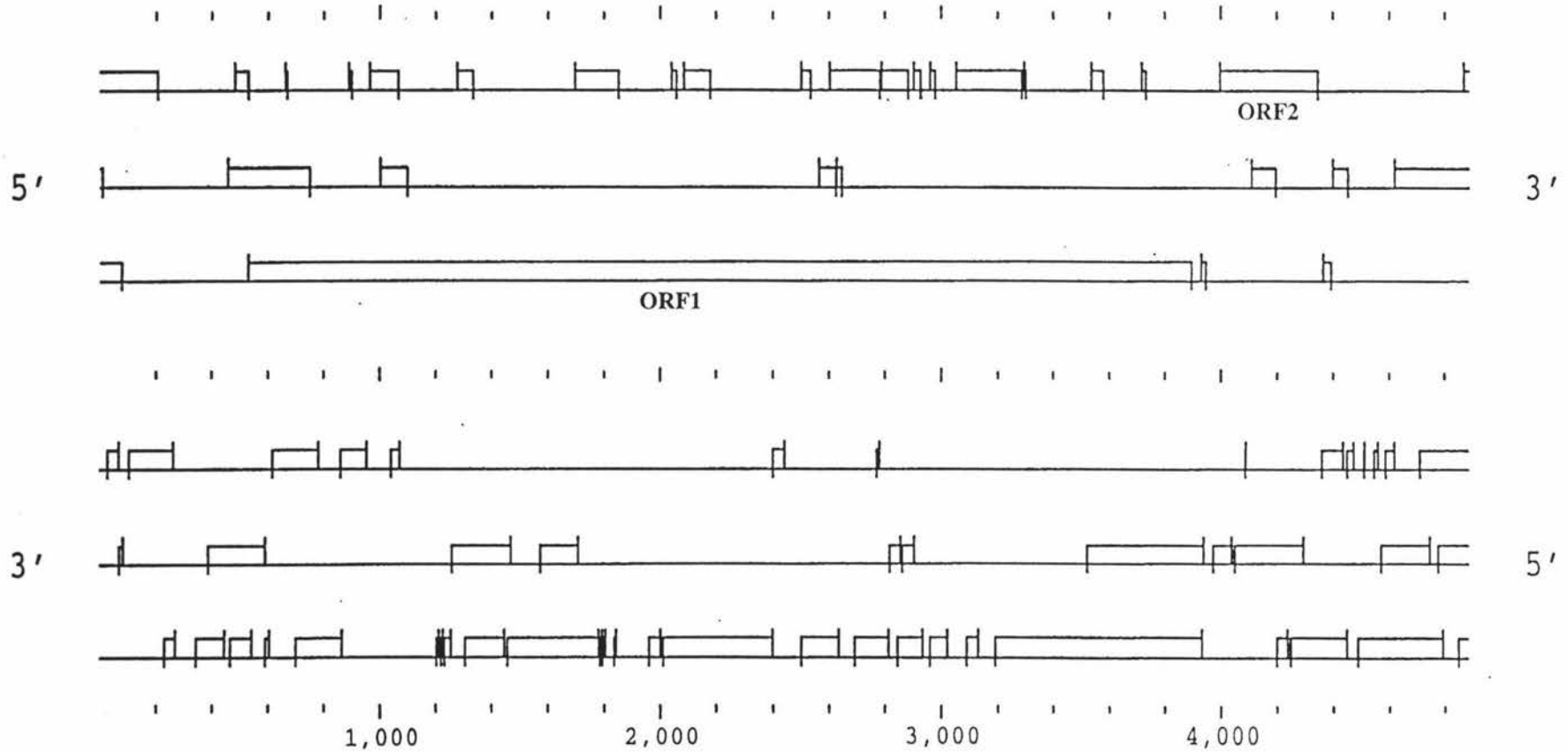
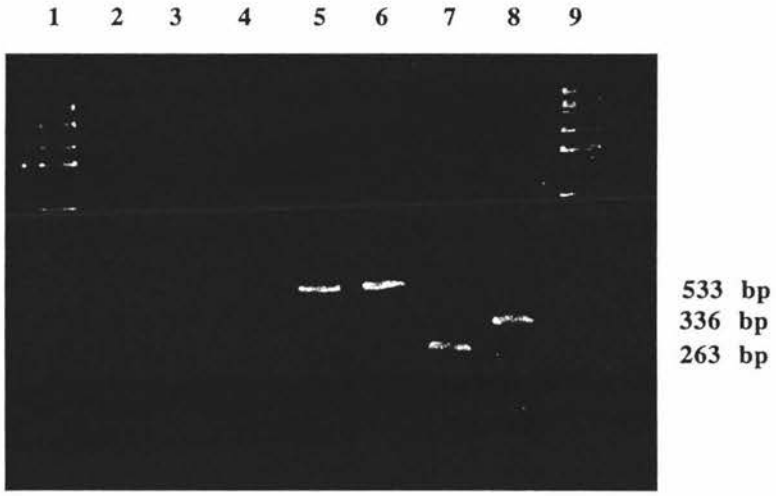

ORF2
5′
3′
ORF1
3′
5′
1,000
2,000
3,000
4,000

Figure 14 Identification of an intron in Lp19 *hmg*.

All reactions were amplified by PCR, with the following templates. Lanes: 2, No RNA; 3, No reverse transcriptase; 4, No DNA. Lane 5 and 6 contain Lp19 cDNA and DNA, respectively, amplified with primers hmg 26 and hmg 27. Lanes 7 and 8 contain Lp19 cDNA and DNA, respectively, amplified with hmg 29 and hmg 30. Lanes 1 and 9 contain the BRL 1 kb ladder



primers amplify a 408 bp fragment from Lp19 DNA (Figure 14). Lp19 cDNA was obtained using random primers (Gibco BRL, Section 2.22.1) and a PCR reaction (Section 2.22.2) was then carried out on 10 µl of Lp19 cDNA (Section 2.22.1) and 25 ng of Lp19 DNA at an annealing temperature of 50°C. Identical reactions were also carried out using primers hmg 26 and hmg 27 as there is no intron within this region. The PCR products were analysed by electrophoresis on a 2% NuSieve gel (Section 2.10) and the product obtained from Lp19 cDNA with primers hmg 29 and hmg 30 was observed to be approximately 70 bp smaller than that obtained from amplification of genomic DNA (Figure 14). The 336 bp RT-PCR product was sequenced (Section 2.13.2) using primer hmg 29 and found to lack a 73 bp region corresponding to the proposed intron.

To confirm that there were no other introns present in this gene, ORF1 (Figure 12) was analysed by RT-PCR. PCR reactions were carried out on Lp19 cDNA and DNA using primers which spanned ORF1: hmg 20 - hmg 10 and hmg 14 - hmg 5 (Figure 10, Table 2) and the products were analysed by gel electrophoresis (Figure 15). The primers hmg 20 and hmg 10 gave a 1561 bp PCR product from cDNA and DNA indicating the absence of any intron in this region. Primers hmg 14 and hmg 5 also gave a PCR product of the same size (1941 bp) when amplifying Lp19 cDNA and DNA, therefore confirming the absence of introns within this region.

3.4.3 Identification of Putative Transcription and Translation Start Sites

The location of the putative transcription start site was determined using RT-PCR (see Appendix 4.0 for methodology). PCR reactions were carried out on Lp19 cDNA with one fixed primer hmg 13 and five primers H-4, H-3, hmg 20, H-2 and hmg 22 at positions upstream. This showed the start site to be between hmg 22 and H-2 due to the presence of an RT-PCR product H-2 and an absence of product with hmg 22 (Figure 16). This was supported by the presence of putative CAAT and TATA consensus sequences within this region (Figure 12).

The proposed translation start site of Lp19 *hmg* ORF1 was identified by upstream discontinuity of the coding sequence and by the presence of specific nucleotides at the start as predicted by Kozak (Figure 12). The sequence around the putative start site GTCGAC**AUG**A, is similar to the vertebrate Kozak consensus sequence, GCCRMC**AUG**G (Kozak 1988), and to two proposed consensus sequences for *N. crassa*: CNNNCAM**U**A**UGGC** (Bruchez *et al.* 1993) and CAMMA **U** GGCU (Edelmann and Staben 1994) (where N=A,T,G or C; M=C or A, R=A or G and U= a

Figure 15 Confirmation that ORF1 lacks introns.

PCR amplification of Lp19 DNA and cDNA was carried out using primers hmg 20 and hmg 10 (lanes 3 and 4) and primers hmg 14 and hmg 5 (lanes 5 and 6). Lane 2 contained no DNA and lane 1 contained the BRL 1 kb ladder.

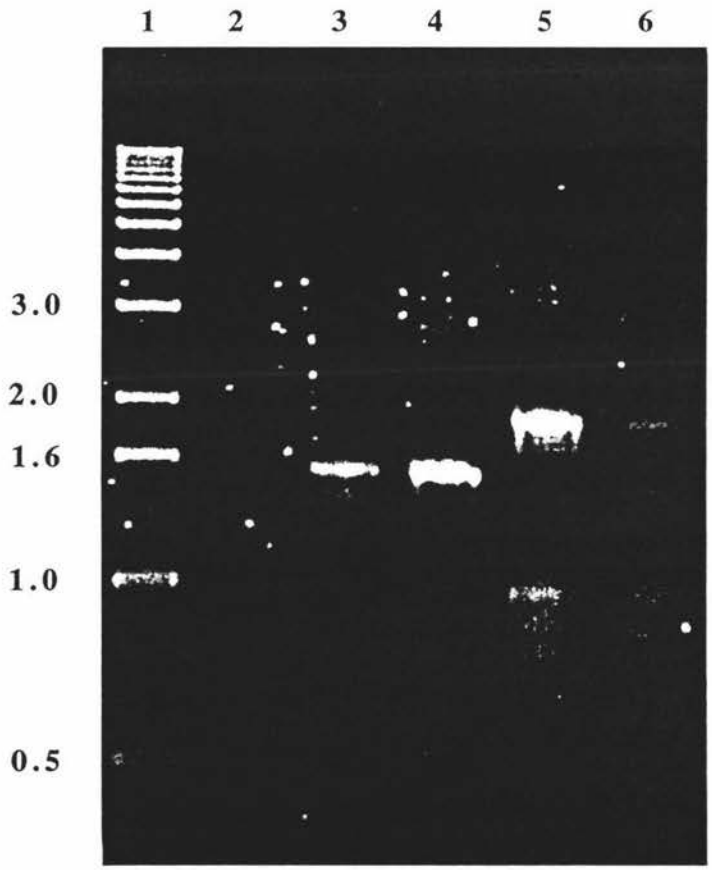
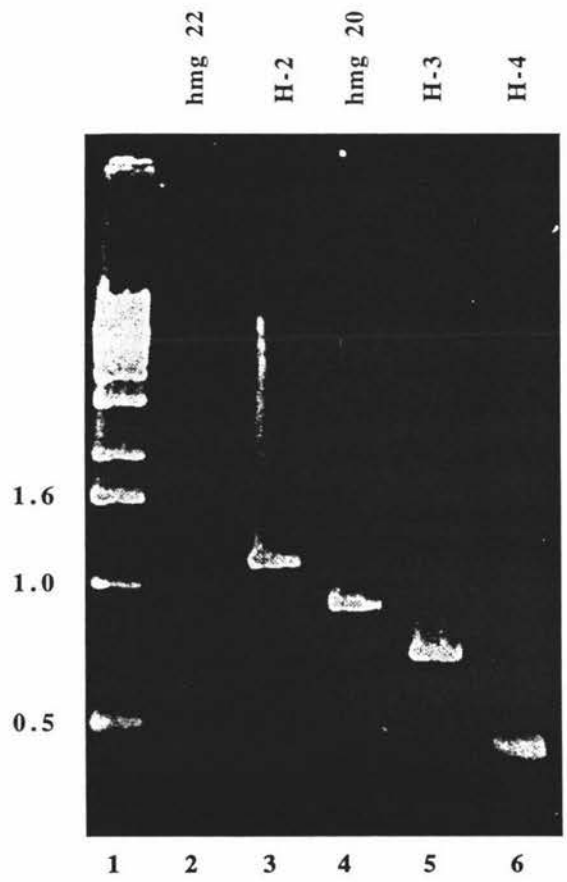


Figure 16 Identification of a putative transcription start site for *hmg*.

The location of the transcription start site was identified using RT-PCR with the fixed primer hmg 13 and five upstream primers (lanes 2-6). Lane 1 contains *Hin*dIII digested λ DNA.



conserved absence of that codon). Determining the position of the translation start proved difficult due to an absence of conservation in the 5' coding region of *hmg* genes.

3.4.4 Identification of the Poly(A) Termination Sequence

A putative polyadenylation signal (AAUAAA) was identified at position 3683 of Lp19 *hmg*, 43 bp downstream of the putative translation stop codon (Figure 12). RT-PCR was employed in an attempt to confirm this site, which is usually located 10 to 30 bp upstream of the poly(A) tail in mammals and yeast (Bruchez *et al.* 1993) which indicates the end of the mRNA

Lp19 cDNA was produced using A, G, and C anchored poly T primers in a single reaction (Section 2.22.1). PCR was then performed using primer hmg 28 (Figure 10, Table 2) in 3 separate reactions with each of the anchored primers (Section 2.22.2, Figure 17). A single RT-PCR product was not obtained in any of these reactions possibly due to the Tm differences of the primers (Table 2). A number of products were obtained two of which were more strongly amplified (approximately 357 bp and 575 bp). A putative poly(A) signal is located immediately upstream of the 3' end of the 357 bp sequence, but no such sequence could be identified adjacent to the end of the 575 bp fragment.

A microsatellite was located 304 bp downstream of the translation termination codon for Lp19 *hmg*. This consists of the four base sequence, TGTC, repeated 18 times within the 3' non-coding region of Lp19 Hmg (Figure 12).

3.4.5 Comparison of Lp19 *hmg* with Other *hmg* Genes

A total of 4464 bp of the *hmg* region of Lp19 was sequenced (Figure 13). Using the PILEUP program of the GCG package the deduced polypeptide sequence of the Lp19 Hmg was aligned with polypeptide sequences of *S. cerevisiae* HMG1 (Basson *et al.* 1986), *S. cerevisiae* HMG2 (Basson *et al.* 1986), *Schizosaccharomyces pombe* (Lum *et al.* 1996), *U. maydis* (Croxen *et al.* 1994), *G. fujikuroi* (Woitek *et al.* 1997) and human (Luskey and Stevens 1985) (Figure 18). The dendrogram from this comparison (Figure 19) agrees with the phylogeny of *hmg* published elsewhere (Bochar *et al.* 1997; Lum *et al.* 1996; Woitek *et al.* 1997), with the Hmg from Lp19 grouping closest to that from the ascomycete *G. fujikuroi*.

Figure 17 Identification of the 3' end of Lp19 *hmg* mRNA.

Lanes: 2, No DNA; 3, No RNA and 4, No reverse transcriptase. Lanes 5-7 contain RT-PCR products obtained with hmg 28 and: A-anchored poly T primer, lane 5; G-anchored poly T primer, lane 6 and C-anchored poly T primer, lane 7. These are flanked by the BRL 1 kb ladder.

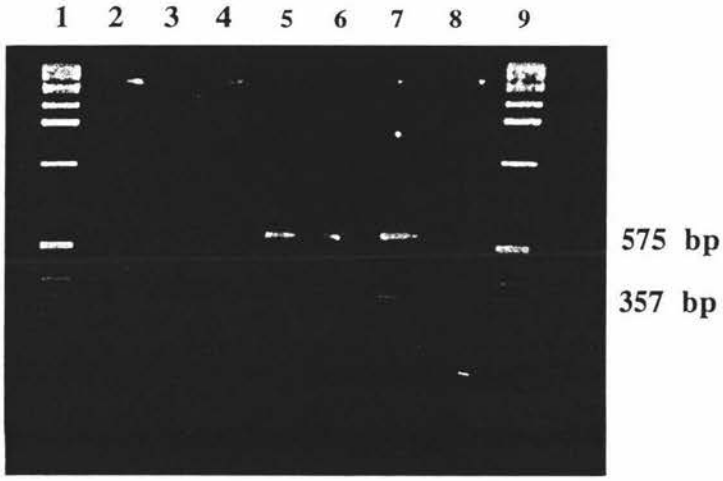


Figure 18 Alignment of deduced Hmg polypeptide sequences. Identity to Lp19 Hmg is shown in blue.

```
                1                                                   50
        Lp19    MISSSFLPNR FRGEPDRSQT SAAPSRIGKK LSPLLQFLAK VACSHPIHTV
  G.fujikuroi   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
S.cerevisiae2   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
S.cerevisiae1   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
      S.pombe   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~MIYK LAARYPIQVI
     U.maydis   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~MAYAVKAL

                51                                                  100
        Lp19    VTIAVLASTS YVGLIQDSLF EG.PARLGKA DWSSLVDGSR DLIASADTKW
  G.fujikuroi   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
S.cerevisiae2   ~~~~MSLPLK TIVHLVKPFA CTARFSARYP IHVIVWAVLL SAAAYLSVTQ
S.cerevisiae1   ~~~~MPPLFK GLKQMAKPIA YVSRFSAKRP IHIILFSLII SAFAYLSVIQ
      S.pombe   AIVGILVSMA YFSFLEALTQ EDFPVLIRAL KRFGILDGFP NTRLPNEMIL
     U.maydis   VIRFWALLKK ADSADIFVML SAYILMHGTF VNLFLSMRKF GSNFWLGASV

                101                                                 150
        Lp19    QWSKVEQDSA SVKNSTHLAL LTFVFPDTLS SESASSAPRS HVVPTPQNLS
  G.fujikuroi   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
S.cerevisiae2   SYLNEWKLDS NQ.YSTYLSI KPDELFEKCT HYYRSPVSDT WKLLSSKEAA
S.cerevisiae1   YYFNGWQLDS NSVFETAPNK DSNTLFQECS HYYRDSSLDG WVSITAHEAS
      S.pombe   KLSSVQGEDA SVWEQIPAAE LGGEGFVDFD ITQWYYPANA KVDVAQLVEP
     U.maydis   LMSSTFAFLL AITFASLLGV TVDPICLSEA LPFLVILVGF EKPYLLTRAI

                151                                                 200
        Lp19    ITPIPATENS FTTYTQDSIL AYSIPYTQAP EFISAAQEIP DEDAEEITIQ
  G.fujikuroi   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
S.cerevisiae2   DIYTPFHYYL STISFQSKDN STTLPSLDDV IYSVDHTRYL LSEEPKIPIE
S.cerevisiae1   ELPAPHHYYL LNLNFNSPNE TDSIPELANT VFEKDNTKYI LQEDLSVSKE
      S.pombe   YRNDCIFHDA SGACHFFFKE VGNWTVSSIA LPSNLANPPI DYFLDSSSIV
     U.maydis   FTHPDITPSA NSLSNRERSQ LILSNLDQAA KPEIGTLGKT SLAVANGTIA

                201                                                 250
        Lp19    HGREKKKWIM KAAKVNSRSS ITHWLSNAWV GFIDLLKNAE TLDIVIMVLG
  G.fujikuroi   ~~~~~~~MDH EGCQGQHPQQ CCQWVSNAWS EFLDLLKNAE TLDIVIMLLG
S.cerevisiae2   LVSENGTKWR LRNNSNFILD LHNIYRNMVK QFSNKTSEFD QFDLFIILAA
S.cerevisiae1   ISSTDGTKWR LRSDRKSLFD VKTLAYSLYD VFSENVTQAD PFDVLIMVTA
      S.pombe   IQRILPAIRE HGI...SWSW LLQLIARTWM NTLKIASQAS KTELLIVGTA
     U.maydis   MSREESFVRA LTERLKSEDS LRWAEPTPMP AHEIVVSAVN RVGVPIIRDY

                251                                                 300
        Lp19    YLSMHLTFVS LFLSMRRMGS NFWLGTSTLF SSVFAFLFGL AVTTK.LGVP
  G.fujikuroi   YIAMHLTFVS LFLSMRKMGS KFWLGICTLF SSVFAFLFGL VVTTK.LGVP
S.cerevisiae2   YLTLFYTLCC LFNDMRKIGS KFWLSFSALS NSACALYLSL YTTHSLLKKP
S.cerevisiae1   YLMMFYTIFG LFNDMRKTGS NFWLSASTVV NSASSLFLAL YVTQCILGKE
      S.pombe   YACMLISIVS LYLKMRRLGS KFWLFFSVLL STLFSVQFAM TLVRA.SGVR
     U.maydis   LIEIAVMAVG ATSGVSGLRE FCQLAALILL FDCIF.LFAF YVSILTVMVE
```

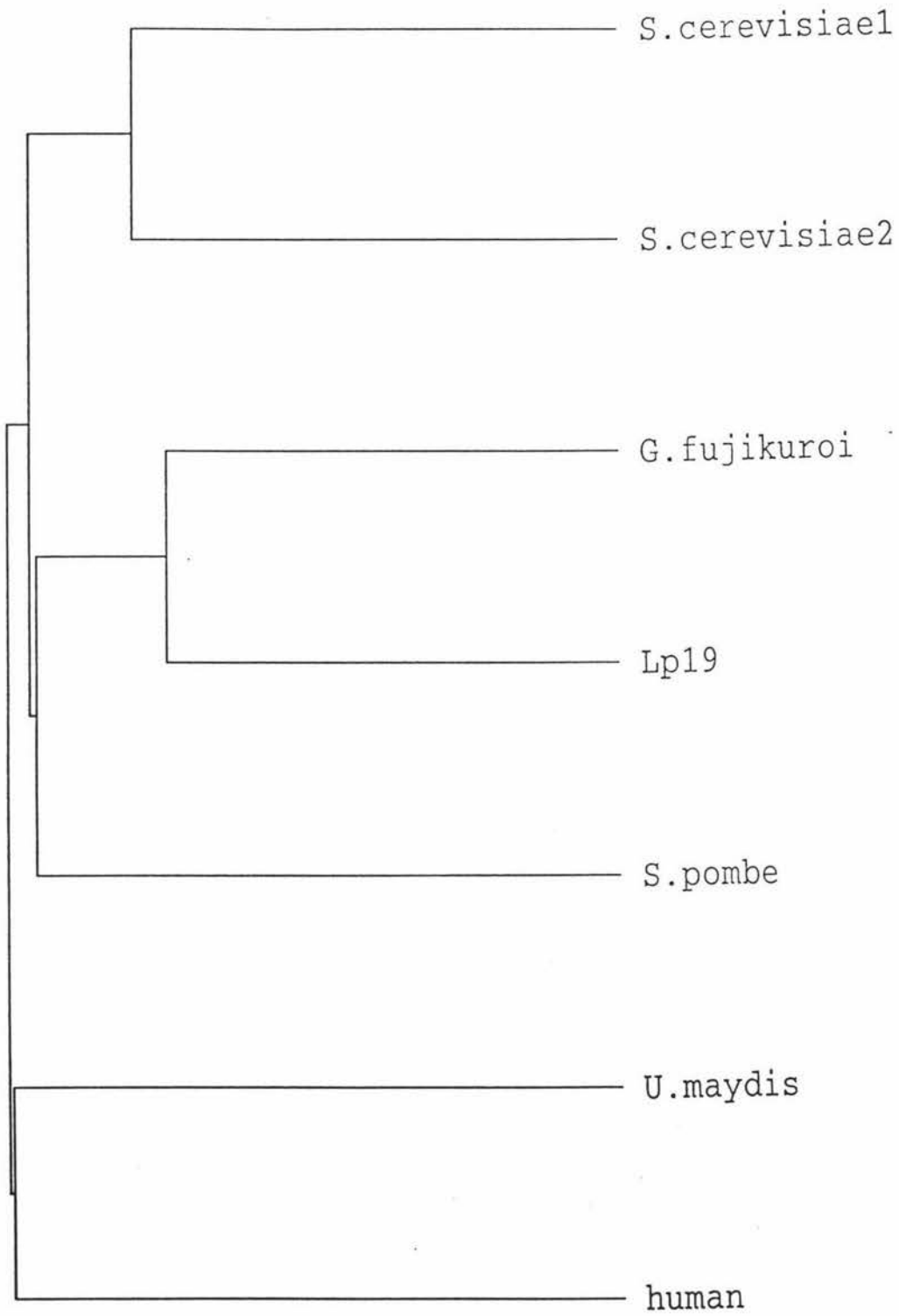
```
              301                                                350
Lp19                                        V    T  G.GREVQ  N
G.fujikuroi                          M    I    I  ..AQNSKS K
S.cerevisiae2 A LLS VI     I  I   K HKVR AAFS  Q...KFH IS I.DKKIT...
S.cerevisiae1 V ALT FE     I  VV  K HK KIAQYA  E...KFE VG L.SKRIT...
S.pombe         LVS I S     INVVALD  AAE   Q I T......... ..........
U.maydis      VHR KVKR V RKPK SAELC LDSDSAP SP ISDSSTS D  SPSPNSR FW

              351                                                400
Lp19            GGDK   I            D  Y           .   SL
G.fujikuroi   R PDGST  M  Q  V      E  F  I         .V  VI
S.cerevisiae2 .......... V NI YE MF QE AYLI    LFY SSFIGC AIYARHLP
S.cerevisiae1 .......... TDEIVFESVS EE GRLIQ H LLC FAFIGC SMYAHQLKT
S.pombe       RCSVSD HSP MHED AK CR NAAPP   HF SFG .VV AI ..F YCNF I
U.maydis       KVFSVILG. SARSEKRRKS ENPLGR KLL LVASF V HS FNLISTLTRQ

              401                                                450
Lp19                         C                         K HYE M MA  A
G.fujikuroi            T     F                R.STVSSVMS ICVWPLRMMA
S.cerevisiae2 VN  I STFM  V  LL SA    S    M      I H STV I QT  E
S.cerevisiae1 TN  I SAF    I EL  TP    S   ALR   M V H STI IKQT  E
S.pombe       K  .  F AV MIY LL   S  FV   TL    MR YNAKDD V KV IEE L
U.maydis      SAITRHH TY EPSPYQIVAN QSPIFAPAAS SAIAPLLAQL SQSQPDDLDL

              451                                                500
Lp19               K     S  DWTQ      ESKNTT     MR                 S
G.fujikuroi        N     GD ELNRVR  ....DAP     .K   I          L  IF
S.cerevisiae2 VPTT DIIY  DETASEPHF. .......... LRSNVAIILG  ASV GLLL
S.cerevisiae1 VPST RIIS  AEKKSVS F. .......... LNL VV IIM  LSV LL  F
S.pombe        EST RH  D G  SSATT A GSRYFKVRY  T.........  IILFI. IA
U.maydis      VA IS PIIL VLLESNEVVS  SRSTLANAA LNA GVYAQA TSTGVRPVPV

              501                                                550
Lp19            VI   T     S  L  L     GG     VS PP   V    GK
G.fujikuroi   V IV   S     NP  M  I  T  SS     IAPLS   F    SN
S.cerevisiae2   LYVFTDKL NATILNTVYF D TIYS P.. .......... ......NF N
S.cerevisiae1   FY FGANW VND FNSLYF DKERVS P.. .......... ...... F T
S.pombe       F LFEL S    KHYAAT AA AARLIP VRS QYPDFKSQRL LDDGVF DV
U.maydis      GRGSSSLAVL DSLM GWTVI VGDPVISKWM SL LGISIFL N YLLKGIAT

              551                                                600
Lp19          T  A G A   LVTVLTPIKY ELEYPSVHYA  S   RDG A GAV SAVQ D
G.fujikuroi   PT  SNNRP  LVTVLTPIKY ELEYPSIHYA  G AASNP Y NDAFHH.H Q
S.cerevisiae2 YKDIGNLSNQ VIISVL KQ  YTPLKKY QI ED V ..... ..........
S.cerevisiae1 SN SENF EQ AIVSV  LL  YKPIK YQRI EDMV ..... ..........
S.pombe       S ISSMSNIE SPS RL.... ...L A F G AEL STSFLS TIH FINNWS
U.maydis      GNAAL EGNA AGAAAYAAAR F...IGA LD SEAKAIEEHS ERNRRRWSAR

              601                                                650
Lp19          NYGVGGRMVG SLLKSLEDPV LSKWIVIALA LSVGLNGYLF NVARWSIKDP
G.fujikuroi   GYGVGGRMVG GILKSLEDPV LSKWIVIALA LSVALNGYLF NVARWGIKDP
S.cerevisiae2 ......LIID  VSNAIR QF I  LLFF F  V ISI VY L  A KIHTGYM
S.cerevisiae1 ......LLLR NVSVAIR RF V  LVLS  V C AVI V  L  A  IHTSYT
S.pombe       H ........ .....ISASF       C  S   IAV VF L  A  LN.SIK
U.maydis      PENGKPGVQT RGHSADGISD YHNRLAQIS  TQASVPLSKL TRDNDKLASK
```

```
                651                                                  700
        Lp19    NVPDHGIDRK ELARAQRFNE TESATLPLGE YVPPTPSCTE PATPALTDDE
  G.fujikuroi   NVPEHNIDRN ELARAREFND TGSATLPLGE YVPPTPMRTQ PSTPAITDDE
S.cerevisiae2   NFQPQSNKI. .DDLVVQQKS ATIEFSETRS MPASSGLETP VTAKDIIISE
S.cerevisiae1   ADQLVKTEVT KKSFTAPVQK ASTPVLTNKT VISGSKVKSL SSAQSSSSGP
      S.pombe   EEPEKKVVEK VV........ .........E VVKYIPSSNS SSIDDIQKDE
     U.maydis   PTSMTGAAVN AVPADAIKPD AMDAVKKHDV ....APSSDA DGAVATPGQQ

                701                                                  750
        Lp19    GDGLSMSKLR SSQSRSQFEH RSIEELEKLI VEKR..THEL NDEEVVTMSM
  G.fujikuroi   AEGLHMTKAR P....ANLPN RSNEELEKLL SENA..LREM TDEEVISLSM
S.cerevisiae2   EIQNNE.CVY ALSSQDE.PI RPLSNLVELM EKEQ..LKNM NNTEVSNLVV
S.cerevisiae1   SSSSEEDDSR DIESLDK.KI RPLEELEALL SSGN..TKQL KNKEVAALVI
      S.pombe   IAQESVV... .......... RSLEECITLY NNGQ..ISTL NDEEVVQLTL
     U.maydis   GQQICLQDAT YVTPDGETVV RPLEELVEIY AGG.AGVFFL TDEEIITLSQ

                751                                                  800
        Lp19    RGKVPGYALE KALKDFTRAV KIRRTIISRT KATSEITNGL DRSKLPFENY
  G.fujikuroi   RGKIPGYALE KTLGDFTRAV KIRRSIIARN KAAADITHSL DRSKLPYENY
S.cerevisiae2   NGKLPLYSLE KKLEDTTRAV LVRRKALSTL AESPI....L VSEKLPFRNY
S.cerevisiae1   HGKLPLYALE KKLGDTTRAV AVRRKALSIL AEAPV....L ASDRLPYKNY
      S.pombe   AKKIPLYALE RVLKDVTRAV VIRRTVVSRS SR....TKTL ESSNCPVYHY
     U.maydis   NGKIAAYALE KVLQDHERAV RVRRALVSRA S....ATQTL ETSLLPHRDY

                801                                                  850
        Lp19    NWERVFGACC ENVIGYLPLP VGVAGPLVID GQSYFIPMAT TEGVLVASTS
  G.fujikuroi   NWERFFGACC ENVIGYMPLP VGVAGPLVID GQSYFIPMAT TEGVLVASAS
S.cerevisiae2   DYDRVFGACC ENVIGYMPIP VGVIGPLIID GTSYHIPMAT TEGCLVASAM
S.cerevisiae1   DYDRVFGACC ENVIGYMPLP VGVIGPLVID GTSYHIPMAT TEGCLVASAM
      S.pombe   DYSRVLNACC ENVIGYMPLP LGVAGPLIID GKPFYIPMAT TEGALVASTM
     U.maydis   DYGKVMGACC ENVVGYMPIP LGIAGPLNID GQFMPIPMAT TEGTLVASTS

                851                                                  900
        Lp19    RGCKAINSGG GAITVLTSDG MTRGPCVSFE TLERAGAAKL WLDSEAGQNT
  G.fujikuroi   RGCKAINSGG GAITVLTADG MTRGPCVAFE TLERAGAAKL WLDSEAGQDM
S.cerevisiae2   RGCKAINAGG GATTVLTKDG MTRGPVVRFP TLIRSGACKI WLDSEEGQNS
S.cerevisiae1   RGCKAINAGG GATTVLTKDG MTRGPVVRFP TLKRSGACKI WLDSEEGQNA
      S.pombe   RGCKAINAGG GAVTVLTRDQ MSRGPCVAFP DLTRAGRAKI WLDSPEGQEV
     U.maydis   RGCKALNAGG GVTTVLTQDA MTRGPALEFP SVVQAAKAKR WIDSQEGAQT

                901                                                  950
        Lp19    MKKAFNSTSR FARLQHMKTA LAGTNLYIRF KTTTGDAMGM NMISKGVEHA
  G.fujikuroi   MKKAFNSTSR FARLQSMKTA LAGTNLYIRF KTTTGARMGM NMISKGVEHA
S.cerevisiae2   IKKAFNSTSR FARLQHIQTC LAGDLLFMRF RTTTGDAMGM NMISKGVEYS
S.cerevisiae1   IKKAFNSTSR FARLQHIQTC LAGDLLFMRF RTTTGDAMGM NMISKGVEYS
      S.pombe   MKKAFNSTSR FARLQHIKTA LAGTRLFIRF CTSTGDAMGM NMISKGVEHA
     U.maydis   IKAAFDSTSR FARLSSLRCV LAGRTLYVRF ATSTGDAMGM NMISKGVEKA

                951                                                 1000
        Lp19    LNVMATDGGF DDMNIITVSG NFCIDKKPAA MNWIDGRGKG IVAEAIIPAD
  G.fujikuroi   LSVMANDGGF DDMQIISVSG NYCTDKKAAA LNWIDGRGKG VVAEAIIPGE
S.cerevisiae2   LKQMVEEYGW EDMEVVSVSG NYCTDKKPAA INWIEGRGKS VVAEATIPGD
S.cerevisiae1   LKQMVEEYGW EDMEVVSVSG NYCTDKKPAA INWIEGRGKS VVAEATIPGD
      S.pombe   LVVMSNDAGF DDMQVISVSG NYCTDKKPAA INWIDGRGKS VIAEAIIPGD
     U.maydis   LGLMTEQY.F PEMKVLSLSG NYCTDKKPAA INWIEGRGKS VVAEAVVPGN
```

```
                1001                                                1050
          Lp19  VVKSVLKSDV DALVELNIAK NLIGSAMAGS IGGFNAHAAN IVAAIFLATG
  G.fujikuroi   VVRSVLKSDV DSLVELNVAK NLIGSAMAGS VGGFNAHAAN IVAAIFLATG
S.cerevisiae2   VVKSVLKSDV SALVELNISK NLVGSAMAGS VGGFNAHAAN LVTALFLALG
S.cerevisiae1   VVRKVLKSDV SALVELNIAK NLVGSAMAGS VGGFNAHAAN LVTAVFLALG
      S.pombe   AVKSVLKTTV EDLVKLNVDK NLIGSAMAGS VGGFNAHAAN IVTAVYLATG
     U.maydis   VVRSVLKCTV RDLVNLNTKK NLIGSAMAGS VGGFNAHAAN ILTAIYLATG

                1051                                                1100
          Lp19  QDPAQVVESC NCITTM.NL. ..HGSLQIAV SMPSLEVGTL GGGTILEPQS
  G.fujikuroi   QDPAQVVESA NCITIMKNL. ..NGALQISV SMPSLEVGTL GGGTILEPQG
S.cerevisiae2   QDPAQNVESS NCITLMKEV. ..DGDLRISV SMPSIEVGTI GGGTVLEPQG
S.cerevisiae1   QDPAQNVESS NCITLMKEV. ..DGDLRISV SMPSIEVGTI GGGTVLEPQG
      S.pombe   QDPAQNVESS NCITLMDNV. ..DGNLQLSV SMPSIEVGTI GGGTVLEPQG
     U.maydis   QDPAQNVESS NCITLMEAIN .DDEDLLITV SMPSIEVGTV GGGTVLPPQR

                1101                                                1150
          Lp19  AMLDMLGVRG SHPTNPGDNA RRLARIIGAS VLAGELSLCS ALQAGHLVKA
  G.fujikuroi   AMLDILGVRG SHPTNPGDNA RRLARIIGAA VLAGELSLCS ALAAGHLVRA
S.cerevisiae2   AMLDLLGVRG PHPTEPGANA RQLARIIACA VLAGELSLCS ALAAGHLVQS
S.cerevisiae1   AMLDLLGVRG PHATAPGTNA RQLARIVACA VLAGELSLCA ALAAGHLVQS
      S.pombe   AMLDLLGVRG AHMTSPGDNS RQLARVVAAA VMAGELSLCS ALASGHLVKS
     U.maydis   SMLEMMGIAG AHSTTPGANA QRLARIIAAS VMAGELSLMG ALCAGHLIQA

                1151                                                2000
          Lp19  HMQHNRSAAP SRSTTPAPPP MTPVSLAMTI AQDKSSKSAA AQQRSKR~~~
  G.fujikuroi   HMQHNRSAAP SRSTTPGSSH DARLTGHDQC PRALSVNNVD ERRRYSEVKA
S.cerevisiae2   HMTHNRKTNK ANELPQPSNK GPPCKTSALL ~~~~~~~~~~ ~~~~~~~~~~
S.cerevisiae1   HMTHNRKPAE PTKPNNLDAT DINRLKDGSV TCIKS~~~~~ ~~~~~~~~~~
      S.pombe   HIGLNRSALN TPAMDSSAKK PATDALKSVN SRVPGR~~~~ ~~~~~~~~~~
     U.maydis   HMKHNRSVPS TPGTMTPLPR PETPKHFAAS MTPLVAPSVG QSTRGSPTGT

                1201                1223
          Lp19  ~~~~~~~~~~ ~~~~~~~~~~ ~~~
  G.fujikuroi   IDE~~~~~~~ ~~~~~~~~~~ ~~~
S.cerevisiae2   ~~~~~~~~~~ ~~~~~~~~~~ ~~~
S.cerevisiae1   ~~~~~~~~~~ ~~~~~~~~~~ ~~~
      S.pombe   ~~~~~~~~~~ ~~~~~~~~~~ ~~~
     U.maydis   RTMSMTNLSI PASDSTGRAS SLT
```

Figure 19 Dendrogram showing relatedness of Lp19 Hmg to Hmgs of other organisms.

Dendrogram from PILEUP program of GCG showing clustering relationships used to determine the order of pairwise alignments that together create the final multiple sequence alignment (Figure 18). Distance along the vertical axis is proportional to the difference between sequences. This dendrogram is from a pairwise alignment using the UPGMA clustering strategy. Hence sequences and clusters of sequences occur in pairs.


S.cerevisiae1
S.cerevisiae2
G.fujikuroi
Lp19
S.pombe
U.maydis
human

3.4.6 Hmg Sequence Features

Hmg contains three domains, a catalytic domain (1-641), a linker region (642-735) and a membrane spanning domain (736-1188). These domains were determined by comparison to the domains of the HMG1 and HMG2 proteins of *S. cerevisiae* (Basson *et al.* 1986). The National Centre for Biotechnology Information (NCBI) program, TMPred, which determines possible membrane spanning regions by analysing amino acid hydrophobicity was employed to determine the putative transmembrane domains of Lp19 Hmg (von Heijne 1992). Seven hydrophobic regions were recognised between amino acid residues 1-641 (Figure 20) corresponding to seven putative transmembrane domains.

Using mutagenesis and kinetic analysis several amino acid residues have been shown to be important for catalysis of *P. mevalonii* and hamster Hmg (Bischoff and Rodwell 1997; Darnay *et al.* 1992; Frimpong and Rodwell 1994). These residues have been located in the catalytic domain of Lp19 Hmg and include: histidine, His^{1141}; glutamic acid, Glu^{836} and aspartic acid, Asp^{1046} (Figure 21). Based on the crystal structure of *P. mevalonii* Hmg, determined by Lawrence *et al.* (1995), four regions were identified which form the binding sites for HMG-CoA (I and II) and for reduced and oxidised NAD(P) (III and IV respectively). These regions are highlighted in Figure 21. Of particular interest with respect to the regulation of Hmg activity is the serine at position 1147 of Lp19 Hmg. This residue is located six residues downstream of the catalytically active histidine (His^{1141}), a spacing which appears to be strictly conserved in higher eukaryotes and fungi (Friesen and Rodwell 1997). In eukaryotic Hmgs this serine residue is known to be phosphorylated by AMP-activated protein kinase, resulting in a loss of catalytic activity but does not appear to play a role in catalysis or substrate recognition.

Using the GAP program of the GCG package the sequence of the putative catalytic domain, from position 736 to 1188 of Lp19 Hmg, was compared to that of other organisms. The amino acid sequence showed 83% similarity to *G. fujikuroi*, 69% to *S. cerevisiae* HMG2, 69% to *S. pombe*, 67% similarity to *S. cerevisiae* HMG1 sequence and 61% to *U. maydis* .

The region between the proposed membrane-bound and catalytic domains is a linker region, which has been shown to contain little sequence similarity between different organisms. Although some sequence similarity can be seen in this region (650-750) between *G. fujikuroi* and Lp19 Hmg (Figure 18).

Figure 20 Deduced amino acid sequence of the Lp19 Hmg transmembrane domain. Putative transmembrane domains are shown in blue.

```
  1  MISSSFLPNR FRGEPDRSQT SAAPSRIGKK LSPLLQFLAK VACSHPIHTV
 51  VTIAVLASTS YVGLIQDSLF EGPARLGKAD WSSLVDGSRD LIASADTKWQ
101  WSKVEQDSAS VKNSTHLALL TFVFPDTLSS ESASSAPRSH VVPTPQNLSI
151  TPIPATENSF TTYTQDSILA YSIPYTQAPE FISAAQEIPD EDAEEITTQH
201  GREKKKWIMK AAKVNSRSSI THWLSNAWVG FIDLLKNAET LDIVIMVLGY
251  LSMHLTFVSL FLSMRRMGSN FWLGTSTLFS SVFAFLFGLA VTTKLGVPIS
301  VILLSEGLPF LVVTIGFEKN IVLTRAVLSH AVEHRRTQGG REVQPGNKSG
351  GDKSQNIISY AIQAAIKDKG YEILRDYAIE ILILSLGAAS GVQGGLQQFC
401  FLAAWILFFD CILLFTFYTA ILSIKLEINR IKRHYEMRMA LEADGVSRRV
451  AEKVAKSNDD WTQSSGSESK NTTLFGRMRS SSVPKFKVLM ISGFVLINVI
501  NICTIPFRSA SSLSTLRSWA GGLGGVVSAP PVDPVKVAGK GLDAILTAAK
551  ASGKATLVTV LTPIKYELEY PSVHYALSSS LRDGAAGAVS SAVQFDNYGV
601  GGRMVGSLLK SLEDPVLSKW IVIALALSVG LNGYLFNVAR W
```

Figure 21 Deduced amino acid sequence of the Hmg catalytic domain. Catalytic residues are shown in **bold** and residues involved in substrate binding are shown in blue. The serine residue involved in regulation by phosphorylation and dephosphorylation is shown in red.

```
 736  DEEVVTMSMR GKVPGYALEK ALKDFTRAVK IRRTIISRTK ATSEITNGLD
                                    I
 786  RSKLPFENYN WERVFGACCE NVIGYLPLPV GVAGPLVIDG QSYFIPMATT
        II
 836  EGVLVASTSR GCKAINSGGG AITVLTSDGM TRGPCVSFET LERAGAAKLW
                                                       III
 886  LDSEAGQNTM KKAFNSTSRF ARLQHMKTAL AGTNLYIRFK TTTGDAMGMN

 936  MISKGVEHAL NVMATDGGFD DMNIITVSGN FCIDKKPAAM NWIDGRGKGI

 986  VAEAIIPADV VKSVLKSDVD ALVELNIAKN LIGSAMAGSI GGFNAHAANI
                                                      IV
1036  VAAIFLATGQ DPAQVVESCN CITTMNLHGS LQIAVSMPSL EVGTLGGGTI

1086  LEPQSAMLDM LGVRGSHPTN PGDNARRLAR IIGASVLAGE LSLCSALQAG

1136  HLVKAHMQHN RSAAPSRSTT PAPPPMTPVS LAMTIAQDKS SKSAAAQQRS

1186  KR
```

3.4.7 Codon Usage of Lp19 *hmg*

Codon usage was measured using the CODONFREQUENCY program of the GCG package and was clearly biased. There was a preference for C in the third nucleotide position and where a purine was found in the third position, G was generally used in preference to A. With the exception of the stop codons UAG and UGA, all codons were used (Table 6). This is as expected as highly expressed genes show a more marked codon bias than genes expressed at lower levels which is thought to be due to the need of highly expressed genes for more common isoaccepting tRNAs (Gurr *et al.* 1987).

Table 6 Combined codon bias for Lp19 *hmg*.

Amino Acid	Codon	Number	Fraction	Amino Acid	Codon	Number	Fraction
Gly	GGG	8.00	0.09	Thr	ACG	10.00	0.13
	GGA	13.00	0.14		ACA	9.00	0.12
	GGU	34.00	0.37		ACU	18.00	0.23
	GGC	36.00	0.40		ACC	40.00	0.52
Glu	GAG	34.00	0.61	STOP	UAG	0.00	0.00
	GAA	22.00	0.39	STOP	UAA	1.00	1.00
	GAU	18.00	0.36	Tyr	UAU	10.00	0.48
	GAC	32.00	0.64		UAC	11.00	0.52
Val	GUG	14.00	0.17	Leu	UUG	19.00	0.17
	GUA	11.00	0.13		UUA	1.00	0.01.
	GUU	30.00	0.36	Phe	UUU	15.00	0.34
	GUC	29.00	0.35		UUC	29.00	0.66
Ala	GCG	10.00	0.09	Ser	UCG	16.00	0.13
	GCA	13.00	0.11		UCA	8.00	0.15
	GCU	35.00	0.30		UCU	26.00	0.21
	GCC	59.00	0.50		UCC	21.00	0.17
Arg	AGG	3.00	0.05	Arg	CGG	8.00	0.14
	AGA	6.00	0.11		CGA	16.000.28	
Ser	AGU	16.00	0.13		CGU	10.00	0.18
	AGC	26.00	0.21		CGC	14.00	0.25
Lys	AAG	45.00	0.73	Gln	CAG	23.00	0.68
	AAA	17.00	0.27		CAA	11.00	0.32
Asn	AAU	9.00	0.20	His	CAU	8.00	0.36
	AAC	35.00	0.80		CAC	14.00	0.64
Met	AUG	32.00	1.00	Leu	CUG	25.00	0.22
Ile	AUA	3.00	0.04		CUA	12.00	0.11
	AUU	32.00	0.41		CUU	19.00	0.17
	AUC	43.00	0.55		CUC	38.00	0.33
Trp	UGG	15.00	1.00	Pro	CCG	9.00	0.17
STOP	UGA	0.00	0.00		CCA	8.00	0.15
Cys	UGU	3.00	0.23		CCU	8.00	0.33
	UGC	10.00	0.77		CCC	19.00	0.35

Chapter 4

DISCUSSION

4.1 Lp19 HMG

The endophyte (Lp19) which was isolated from perennial ryegrass, is classified as a *Neotyphodium* sp., and placed in taxon *L. perenne* taxonomic group one, otherwise known as LpTG-1 (Christensen *et al.* 1993). The *hmg* gene was isolated from this *N. lolii* isolate by PCR and the genomic region containing this gene was cloned. This allowed sequencing of the entire *hmg* gene as well as 5' and 3' non-translated regions.

The *hmg* gene of Lp19 was isolated by PCR using the degenerate primers hmg 1 and hmg 3, containing the neutral base inosine. A range of different sized products were amplified, which may be due to the low annealing temperature used and the degeneracy of the primers. The most abundant products were a 210 bp fragment and the expected 359 bp fragment (Figure 3D). These two fragments were purified and used to probe a Southern blot containing Lp1 and Lp19 digested genomic DNA (Figure 4). Both probes hybridised to fragments within the *Neotyphodium* genome, one fragment in the Lp19 digests and two within the Lp1 digests. The 359 bp fragment was sequenced and shown to be part of a unique *hmg* gene (Figure 6). The 210 bp fragment also weakly hybridised to a number of other fragments within each genomic digest and sequencing is required to further determine the nature of this PCR isolated fragment.

Lp19 possesses a haploid genome which has been shown, by Southern analysis, to contain a single copy of the *hmg* gene (Figure 4B). Lp1, another perennial ryegrass isolate, has been shown to contain two *hmg* genes, one from *N. lolii* and the other from *E. typhina* (Figure 11B). This provides further evidence for the conclusion that this isolate is a hybrid whose ancestors are *N. lolii* and *E. typhina* (Schardl *et al.* 1994). The ascomycetous fungi *G. fujikuroi*, *S. manihoticola* (Woitek *et al.* 1997) and the fission yeast *S. pombe* (Lum *et al.* 1996) have also been shown to contain only one *hmg* gene, in contrast to *S. cerevisiae* (Basson *et al.* 1986) and *U. maydis* (Croxen *et al.* 1994) which contain two structural genes encoding Hmg isozymes. The two genes in *S. cerevisiae* may be the result of a gene duplication event (Wolfe and Shields 1997).

4.2 *HMG* SEQUENCE FEATURES

4.2.1 Lp19 *hmg* Nucleotide Sequence

Lp19 *hmg* consists of two open reading frames of 3184 bp (ORF1) and 378 bp (ORF2), separated by a 73 bp intron. The nuclear genes of filamentous fungi are often punctuated by non-translated intervening sequences which are usually less than 100 bp in length (Gurr *et al.* 1987). The intronic region in the Lp19 *hmg* gene was determined by genomic sequence comparison to cDNA, and is located between positions 3186 and 3258 of the nucleotide coding sequence (Figure 14). ORF1 was shown to be free of any introns which may have left the sequence in frame after processing (Figure 15), thus proving the presence of a single intronic region within the *hmg* gene.

In general, introns begin with the dinucleotide GT and end with AG, forming the splice site junctions (Lewin 1990). These conserved nucleotides are present at the intron boundaries of Lp19 *hmg*. In some instances the dinucleotide GC has also been found to be present at the 5'-splice site. This has been described for one of the introns in the *E. typhina* β-tubulin (*tub*2) gene (Byrd *et al.* 1990) and also in *S. cerevisiae* where the introns in question are still functional (Hodge and Cumsky 1989). During splicing of the intronic region, the left exon is cleaved to produce a linear molecule. The remaining right intron-exon molecule is not linear and becomes linked to a consensus sequence within the intron, thus forming a lariat intermediate (Lewin 1990). This consensus sequence is referred to as the TACTAAC box and was found between positions 3235 and 3241 of Lp19 *hmg*. Mutations within this sequence prevent splicing in yeast, however this sequence is not as well conserved in higher eukaryotes (Lewin 1990). The *hmg* gene of *G. fujikuroi* is 2928 bp in length and also contains an intron, though this is smaller, being only 47 bp in length (Woitek *et al.* 1997). This intron occurs in approximately the same position as that in Lp19 and also contains the standard intron consensus sequences. In contrast, the *S. cerevisiae, S. pombe* and *U. maydis hmg* genes are continuous open reading frames. *S. cerevisiae HMG*1 and *HMG*2 are 3162 bp and 3135 bp respectively, *S. pombe* contains a 3159 bp gene and the *U. maydis* gene is 3495 bp in length. The occurrence and position of introns within corresponding genes of different organisms provides a basis for speculation as to the origin of these DNA elements, as well implying relationships between fungal genera. Filamentous fungal genes, in general, possess introns whilst the equivalent genes of the related ascomycete, *S. cerevisiae*, do not (Gurr *et al.* 1987).

It was determined that codon usage within Lp19 *hmg* was biased. There was a preference for C in the third position and where a purine was found in the third position, G was generally used in preference to A. This codon bias was also seen in the *tub*2 gene of *E. typhina* (Byrd *et al.* 1990). This was as expected, since highly expressed genes show a more marked codon bias than genes expressed at lower levels (Gurr *et al.* 1987). The *pyr*4 genes of isolate Lp1 showed codon bias, however this was not as extreme as in the *hmg* and *tub*2 genes; only a preference against A in the third position was observed, maybe due to low expression levels. Varying patterns of overall codon usage may be due to the abundance of codon isoacceptor tRNA molecules (Gurr *et al.* 1987). All of the codons were used in the *hmg* coding sequence of Lp19 (with the exception of two of the STOP codons) and in the Lp1 *pyr*4 genes. However, 15 codons were not used in the *E. typhina tub*2 gene, including two of the stop codons. The GC content of the coding sequence was determined for the Lp19 *hmg* gene. The open reading frames showed a preference for GC, with this content at 54%. The non-coding sequence of the intronic region showed a preference for AT, as the GC content was only 44%. This correlates with the open reading frames of the two *pyr*4 genes of Lp1, which are both approximately 54% GC rich. In comparison, the non-coding regions of these genes have a preference for AT, with a GC content of approximately 40% (Collett 1994). This GC richness of the coding regions is due to the low frequency with which A and T appear in the third position of codons.

4.2.2 5' Sequence Analysis

A 287 bp portion of the Lp19 *hmg* promoter region was analysed (-1 to -287) for putative regulatory elements. Regulation of filamentous fungal gene expression is exerted mainly at the stage of transcription, with a tendency toward integrative control of transcription initiation, as both major and minor initiation sites can be present (Gurr *et al.* 1987). The approximate transcription initiation site within the promoter region of Lp19 *hmg* was determined to occur between positions -215 and -76 (Figure 16). There is no extensive sequence homology at the start point of eukaryotic genes, however there is a tendency for the first base of mRNA to be A, flanked by pyrimidines. The exact position of the start site awaits further experimental analysis by primer extension.

Upstream of the transcription start site, short sequences occur which are recognised and bound by particular transcription factors, thus facilitating transcriptional activation. These sequences are distributed within approximately 150 bp immediately upstream of the start point of Lp19 *hmg*. The spacing between these regions, rather than the actual sequence, has been reported to be important in determining binding specificities of these regulatory regions. There is variability in the position and occurrence of these

fungal core promoter sequences and this raises questions as to their contribution and functional significance as promoters. As they may or may not be present within the fungal gene promoter, it is unclear whether these elements affect the level of expression or function of a gene, or which elements are essential for transcription (Gurr *et al.* 1987).

Two of the best characterised of these promoter region modules are the CAAT and TATA boxes. Putative CAAT and TATA consensus sequences were located near the position of the transcription start site of the Lp19 *hmg* gene. The TATA box is usually found approximately 30 bp upstream of the major transcriptional initiation point of most higher eukaryotic genes (Lewin 1990). This AT-rich motif is not a ubiquitous feature of filamentous fungal promoters, but when present is found in single copy, upstream of the major start site. This resembles the consensus of higher eukaryotes and *S. pombe*, but is distinct from the *S. cerevisiae* TATA consensus which often occurs in multi copy and in variable positions (Gurr *et al.* 1987). The putative TATA consensus sequence, TAATCATAAT, of Lp19 occurs from position -173 to -164, within the promoter region. The CAAT motif occurs at around -70 to -90 bp upstream of the mRNA start site in higher eukaryotes, but is rarely seen in *S. cerevisiae*. In filamentous fungal genes the putative CAAT consensus, when present, is usually observed between 60-120 bp upstream from the start site. Within the Lp19 *hmg* gene promoter a putative CAAT consensus sequence was located at position -255. The putative CAAT box of Lp19 *hmg* occurs 77 bp upstream of the determined putative TATA box. This is the correct spacing for these modules to be used as promoter elements with respect to the designated transcription initiation region.

A number of other elements are located within the 5' non-transcribed regions of genes which are involved in the control of transcription. Further experimental analysis of the promoter region of *hmg* will be required to elucidate these regions. No coding sequence has been identified 5' of the Lp19 *hmg* mRNA so the entire regulatory region cannot, as yet, be defined.

4.2.3 3' Sequence Analysis

Many higher eukaryotic genes are transcribed far beyond the site of polyadenylation of the mature mRNA. The transcript is subsequently cleaved, followed by the addition of a poly(A) tail approximately 10-30 bp downstream of the sequence AAUAAA. This consensus sequence is thought to be involved in polyadenylation of the 3' terminus of mature mRNA (Lewin 1990). This element is not necessarily an important regulatory feature, but it has been found to occur in a number of filamentous fungal genes (Gurr *et*

al. 1987). An AAUAAA consensus sequence was located within the *hmg* gene, from position 3683 to 3688, 42 bases downstream of the translational STOP codon, UAA. The 3'-terminus of *hmg* RNA was shown experimentally to correlate to this determined polyadenylation signal (Figure 17).

A microsatellite was located within the 3' non-coding region of Lp19 *hmg*. Microsatellites are tandemly repeated units of 2-5 bp of DNA which are highly abundant within eukaryotic genomes, and which can reach a length of up to 150 bp. As they are distributed throughout the eukaryotic genome, they are very useful as genomic markers and for genome mapping (Schlotter and Pemberton 1994). The microsatellite located downstream of the *hmg* gene consisted of the four base sequence TGTC which was tandemly repeated 18 times, creating a microsatellite 72 bp in length. The microsatellite has been shown in further analysis to be a highly polymorphic genetic marker for the analysis and fingerprinting of *N. lolii* isolates (Moon 1997).

4.2.4 Lp19 Hmg, Polypeptide Sequence

The *hmg* gene of Lp19 encodes a predicted protein of 1189 amino acids. In higher eukaryotes, recognition of the correct translation start site depends on the sequence context around the AUG codon, as this consensus sequence is necessary to target the ribosome to the correct start site (Kozak 1981). The translational start site for Lp19 *hmg* is proposed to occur at the AUG at nucleotide position +1 (Figure 12). The Hmg polypeptide sequence exhibits the greatest overall similarity to *G. fujikuroi* Hmg (79%), *U. maydis* (55%) and *S. cerevisiae* HMG2 (51%). The sequence similarity within the C-terminus is high and this corresponds to the catalytic domain of the enzyme, whereas there is very little sequence similarity within the amino-terminal domain, or the membrane bound region of the enzyme.

Sequence conservation within the catalytic domain of this enzyme is a feature of *hmg* sequences from a range of diverse organisms, which may indicate high sequence conservation during evolution. This may be associated with substrate and/or co-factor binding sites and catalytic activity. The catalytic domain of Lp19 Hmg was defined by comparison to the domains of the two *S. cerevisiae* Hmg polypeptides (Basson *et al.* 1986) and was determined to begin at residue 736 in Lp19 Hmg. The most sequence identity within this region was to *G. fujikuroi*, with 83% at the amino acid level, followed by 69% to *S. pombe* and *S. cerevisiae* HMG2, 67% to *S. cerevisiae* HMG1 and 61% to the isolated *U. maydis* Hmg. The high sequence similarity between the *hmg* genes of the two ascomycetous fungi, *N. lolii* and *G. fujikuroi*, alluded to their being closely related with little sequence divergence. These organisms are of the same

Ascomycete class, pyrenomycete, and were grouped together on a dendrogram (Figure 19). It was also seen that *S. pombe* Hmg grouped closer to the ascomycetous fungal Hmg proteins than to the Hmgs of the budding yeast *S. cerevisiae*.

Within the C-terminal domain of this enzyme, a number of amino acids have been found to be functionally conserved, and these play a role in the structural conformation and/or catalytic properties of the enzyme. Amino acid residues important for catalysis in the Hmg of *P. mevalonii, Haloferax volcanii* and hamster include histidine, glutamate and aspartate residues (Bischoff and Rodwell 1997; Friesen and Rodwell 1997; Frimpong and Rodwell 1994). These residues are also present in Lp19 Hmg at positions His^{1141}, Glu^{836} and Asp^{1046} (Figure 21). Also of particular interest with respect to the regulation of Hmg is the serine residue located at position 1147. This residue is located close to the catalytically important histidine residue at position 1141, and this six residue spacing between the histidine and serine has been shown to be conserved in most eukaryotes (Friesen and Rodwell 1997).

The activity of Hmg in higher eukaryotes is regulated by the reversible phosphorylation of the serine residue at position 1147 by an AMP-activated protein kinase, which results in a loss of catalytic activity. Interestingly, neither of the *S. cerevisiae* Hmgs contain a serine residue at this position in the protein. Protein kinase recognition motifs frequently contain basic residues on the N-terminal side of the target sequence and investigations of *P. mevalonii* Hmgs have implicated basic residues as being involved in substrate recognition by AMP-activated protein kinase. All Hmgs containing the serine residue have been found to have at least two basic amino acids at positions P-1 through P-4 (residues 1-4 upstream of the serine residue). The mammalian enzymes all have histidine at P-3 and arginine at P-1, while the plant enzymes have lysine at P-4 and arginine at P-1 (Friesen and Rodwell 1997). Lp19 Hmg contains a histidine at P-3 and arginine at P-1, like that of the mammalian enzymes. Kinetic analysis of the Hmg-catalysed reaction led to the hypothesis that the introduced phosphate interacts ionically with the active site histidine, inhibiting its ability to protonate the potentially inhibitory $CoAS^-$, thus arresting its release from the active site and halting the overall reaction (Omkumar and Rodwell 1994).

Residues involved in substrate binding have also been identified by studies of the crystal structure of *P. mevalonii* Hmg, and these residues appear to be mostly conserved over a range of organisms (Lawrence *et al.* 1995). Within Lp19 Hmg, the proposed binding sites for HMG-CoA occur at positions 805-814 and 835-841, while the reduced and oxidised forms of NAD(P) bind at 930-935 and 1078-1084, respectively (Figure

21). These binding regions conform to the consensus as determined by Lawrence *et al.* (1995).

The N-terminal domain of Lp19 Hmg (residues 1-641) contains seven putative transmembrane domains, as determined using TMpred program of NCBI (von Heijne 1992). This number was also predicted within the *G. fujikuroi* Hmg N-terminal domain, using Kyte and Doolittle hydropathy plots. In the *U. maydis* and *S. cerevisiae* Hmg proteins, eight hydrophobic regions have been predicted, but only seven of these have been recognised as possible transmembrane domains (Senstag *et al.* 1990). The N-terminal domain is important for anchoring the enzyme to the endoplasmic reticulum membrane, as well as being required for the sterol-regulated proteolytic degradation of the enzyme. However, the domain is not required for catalytic functioning of the enzyme (Gil *et al.* 1985).

4.3 *G. FUJIKUROI* HMG

The coding sequence of the *G. fujikuroi hmg* gene is somewhat shorter than that of other fungal *hmgs* and Woitek *et al.* (1997) described the gene as having a truncated N-terminal region. A 6-frame translation of the *G. fujikuroi hmg* nucleotide sequence was performed, using the MAP program of the GCG package. This showed the two open reading frames as determined by Woitek *et al.* (1997), but in addition a region of openness was also determined in a different frame, overlapping the 5' end of *G. fujikuroi* ORF1 (see Appendix 5.0). The nucleotide sequence in this frame also shows similarity to the sequence of Lp19 *hmg*, supporting the possibility that the N-terminus of this gene is not truncated, but that a sequencing error forced a frame shift (Figure 22). If this is part of the actual coding sequence, the polypeptide sequence of *G. fujikuroi* Hmg would increase to become larger than the *S. cerevisiae* polypeptide sequences. There are a number of methionines present upstream of the determined translation start methionine of *G. fujikuroi* Hmg, and interestingly, if the coding sequence begins at the first methionine of this new open reading frame, the first four amino acids of the new sequence are identical to those of *S. cerevisiae* HMG2 (Figure 18).

The *hmg* gene of *N. lolii* isolate Lp19 has been isolated and the upstream regulatory region identified and characterised. Further experimental analysis can now be undertaken to determine hmg expression *in planta*. This will enable some aspects of the regulatory influence of *hmg* on the isoprenoid pathway and toxin production to be elucidated.

```
Lp19            MISSSFLPNR FRGEPDRSQT SAAPSRIGKK LSPLLQFLAK VACSHPIHTV

Lp19            VTIAVLASTS YVGLIQDSLF EGPARLGKAD WSSLVDGSRD LIASADTKWQ

Lp19            WSKVEQDSAS VKNSTHLALL TFVFPDTLSS ESASSAPRSH VVPTPQNLSI
G.fujikuroi2        MSLPML ARMPTTLPSG TLVFQQA.SA ESSSTAPRSH HVPIPQNLSI

Lp19            TPIPATENSF TTYTQDSILA YSIPYTQAPE FISAAQEIPD EDAEEITTQH
G.fujikuroi2    TSLPSTENPF TAYSQDSILA YALPYSEGPE FLAAAQEIPN EDAVEIETKH

Lp19            GREKKKWIMK AAKVNSRSSI THWLSNAWVG FIDLLKNAET LDIVIMVLGY
G.fujikuroi2    GREKKTWIMK AAKVNTRNSV VNGLAMPGAS SSTFSRTPRL SISSSCFLVT
G.fujikuroi1          MDHE GCQGQHPQQC CQWVSNAWSE FLDLLKNAET LDIVIMLLGY

Lp19            LSMHLTFVSL FLSMRRMGSN FWLGTSTLFS SVFAFLFGLA VTTKLGVPIS
G.fujikuroi2    LPCT*
G.fujikuroi1    IAMHLTFVSL FLSMRKMGSK FWLGICTLFS SVFAFLFGLV VTTKLGVPIS

Lp19            VILLSEGLPF LVVTIGFEKN IVLTRAVLSH AVEHRRTQGG REVQPGNKSG
G.fujikuroi1    VILLSEGLPF LVVTIGFEKN IVLTRAVMSH AIEHRRIQ.A QNSKSGKRSP

Lp19            GDKSQNIISY AIQAAIKDKG YEILRDYAIE ILILSLGAAS GVQGGLQQFC
G.fujikuroi1    DGSTQNMIQY AVQAAIKEKG FEIIRDYAIE IVILVIGAAS GVQGGLQQFC
```

Figure 22 Alignment of *G. fujikuroi* Hmg published coding sequence and upstream sequence with Lp19 Hmg.

The published sequence of *G. fujikuroi* (*G. fujikuroi*1, Woitek *et al.* 1997) is aligned to Lp19 Hmg and the identity is shown in blue. The upstream sequence (*G. fujikuroi*2), which was in another frame is also aligned to the Hmg polypeptide sequence of Lp19. This upstream open reading frame showed 62% identity to Lp19 Hmg and this is highlighted in red.

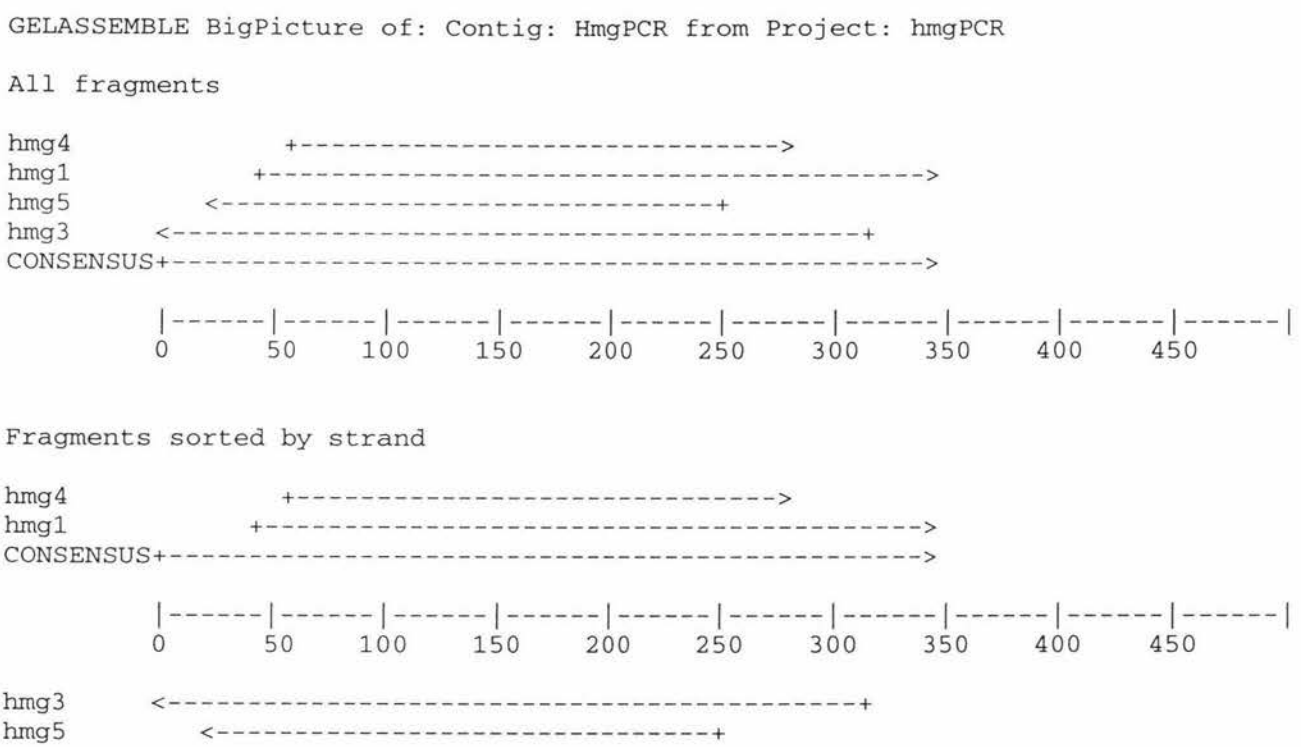
APPENDIX

Appendix 1.0 SEQUENCE DATA

A 1.1 PRETTYOUT from *hmg* PCR Contig

```
GELASSEMBLE PrettyOut of Contig: hmg PCR from Project: hmgPCR
From: 1  To: 345

hmg1        >                                                  tGGCCACCGACGGCgtg 17
hmg3        < atgaagatgatttcaaagggtgttgAGCATGCGCTCAACGTCATG..cACCGACGGTGGC 60
CONSENSUS   > ATGAAGATGATTTCAAAGGGTGTTGAGCATGCGCTCAACGTCATGgcCACCGACGGyGks 60

                .........+.........+.........+.........+.........+.........+

hmg1        > tttGATGACATGAACATCATCACCGTCTCTGGCaACTTCTGtaTTGAcaAGAAACCCGcC 77
hmg3        < TTTGATGACATGAACATCATCACCGTCTCTGGCAACTTCTGTATTGACAAGAAACCCGCC 120
CONSENSUS   > TTTGATGACATGAACATCATCACCGTCTCTGGCAACTTCTGTATTGACAAGAAACCCGCC 120

                .........+.........+.........+.........+.........+.........+

hmg1        > GcatTGAACTGGATCGACGGT.CGCGG.acaGGGacTTGTTGCTGAGGacctacTTcCTG 137
hmg3        < GCTATGAACTGGATCGACGGTACGCGGcaYAGGGCATTGTTGCTGAGGCCATCATTCCTG 180
CONSENSUS   > GCwwTGAACTGGATCGACGGTaCGCGGcAcAGGGmmTTGTTGCTGAGGmCmTmmTTCCTG 180

                .........+.........+.........+.........+.........+.........+

hmg1        > CCGACGTGGTCAAATCCGTGTTGCAAGAGTGATGTTGATGCTTTGGTAGAGCTCAATATT 197
hmg3        < CCGACGTGGTCAAATCCGTGTTG.AAGAGTGATGTTGATGCTTTGGTAGAGCTCAATATT 240
CONSENSUS   > CCGACGTGGTCAAATCCGTGTTGcAAGAGTGATGTTGATGCTTTGGTAGAGCTCAATATT 240

                .........+.........+.........+.........+.........+.........+

hmg1        > GCTAAGAACCTGATCGGGTCTGCCATGGCCGGTTCGATCGGTGGCTTCAACGcCCACGCT 257
hmg3        < GCTAAGAACCTGATCGGGTCTGCCATGGCCGGTTCGATCGGTGGCTTCAACGCCCACGCT 300
CONSENSUS   > GCTAAGAACCTGATCGGGTCTGCCATGGCCGGTTCGATCGGTGGCTTCAACGCCCACGCT 300

                .........+.........+.........+.........+.........+.........+

hmg1        > GCCAACATCGTAGCGGCCATATTCCTCGcCACCGGCCAAGACCcc                 302
hmg3        < GCCAACATCGTAtacg                                              316
CONSENSUS   > GCCAACATCGTAkmsGCCATATTCCTCGCCACCGGCCAAGACCCC                 345

                .........+.........+.........+.........+.........+.........+
```

A 1.2 BIGPICTURE of *hmg* PCR Contig



A 1.3 PRETTYOUT from *hmg* Contig

```
GELASSEMBLE PrettyOut of Contig: hmg.fil from Project: hmg4
From: 1  To: 4885

p41hmg23    > ATCCTTGTGATCAATATTCAAAGTTCAAAcAAGTCCAGGAAAAGGGGGGCAGTTCCTTCG 60
CONSENSUS   > ATCCTTGTGATCAATATTCAAAGTTCAAACAAGTCCAGGAAAAGGGGGGCAGTTCCTTCG 60

                .........+.........+.........+.........+.........+.........+

p41hmg23    > GTGTCTCATTTTGGCTGTAACATCAATTGCTTTTCGTGGTTGCTATTGTTCTGCAGTTTT 120
CONSENSUS   > GTGTCTCATTTTGGCTGTAACATCAATTGCTTTTCGTGGTTGCTATTGTTCTGCAGTTTT 120

                .........+.........+.........+.........+.........+.........+

p41hmg23    > CAATTCAGGGTCCAGTTCTGTGTCGGCGCAACCCGCTTTTGCTTGATTGGCCCCGCGATC 180
CONSENSUS   > CAATTCAGGGTCCAGTTCTGTGTCGGCGCAACCCGCTTTTGCTTGATTGGCCCCGCGATC 180

                .........+.........+.........+.........+.........+.........+

p41hmg23    > CCTCCTTTCACAACCACAATCGATTGAGAATCTCGTCCGCCACTTTTCAACACACGACAT 240
CONSENSUS   > CCTCCTTTCACAACCACAATCGATTGAGAATCTCGTCCGCCACTTTTCAACACACGACAT 240

                .........+.........+.........+.........+.........+.........+

p41hmg23    > CTGTCTCCCGTCTCCCGTCATCCGTCATCGCAACCTTCGCGCCTGCCAATTCCTATCAAG 300
CONSENSUS   > CTGTCTCCCGTCTCCCGTCATCCGTCATCGCAACCTTCGCGCCTGCCAATTCCTATCAAG 300

                .........+.........+.........+.........+.........+.........+

p41hmg23    > TGCACGGCGGTGCTCTCTACACCCGAGGTCAGTAAACCTGTTACTTGCTCTTACGTCTTG 360
CONSENSUS   > TGCACGGCGGTGCTCTCTACACCCGAGGTCAGTAAACCTGTTACTTGCTCTTACGTCTTG 360

                .........+.........+.........+.........+.........+.........+

p41hmg21    >                                                GAGACGcTGCGGTGCACCAC 20
p41hmg23    > CTGCGTCGTGGTCTGGTCTTGCCTCACTCGACAGATCGTCGAGACGCTGCGGTGCACCAC 420
CONSENSUS   > CTGCGTCGTGGTCTGGTCTTGCCTCACTCGACAGATCGTCGAGACGCTGCGGTGCACCAC 420

                .........+.........+.........+.........+.........+.........+

p41hmg21    > CACGACGCCCGAAAAATCCcTCCATACGACATCAATGcCTTCcTCATcTGGTGCAACACC 80
p41hmg23    > CACGACGCCCGAAAAATCCCTCCATACGACATCAATGCCTTCCTCATCTGGTGCAACACC 480
CONSENSUS   > CACGACGCCCGAAAAATCCCTCCATACGACATCAATGCCTTCCTCATCTGGTGCAACACC 480

                .........+.........+.........+.........+.........+.........+

p41hmg22    >                                          TCcAAGTCAGCcACATGACGATAATC 26
p41hmg21    > ATGGCTGCCCCATCTCTTGCTTTACACGgGACGATCCAAGTCAGCCACATGACGATAATC 140
p41hmg23    > ATGGCTGCCCCATCTCTTGCTTTACACgGGACGAtCCAA                       519
CONSENSUS   > ATGGCTGCCCCATCTCTTGCTTTACACGGGACGATCCAAGTCAGCCACATGACGATAATC 540

                .........+.........+.........+.........+.........+.........+
```

```
p41hmg22    > ATAATCAGTCCCCATCAACCCCATCACAACGTTTACCACTACCCGCCTCATTTACTTACA 86
p41hmg21    > ATAATCAGTCCCCATCAACCCCATCACAACGTTTACCACTACCCGCCTCATTTACTTACA 200
CONSENSUS   > ATAATCAGTCCCCATCAACCCCATCACAACGTTTACCACTACCCGCCTCATTTACTTACA 600

                .........+.........+.........+.........+.........+.........+

p41hmg22    > CGCATTCAGCTCTTTACTTGGACAGAACCGTTGCCGATAACGAGGATCCACGCAAGaCCC 146
p41hmg21    > CGCATTCAGCTCTTTACTTGGACAGAACCGTTGCCGATAACGAGGATCCACGCAAGACCC 260
CONSENSUS   > CGCATTCAGCTCTTTACTTGGACAGAACCGTTGCCGATAACGAGGATCCACGCAAGACCC 660

                .........+.........+.........+.........+.........+.........+

p41hmg22    > GATGTCGTAAACCGGACCGCCGAGTTACTTTTTTCCGATCAAGTCGACATGATATCTTCT 206
p41hmg21    > GATGTCGTAAACCGGACCGCCGAGTTACTTTTTTCCGATCAAGTCGACATGATATCTTCT 320
CONSENSUS   > GATGTCGTAAACCGGACCGCCGAGTTACTTTTTTCCGATCAAGTCGACATGATATCTTCT 720

                .........+.........+.........+.........+.........+.........+

p41hmg22    > TCATTCCTACCGAACCGCTTTCGCGGTGAACCTGACCGCTCCCAAACCTCAGCCGCTCCA 266
p41hmg21    > TCATTCCTACCGAACCGCTTTCGCGGTGAACCTGACCGCTCCCAAACCTCAGCCGCTCCA 380
CONSENSUS   > TCATTCCTACCGAACCGCTTTCGCGGTGAACCTGACCGCTCCCAAACCTCAGCCGCTCCA 780

                .........+.........+.........+.........+.........+.........+

p41hmg22    > TCGCGCATCGGCAAAAAGCTCTCGCCTCTGCTGCAGTTCCTAGCTAAAGTGGCTTGCTCG 326
p41hmg21    > TCGCGCATCGGCAAAAAGCTCTCGCCTCTGCTGCAGTTCCTAGCTAAAGTGGCTTGCTCG 440
CONSENSUS   > TCGCGCATCGGCAAAAAGCTCTCGCCTCTGCTGCAGTTCCTAGCTAAAGTGGCTTGCTCG 840

                .........+.........+.........+.........+.........+.........+

p41hmg20    >                                                   TCATATGTTGGCCTG 15
p41hmg18    >  ACCCAATCCACACCGTCGTTACCATCGCCGTTCTAGCCAGCACGTCATATGTTGGCCTG 59
p41hmg22    > CACCCAATCCACACCGTCGTTACCATCGCCGTTCTAGCCAGCACGTCATATGTTGGCCTG 386
p41hmg21    > CACCCAATCCACACCGTCGTTACCATCGCCGTTCTAGCCAGCACGTCATATGTTGGCCTG 500
CONSENSUS   > CACCCAATCCACACCGTCGTTACCATCGCCGTTCTAGCCAGCACGTCATATGTTGGCCTG 900

                .........+.........+.........+.........+.........+.........+

p41hmg20    > ATCCAGGACAGCCTGTTCGAAGGCCCTGCGAGGCTCGGCAAGGCTGACTGGTCATCTCTG 75
p41hmg18    > ATCCAGGACAGCCTGTTCGAAGGCCCTGCGAGGCTCGGCAAGGCTGACTGGTCATCTCTG 119
p41hmg22    > ATCCAGGACAGCCTGTTCGAAgGCCCTGCGAgGCTCGGCAAGGCTGACTGGTCATCTCTG 446
p41hmg21    > ATCCAGGACAGCCTGTTCG                                           519
CONSENSUS   > ATCCAGGACAGCCTGTTCGAAGGCCCTGCGAGGCTCGGCAAGGCTGACTGGTCATCTCTG 960

                .........+.........+.........+.........+.........+.........+

p41hmg20    > GTGGATGGTAGCAGAgATCTTATTGCCAGTGCTGATACCAAATGGCAGTGGTCAAAAGTC 135
p41hmg18    > GTGGATGGTAGCAGAGATCTTATTGCCAGTGCTGATACCAAATGGCAGTGGTCAAAAGTC 179
p41hmg22    > GTGGATGGTAgCAgAgaTCTTATTGCCAGTGCTGATACCAAATGGCAGTGGTCAAAAgTC 506
CONSENSUS   > GTGGATGGTAGCAGAGATCTTATTGCCAGTGCTGATACCAAATGGCAGTGGTCAAAAGTC 1020

                .........+.........+.........+.........+.........+.........+

p41hmg20    > GAGCAAGATTCTGCCTCTGTCAAGAACAGCACTCACCTGGCTCTTCTGACATTCGTCTTC 195
p41hmg18    > GAGCAAGATTCTGCCTCTGTCAAGAACAGCACTCACCTGGCTCTTCTGACATTCGTCTTC 239
p41hmg22    > GAgCAAgATTCTGCCTCTGT                                          526
CONSENSUS   > GAGCAAGATTCTGCCTCTGTCAAGAACAGCACTCACCTGGCTCTTCTGACATTCGTCTTC 1080

                .........+.........+.........+.........+.........+.........+
```

```
p41hmg20    > CCCGACACTCTCTCGTCTGAATCAGCCAGCTCTGCTCCTCGATCCCACGTCGTCCCGACT 255
p41hmg18    > CCCGACACTCTCTCGTCTGAATCAGCCAGCTCTGCTCCTCGATCCCACGTCGTCCCGACT 299
CONSENSUS   > CCCGACACTCTCTCGTCTGAATCAGCCAGCTCTGCTCCTCGATCCCACGTCGTCCCGACT 1140

                .........+.........+.........+.........+.........+.........+

p41hmg20    > CCCCAGAACCTCTCGATCACCCCAATCCCTGCCACTGAGAACTCGTTCACGACCTACACT 315
p41hmg18    > CCCCAGAACCTCTCGATCACCCCAATCCCTGCCACTGAGAACTCGTTCACGACCTACACT 359
CONSENSUS   > CCCCAGAACCTCTCGATCACCCCAATCCCTGCCACTGAGAACTCGTTCACGACCTACACT 1200

                .........+.........+.........+.........+.........+.........+

p41hmg13    >                                                    TTTCAGCG 8
p41hmg20    > CAGGATAGCATTCTTGCCTATTCCATTCCTTACACCCAGGCACCGGAGTTCATTTCAGCG 375
p41hmg18    > CAGGATAGCATTCTTGCCTATTCCATTCCTTACACCCAGGCACCGGAGTTCATTTCAGCG 419
CONSENSUS   > CAGGATAGCATTCTTGCCTATTCCATTCCTTACACCCAGGCACCGGAGTTCATTTCAGCG 1260

                .........+.........+.........+.........+.........+.........+

p41hmg13    > GCCCAGGAAATTCCCGATGAGGATGcTGAAGAGATCACAAcTCAACATGGGcGAGAGAAG 68
p41hmg20    > GCCCAgGAAATTCCCGATGAGGATGCtGAAGAGATCaCAACTCAaCATGGGCGAGAGAAG 435
p41hmg18    > GCCCAGGAAATTCCCGATGAGGATGCTGAAGAGATCACAACTCAACATGGGCGAGAGAAG 479
CONSENSUS   > GCCCAGGAAATTCCCGATGAGGATGCTGAAGAGATCACAACTCAACATGGGCGAGAGAAG 1320

                .........+.........+.........+.........+.........+.........+

p41hmg13    > AAGAAGTGGATTATGAAGGCTGCCAAGGTCAAcTcTcGAAGTTCCATCACGCACTGGcTC 128
p41hmg20    > AAGAAGTGGATTATGAAGGCTGCCAAGGTCAACTCTCgAAGTTCCATCACGCACTGGCTC 495
p41hmg18    > AAGAAGTGGATTATGAAGGCTGCCAAGGTCAACTCTCGAAGTTCCATCACGCACTGGCTC 539
CONSENSUS   > AAGAAGTGGATTATGAAGGCTGCCAAGGTCAACTCTCGAAGTTCCATCACGCACTGGCTC 1380

                .........+.........+.........+.........+.........+.........+

p41hmg13    > AGCAACGCGTGGGTGGGATTCATCGACCTTCTCAAAAACGCCGAAACCCTCGACATCGTC 188
p41hmg20    > AGCAACGcGTGGGTGGGATTCATCgAcCTTCTCaAAAACgCCgAAAcCCTCGACATCGTC 555
p41hmg18    > AGCAACGCGTGGGTGGGATTCATCGACCTTCTCAAAAACGCCGAAACCCTCG          591
CONSENSUS   > AGCAACGCGTGGGTGGGATTCATCGACCTTCTCAAAAACGCCGAAACCCTCGACATCGTC 1440

                .........+.........+.........+.........+.........+.........+

p41hmg19    >                 TTAtCTCTCCATGCATCTGAcCTTTGTGTCTCTCTTCCTCTCCATG 46
p41hmg13    > ATCATGGTGCTTGGTTATCTCTCCATGCATCTGACCTTTGTGTCTCTCTTCCTCTCCATG 248
p41hmg20    > ATCATGGTGCTTGGTTATCTCTCCATGCATCTgAcCTTTG                      595
CONSENSUS   > ATCATGGTGCTTGGTTATCTCTCCATGCATCTGACCTTTGTGTCTCTCTTCCTCTCCATG 1500

                .........+.........+.........+.........+.........+.........+

p41hmg19    > AGACGAATGGGCTCCAACTTCTGGCTAGGCACGAGCACCCTTTTTTCGTCTGTCTTTGCC 106
p41hmg13    > AGACGAATGGGCTCCAACTTCTGGCTAGGCACGAGCACCCTTTTTTCGTCTGTCTTTGCC 308
CONSENSUS   > AGACGAATGGGCTCCAACTTCTGGCTAGGCACGAGCACCCTTTTTTCGTCTGTCTTTGCC 1560

                .........+.........+.........+.........+.........+.........+

p41hmg19    > TTTCTTTTCGGTCTAGCCGTGACCACTAAGCTCGGCGTACCGATCAGTGTTATTCTCTTG 166
p41hmg13    > TTTCTTTTCGGTCTAGCCGTGACCACTAAGCTCGGCGTACCGATCAGTGTTATTcTCTTG 368
CONSENSUS   > TTTCTTTTCGGTCTAGCCGTGACCACTAAGCTCGGCGTACCGATCAGTGTTATTCTCTTG 1620

                .........+.........+.........+.........+.........+.........+
```

```
p41hmg19    > TCCGAGGGCCTGCCGTTCCTTGTTGTAACAATTGGTTTTGAgAAGAACATTGTCCTCACC 226
p41hmg13    > TCCGAGGGCCTGCCGTTCCTTGTTGTAACAATTGGTTTTGAGAAGAACATTGTCCTCACC 428
CONSENSUS   > TCCGAGGGCCTGCCGTTCCTTGTTGTAACAATTGGTTTTGAGAAGAACATTGTCCTCACC 1680

                .........+.........+.........+.........+.........+.........+

p41hmg19    > AGAGCTGTCCTCAGCCATGCCGTTGAGCATCGTCGAACGCAAGGAGGCCGTGAGGTCCAG 286
p41hmg13    > AGAGCTGTCCTCAGCCATGCCGTTGAGCATCGTCGAACGCAAGGAGGCCGTGAGGTCCAG 488
CONSENSUS   > AGAGCTGTCCTCAGCCATGCCGTTGAGCATCGTCGAACGCAAGGAGGCCGTGAGGTCCAG 1740

                .........+.........+.........+.........+.........+.........+

p41hmg19    > CCCGGCAACAAGTCCGGTGGGGACAAATCACAAAACATCATTTCCTACGCCATTCAAGCT 346
p41hmg13    > CCCGGCAACAAGTCCGGTGGGGACAAATCACAAAACATCATTTCCTACGCCATTCAAGCT 548
CONSENSUS   > CCCGGCAACAAGTCCGGTGGGGACAAATCACAAAACATCATTTCCTACGCCATTCAAGCT 1800

                .........+.........+.........+.........+.........+.........+

p41hmg10    >                     TTATGAAATTCTGCGtGACTATGCCATCGAAATCTTGATTCTA 43
p41hmg19    > GCCATCAAGGACAAGGGTTATGAAATTCTGCGTGACTATGCCATCGAAATCTTGATTCTA 406
p41hmg13    > GCCATCAAGGACAAGGGTTATGAAATTcTGCGTGACTATGCCAT                 592
CONSENSUS   > GCCATCAAGGACAAGGGTTATGAAATTCTGCGTGACTATGCCATCGAAATCTTGATTCTA 1860

                .........+.........+.........+.........+.........+.........+

p41hmg14    >                       TGTTCAGGGAgGTCTCCAGCAGTTCTGCTTCTTGGCTGCG 40
p41hmg10    > TCGTTGGGCGCAGCcTCGGGTGTTCAGGGAGGTCTCCAGCAGTTCTGCTTCTTGGCTGCG 103
p41hmg19    > TCGTTGGGCGCAGCCTCGGGTGTTCAGGGAgGTCTCCAGCAGTTCTGCTTCTTGGCTGCG 466
CONSENSUS   > TCGTTGGGCGCAGCCTCGGGTGTTCAGGGAGGTCTCCAGCAGTTCTGCTTCTTGGCTGCG 1920

                .........+.........+.........+.........+.........+.........+

p41hmg14    > TGGATCTTGTTCTTTGACTGTATTCTCTTGTTTACCTTCTACACTGCAATCCTCAGTATC 100
p41hmg10    > TGGATCTTGTTCTTTGACTGTATTCTCTTGTTTACCTTCTACACTGCAATCCTCAGTATC 163
p41hmg19    > TGGATCTTGTTCTTTGACTGTATTCTCTTGTTTACCTTCTACACTGCAATCCTCAGTATC 526
CONSENSUS   > TGGATCTTGTTCTTTGACTGTATTCTCTTGTTTACCTTCTACACTGCAATCCTCAGTATC 1980

                .........+.........+.........+.........+.........+.........+

p41hmg14    > AAACTCGAAATCAACCGCATCAAGCGACACTACGAgATGCGCATGGCCCTCGAAGCAGAT 160
p41hmg10    > AAACTCGAAATCAACCGCATCAAGCGACACTACGAGATGCGCATGGCCCTCGAAGCAGAT 223
p41hmg19    > AAACTCGAAATCAACCGCATCAAGCGACACTAC                            559
CONSENSUS   > AAACTCGAAATCAACCGCATCAAGCGACACTACGAGATGCGCATGGCCCTCGAAGCAGAT 2040

                .........+.........+.........+.........+.........+.........+

p41hmg14    > GGTGTTAGTCGGCGCGTAgCCGAgAAGGTGGCCAAGAgCAACGATGACTGGACCCAGTCC 220
p41hmg10    > GGTGTTAGTCGGCGCGTAGCCGAGAAGGTGGCCAAGAGCAACGATGACTGGACCCAGTCC 283
CONSENSUS   > GGTGTTAGTCGGCGCGTAGCCGAGAAGGTGGCCAAGAGCAACGATGACTGGACCCAGTCC 2100

                .........+.........+.........+.........+.........+.........+

p41hmg14    > AGTGGATCTGAATCCAAGAACACAACCTTGTTTGGTCGCATGCGGAGTAgCAGCGTTCCA 280
p41hmg10    > AGTGGATCTGAATCCAAGAACACAACCTTGTTTGGTCGCATGCGGAGTAGCAGCGTTCCA 343
CONSENSUS   > AGTGGATCTGAATCCAAGAACACAACCTTGTTTGGTCGCATGCGGAGTAGCAGCGTTCCA 2160

                .........+.........+.........+.........+.........+.........+
```

```
hmg26       >                                TTCTGGGTTCGTTCTCATCAACGTAATCAACATTTGCACG 40
p41hmg14    > AAGTTCAAGGTGCTTATGATTTCTGGGTTCGTTCTCATCAACGTAATCAACATTTGCACG 340
p41hmg10    > AAGTTCAAGGTGCTTATGATTTCTGGGTTCGTTCTCATCAACGTAATCAACATTTGCACG 403
CONSENSUS   > AAGTTCAAGGTGCTTATGATTTCTGGGTTCGTTCTCATCAACGTAATCAACATTTGCACG 2220

                .........+.........+.........+.........+.........+.........+

hmg26       > ATCCCTTTCCGCAGCGCAAGCTCTCTATCAACCCTTCGATCGTGGGCTGGTGGTCTGGGC 100
p41hmg14    > ATCCCTTTCCGCAGCGCAAGCTCTCTATCAACCCTTCGATCGTGGGCTGGTGGTCTGGGC 400
p41hmg10    > ATCCCTTTCCGCAGCGCAAGCTCTCTATCAACCCTTCGAtCGTGGGCTGGTGGTCTGGGC 463
CONSENSUS   > ATCCCTTTCCGCAGCGCAAGCTCTCTATCAACCCTTCGATCGTGGGCTGGTGGTCTGGGC 2280

                .........+.........+.........+.........+.........+.........+

hmg26       > GGTGTTGTGTCTGCTCCTCCGGTTGACCCTGTCAAGGTTGCTGGCAAGGGTCTTGACGCC 160
p41hmg14    > GGTGTTGTGTCTGCTCCTCCGGTTGACCCTGTCAAGGTTGCTGGCAAGGGTCTTGACGCC 460
p41hmg10    > GGTGTTGTGtCTGCtCCTCCGGTTGACCCTGTCAAGGTTGCTGGCAAGGGtCTTGACGCC 523
CONSENSUS   > GGTGTTGTGTCTGCTCCTCCGGTTGACCCTGTCAAGGTTGCTGGCAAGGGTCTTGACGCC 2340

                .........+.........+.........+.........+.........+.........+

p41F        >                    AAAGCAAGTGGGAAGGCAACCcTGGTCACTGTCCTTACACCCATC 45
hmg26       > ATTCTGACCGCCGCCAAAGCAAGTGGGAAGGCAACCCTGGTCACTGTCCTTACACCCATC 220
p41hmg14    > ATTCTGACCGCCGCCAAAGCAAGTGGGAAGGCAACCCTGGTCACTGTCCTTACACCCATC 520
p41hmg10    > ATTCTGACCGCCGCCAAAGCAAGTGGGAAGGCAACc                          559
CONSENSUS   > ATTCTGACCGCCGCCAAAGCAAGTGGGAAGGCAACCCTGGTCACTGTCCTTACACCCATC 2400

                .........+.........+.........+.........+.........+.........+

p41hmg11    >                                                      AGAGACGGC 9
p41F        > AAATACGAACTAGAGTAcCCCTCTGTTCAcTACGCGCTGTCATCcTCGTTGAGAGACGGC 105
hmg26       > AAATACgAACTAgAGTACCCCTCTGTTCACTACGCGCTGTCATCCTCGTTGAgAgACGGC 280
p41hmg14    > AAATACGAACTAGAgTAcCCCTCTGTTCACTACGCGCTGTCATCCTCgTTGAGAGAC    577
CONSENSUS   > AAATACGAACTAGAGTACCCCTCTGTTCACTACGCGCTGTCATCCTCGTTGAGAGACGGC 2460

                .........+.........+.........+.........+.........+.........+

p41hmg11    > GCAGCCGGTGCCGTAAGTTCTGCCGTTCAGTTTGACAATTATGGGGTTGGTGGCCGAATG 69
p41F        > GCAGCCGGTGCCGTAAGTTCTGCCGTTCAGTTTGACAATTATGGGGTTGGTGGCCGAATG 165
hmg26       > GCAGCCGGTGCCGTAAGTTCTGCCGTTCAGTTTGACAATTATGGGGTTGGTGGCCGAATG 340
CONSENSUS   > GCAGCCGGTGCCGTAAGTTCTGCCGTTCAGTTTGACAATTATGGGGTTGGTGGCCGAATG 2520

                .........+.........+.........+.........+.........+.........+

p41hmg11    > GTTGGGAGCCTCTTGAAGAGCCTCGAGGACCCCGTACTTTCAAAATGGATAGTCATCGCC 129
p41F        > GTTGGGAGCcTCTTGAAGAGCCTCGAGGACCCCGTACTTTCAAAATGGATAGTCATCGCC 225
hmg26       > GTTGGGAgCCTCTTGAAgAGCCTCgAGGACCCCGTACTTTCAAAATGGATAgTCATCGCC 400
CONSENSUS   > GTTGGGAGCCTCTTGAAGAGCCTCGAGGACCCCGTACTTTCAAAATGGATAGTCATCGCC 2580

                .........+.........+.........+.........+.........+.........+

p41hmg11    > CTGGCCCTAAGTGTTGGCCTCAATGGATACCTCTTCAATGTCGCTAGATGGAGCATCAAA 189
p41F        > CTGGCCCTAAGTGTTGGCCTCAATGGATACCTCTTCAATGTCGCTAGATGGAGCATCAAA 285
hmg26       > CTGGCCCTAAgTgTTGGCCTCAATGGATACCTCTTCAATGTCGCTAgATGGAGCATC    457
CONSENSUS   > CTGGCCCTAAGTGTTGGCCTCAATGGATACCTCTTCAATGTCGCTAGATGGAGCATCAAA 2640

                .........+.........+.........+.........+.........+.........+

p41hmg11    > GACCCTAACGTTCCGGACCACGGAATTGACCGCAAGGAGCTGGCCCGGGCTCAACGCTTC 249
p41F        > GACCCTAACGTTCCGGACCACGGAATTGACCGCAAGGAGCTGGCCCGGGCTCAACGCTTC 345
CONSENSUS   > GACCCTAACGTTCCGGACCACGGAATTGACCGCAAGGAGCTGGCCCGGGCTCAACGCTTC 2700

                .........+.........+.........+.........+.........+.........+
```

```
p41hmg11    > AACGAGACAGAATCCGCTACCCTACCCCTCGGCGAATATGTTCCCCCAACACCCAGTTGT 309
p41F        > AACGAGACAGAATCCGCTACCCTACCCCTCGGCGAATATGTTCCCCCAACACCCAGTTGT 405
CONSENSUS   > AACGAGACAGAATCCGCTACCCTACCCCTCGGCGAATATGTTCCCCCAACACCCAGTTGT 2760

                .........+.........+.........+.........+.........+.........+

p124hmg8    >                            ctgatgacgaaggtgatggtctgtccatgtcaaag 35
p41hmg11    > ACCGAGCCTGCTACGCCTGCATTGACTGATGACGAAGGTGATGGTCTGTCCATGTCAAAG 369
p41F        > ACCGAGCCTGCTACGCCTGCATTGACTGATGACGAAGGTGATGGTCTGTCCATGTCAAAG 465
CONSENSUS   > ACCGAGCCTGCTACGCCTGCATTGACTGATGACGAAGGTGATGGTCTGTCCATGTCAAAG 2820

                .........+.........+.........+.........+.........+.........+

p124hmg8    > ctcaGGTCCtCgcAaTCTCGGTCTcaGtTTGAGcatcGaTCAATTGAGGAACTGGAGAAG 95
p41hmg11    > CTCAGGTCCTCGCAATCTCGGTCTCAGTTTGAGCATCGATCAATTGAGGAACTGGAGAAG 429
p41F        > CTCAGGTCCTCGCAATCTCGGTCTCAGTTTGAGCATCGATCAATTGAGGAACTGGAGAAG 525
CONSENSUS   > CTCAGGTCCTCGCAATCTCGGTCTCAGTTTGAGCATCGATCAATTGAGGAACTGGAGAAG 2880

                .........+.........+.........+.........+.........+.........+

p6F         >                                GAGCTCAATGACGAGGAAGTTGTAACCATGTCTATG 36
p12hmg8     >                                GAGCTCAATGACGAGGAAGTTGTAACCATGTCTATG 36
p124hmg8    > TTGATCGTGGAGAAGCGAACCCATgagctcaatgacgaggaagttgtAacCATGtCTATg 155
p41hmg11    > TTGATCGTGGAGAAGCGAAcCCATGAGCTC                               459
p41F        > TTGATCGTGGAGAAGCGAACCCATGAGCTC                               555
CONSENSUS   > TTGATCGTGGAGAAGCGAACCCATGAGCTCAATGACGAGGAAGTTGTAACCATGTCTATG 2940

                .........+.........+.........+.........+.........+.........+

p6F         > CGTGGAAAGGTTCCCGGTTATGCACTCGAAAAGGCTCTGAAGGACTTTACTCGGGCCGTC 96
p12hmg8     > CGTGGAAAGGTTCCCGGTTATGCACTCGAAAAGGCTCTGAAGGACTTTACTCGGGCCGTC 96
p124hmg8    > CGTgGaAAGgTtcCCgGTtATGCACTCGAAaaGgCTCTGAaGGACTTtACTCGGgcCGTC 215
CONSENSUS   > CGTGGAAAGGTTCCCGGTTATGCACTCGAAAAGGCTCTGAAGGACTTTACTCGGGCCGTC 3000

                .........+.........+.........+.........+.........+.........+

p6F         > AAGATCCGTCGTACTATCATTTCTCGCACCAAAGCCACTTCCGAAATCACCAATGGACTT 156
p12hmg8     > AAGATCCGTCGTACTATCATTTCTCGCACCAAAGCCACTTCCGAAATCACCAATGGACTT 156
p124hmg8    > aAGATCcGTCGTACTATcatttctcgcacCAAAGCCACTTCCgAAATCACCAATgGACtT 275
CONSENSUS   > AAGATCCGTCGTACTATCATTTCTCGCACCAAAGCCACTTCCGAAATCACCAATGGACTT 3060

                .........+.........+.........+.........+.........+.........+

p6F         > GACCGATCCAAGCTTCCCTTCGAAAACTACAACTGGGAAAGGGTGTTTGGTGCCTGCTGC 216
p12hmg8     > GACCGATCCAAGCTTCCCTTCGAAAACTACAACTGGGAAAGGGTGTTTGGTGCCTGCTGC 216
p124hmg8    > GACCGATCCAAGCtTCCCTTCGaAAACTAcAACTgGGAAAGGGTGTTTGGTGCCTGCTGC 335
CONSENSUS   > GACCGATCCAAGCTTCCCTTCGAAAACTACAACTGGGAAAGGGTGTTTGGTGCCTGCTGC 3120

                .........+.........+.........+.........+.........+.........+

p6F         > GAGAATGTCATCGGCTACCTGCCCCTTCCCGTTGGCGTCGCCGGCCCGCTCGTTATTGAT 276
p12hmg8     > GAGAATGTCATCGGCTACCTGCCCCTTCCCGTTGGCGTCGCCGGCCCGCTCGTTATTGAT 276
p124hmg8    > GAGAATGTCATCGGCtaCCTGCCCCTTCCCGTTgGCGTCGCCGGCCCGCTCGTTATTGAT 395
CONSENSUS   > GAGAATGTCATCGGCTACCTGCCCCTTCCCGTTGGCGTCGCCGGCCCGCTCGTTATTGAT 3180

                .........+.........+.........+.........+.........+.........+
```

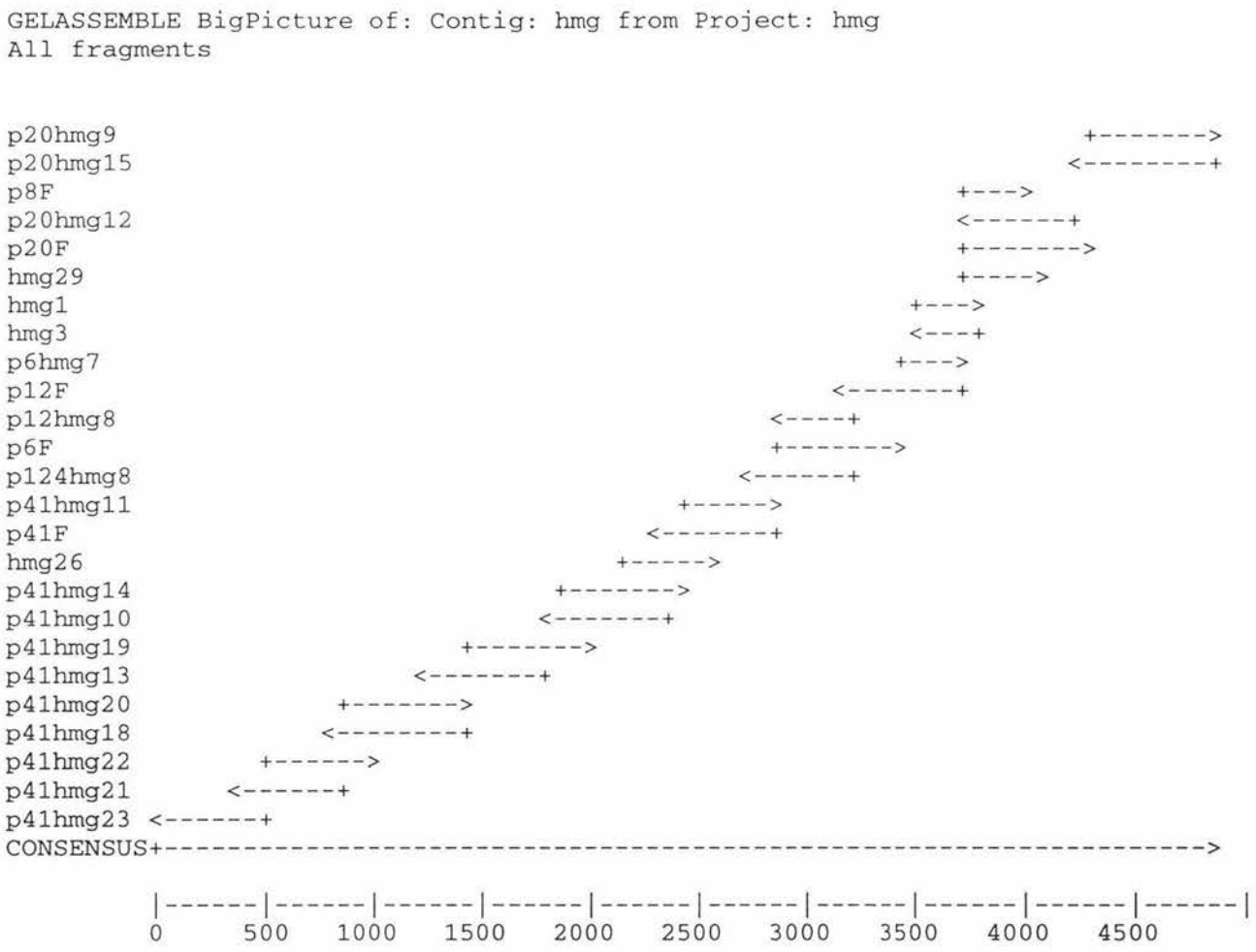
```
p12F        >          CTATTTCATTCCTATGGCCACGACGGAAGGTgTCTTGGTTGCCAGCACTAGT 52
p6F         > GGGCAGAGCTATTTCATTCCTATGGCCACGACGGAAGGTGTCTTGGTTGCCAGCACTAGT 336
p12hmg8     > GGGCAGAGCTATTTCATTCCTATGGCCACGACGGAAGGtGTCTTgGTTGCCAGCACT    333
p124hmg8    > GGGCAGAGCTAtttCAttCCTATGGCCACGACGgAAGGTGTCTTGGTT             443
CONSENSUS   > GGGCAGAGCTATTTCATTCCTATGGCCACGACGGAAGGTGTCTTGGTTGCCAGCACTAGT 3240

                .........+.........+.........+.........+.........+.........+

p12F        > CGTGGCTGCAAAGCGATCAACTCTGGTGGCGGTGCCATTACTGTCTTAACAAGCGATGGT 112
p6F         > CGTGGCTGCAAAGCGATCAACTCTGGTGGCGGTGCCATTACTGTCTTAACAAGCGATGGT 396
CONSENSUS   > CGTGGCTGCAAAGCGATCAACTCTGGTGGCGGTGCCATTACTGTCTTAACAAGCGATGGT 3300

                .........+.........+.........+.........+.........+.........+

p12F        > ATGACCCGTGGACCTTGCGTTAGTTTTGAGACTTTGGAGCGTGCCGGTGCCGCCAAGCTC 172
p6F         > ATGACCCGTGGACCTTGCGTTAGTTTTGAGACTTTGGAGCGTGCCGGTGCCGCCAAGCTC 456
CONSENSUS   > ATGACCCGTGGACCTTGCGTTAGTTTTGAGACTTTGGAGCGTGCCGGTGCCGCCAAGCTC 3360

                .........+.........+.........+.........+.........+.........+

p12F        > TGGCTGGATTCTGAGGCAGGCCAGAACACCATGAAGAAGGCTTTCAACTCGACCAGTCGG 232
p6F         > TGGCTGGATTCTGAGGCAGGCCAGAACACCATGAAGAAgGCTTTCAACTCGACCAGTCGG 516
CONSENSUS   > TGGCTGGATTCTGAGGCAGGCCAGAACACCATGAAGAAGGCTTTCAACTCGACCAGTCGG 3420

                .........+.........+.........+.........+.........+.........+

p6hmg7       >                                  tttggccggcaccaaCTTGTATATTCGATTC 31
p12F        > TTCGCTCGTCTACAGCACATGAAGACTGCTTTGGCCGGCACCAACTTGTATATTCGATTC 292
p6F         > TTCGCTCGTCTACAGCACATGAAGACTGCTTTGGCCGGCACCAACTTGT            565
CONSENSUS   > TTCGCTCGTCTACAGCACATGAAGACTGCTTTGGCCGGCACCAACTTGTATATTCGATTC 3480

                .........+.........+.........+.........+.........+.........+

hmg1        >                                                    AGCAtGCG 8
hmg3        >                                        atgatttcaaagggtgttgAGCATGCG 27
p6hmg7      > AAGACCACCACCGGAGACGCCATGGGCATGAACATGATTTCAAAGGGTGTTGAGCATGCG 91
p12F        > AAGACCACCACCGGAGACGCCATGGGCATGAACATGATTTCAAAGGGTGTTGAGCATGCG 352
CONSENSUS   > AAGACCACCACCGGAGACGCCATGGGCATGAACATGATTTCAAAGGGTGTTGAGCATGCG 3540

                .........+.........+.........+.........+.........+.........+

hmg1        > CTCAACGTCAtGGCCACCGACGGtggctttGATGACATGAACATCATCACCGTCTCTGGC 68
hmg3        > CTCAACGTCATGgccACCGACGGTGGCTTTGATGACATGAACATCATCACCGTCTCTGGC 87
p6hmg7      > CTCAACGTCATGGCCACCGACGGTGGCTTTGATGACATGAACATCATCACCGTCTCTGGC 151
p12F        > CTCAACGTCATGGCCACCGACGGTGGCTTTGATGACATGAACATCATCACCGTCTCTGGC 412
CONSENSUS   > CTCAACGTCATGGCCACCGACGGTGGCTTTGATGACATGAACATCATCACCGTCTCTGGC 3600

                .........+.........+.........+.........+.........+.........+

hmg1        > AACTTCTGTATTGACAAGAAACCCGcCGcTATGAACTGGATCGACGGTCGCGGCAAGGGC 128
hmg3        > AACTTCTGTATTGACAAGAAACCCGCCGCTATGAACTGGATCGACGGTCGCGGcaAGGGC 147
p6hmg7      > AACTTCTGTATTGACAAGAAACCCGCCGCTATGAACTGGATCGACGGTCGCGGCAAGGGC 211
p12F        > AACTTCTGTATTGACAAGAAACCCGCCGCTATGAACTGGATCGACGGTCGCGGCAAGGGC 472
CONSENSUS   > AACTTCTGTATTGACAAGAAACCCGCCGCTATGAACTGGATCGACGGTCGCGGCAAGGGC 3660

                .........+.........+.........+.........+.........+.........+
```

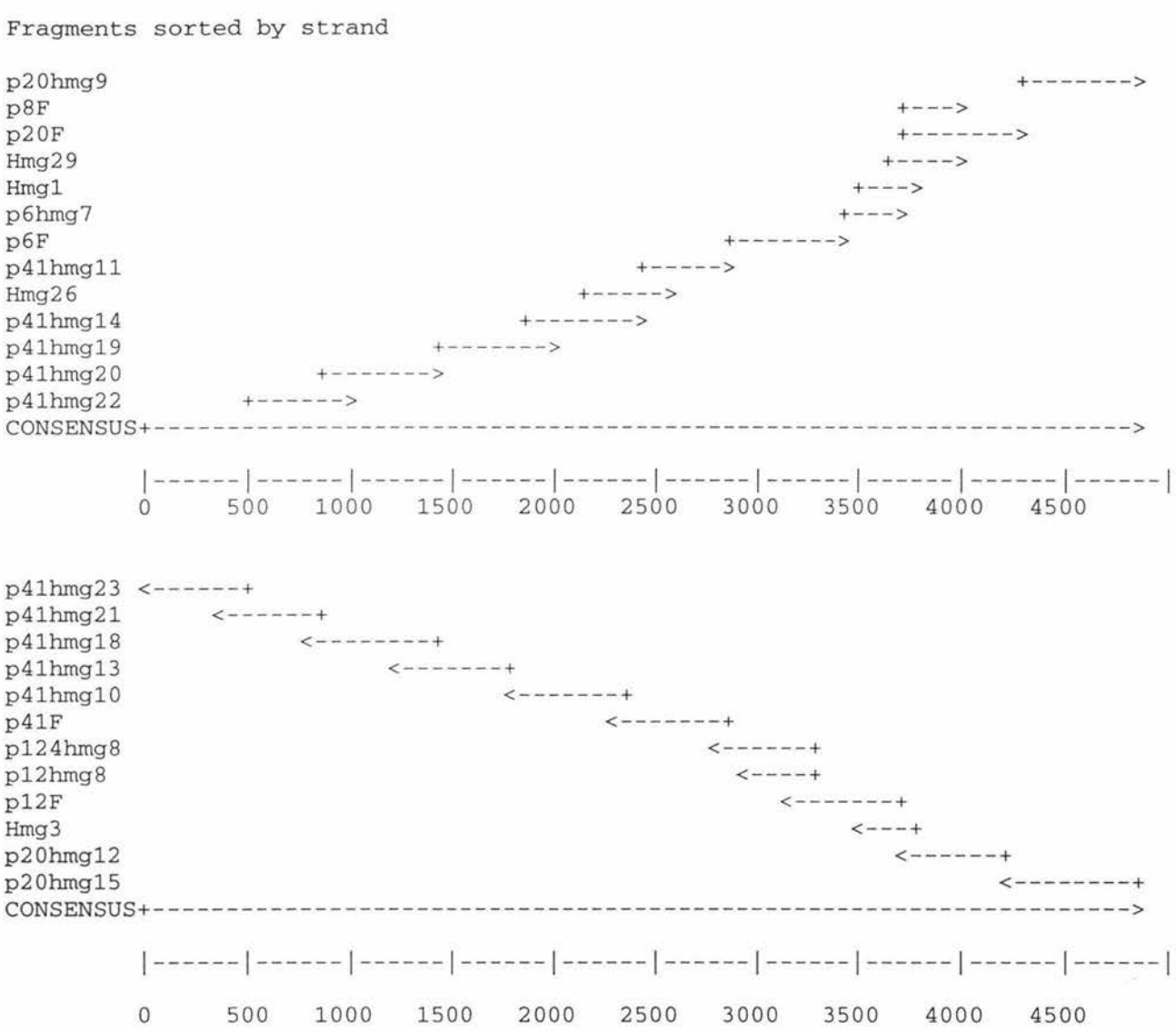
```
hmg29       >                                                          GTT 3
hmg1        > ATTGTTGCTGAGGCCATCATTcCTGCCGACGTGGTCAAATCCGTGTTGAAGAGTGATGTT 188
hmg3        > ATTGTTGCTGAGGCCATCATTCCTGCCGACGTGGTCAAATCCGTGTTGAAGAGTGATGTT 207
p6hmg7      > ATTGTTGCTGAGGCCATCATTCCTGCCGACGTGGTCAAATCCGTGTTGAAGaGTGATGTT 271
p12F        > ATTGTTGCTGAGGCCATCATTCCTGCCGACGTGGTCAAATCCGTGTTGAAGAGTGATGTT 532
CONSENSUS   > ATTGTTGCTGAGGCCATCATTCCTGCCGACGTGGTCAAATCCGTGTTGAAGAGTGATGTT 3720

                .........+.........+.........+.........+.........+.........+

p20hmg12    >              GAGcTCAATATTGCTAAGAACCTGATCGGGTCTGCCATGGCCGGTTCG 48
p20F        >              GAGCTCAATATTGCTAAGAACCTGATCGGGTCTGCCATGGCCGGTTCG 48
hmg29       > GATGCTTTGGTagAGCTCAATATTGCTAAGAAcCTGATCGGGTCTGCCATGGCCGGTTCG 63
hmg1        > GATGCTTTGGTAGAGCTCAATATTGCTAAGAACCTGATCGGGTCTGCCATGGCCGGTTCG 248
hmg3        > GATGCTTTGGTAGAGCTCAATATTGCTAAGAACCTGATCGGGTCTGCCATGGCCGGTTCG 267
p6hmg7      > GATGCTTTGGTAGAGCTC                                           289
p12F        > GATGCTTTGGTAGAGCTC                                           550
CONSENSUS   > GATGCTTTGGTAGAGCTCAATATTGCTAAGAACCTGATCGGGTCTGCCATGGCCGGTTCG 3780

                .........+.........+.........+.........+.........+.........+

p20hmg12    > ATCGGTGGcTTCAACGCCCACGCTGCCAACATcGTAGCGGCCATATTCcTCGCCACCGGC 108
p20F        > ATCGGTGGCTTCAACGCCCACGCTGCCAACATCGTAGCGGCCATATTCCTCGCCACCGGC 108
hmg29       > ATCGGTGGCTTCAACGcCCACGCTGCCAACATCGTAGCGGCCATATTCCTCGcCACCGGC 123
hmg1        > ATCGGTGGCTTCAACGcCCACGCTGCCAACATCGTAGCGGCCATATTCCTCGcCACCGGC 308
hmg3        > ATCGGTGGCTTCAACGCCCACGCTGCCAACATCGTA                          303
CONSENSUS   > ATCGGTGGCTTCAACGCCCACGCTGCCAACATCGTAGCGGCCATATTCCTCGCCACCGGC 3840

                .........+.........+.........+.........+.........+.........+

p20hmg12    > CAAGATCCCGCACAGGTAGTAGAGAGCTGCAACTGCATCACCACCATGAAGAAGTGAGTG 168
p20F        > CAAGATCCCGCACAGGTAGTAGAGAGCTGCAACTGCATCACCACCATGAAGAAGTGAGTG 168
hmg29       > CAAGATCCCGCACAGGTAGTAGAGAGCTGCAACTGCATCACCACCATGAAgAA....... 183
hmg1        > CAAGAtCccgc                                                  319
CONSENSUS   > CAAGATCCCGCACAGGTAGTAGAGAGCTGCAACTGCATCACCACCATGAAGAAGTGAGTG 3900

                .........+.........+.........+.........+.........+.........+

p20hmg12    > CCAGCCTTGTTTCTCCCGTTCTTTGAAAGCATGATTCATTTGTACTAACGACGATCTTTG 228
p20F        > CCAGCCTTGTTTCTCCCGTTCTTTGAAAGCATGATTCATTTGTACTAACGACGATCTTTG 228
hmg29       > ..............................................................  243
CONSENSUS   > CCAGCCTTGTTTCTCCCGTTCTTTGAAAGCATGATTCATTTGTACTAACGACGATCTTTG 3960

                .........+.........+.........+.........+.........+.........+

p20hmg12    > CTCTAGCCTTCACGGATCGCTCCAGATTGCCGTTTCCATGCCGTCTCTCGAGGTCGGCAC 288
p20F        > CTCTAGCCTTCACGGATCGCTCCAGATTGCCGTTTCCATGCCGTCTCTCGAGGTCGGCAC 288
hmg29       > ......CCTTCACGGATCGCTCCAGATTGCCGTTTCCATGCCGTCTCTCGAGGTCGGCAC 303
CONSENSUS   > CTCTAGCCTTCACGGATCGCTCCAGATTGCCGTTTCCATGCCGTCTCTCGAGGTCGGCAC 4020

                .........+.........+.........+.........+.........+.........+

p20hmg12    > CCTCGGCGGCGGCACCATCTTGGAGCCTCAGAGCGCTATGCTGGACATGCTCGGTGTGCG 348
p20F        > CCTCGGCGGCGGCACCATCTTGGAGCCTCAGAGCGCTATGCTGGACATGCTCGGTGTGCG 348
hmg29       > CCTCGGCGGCGGCACCATCTTGGAGCCTCAGAgCGCTATGCTGGACATGCTCGGTGTGCG 363
CONSENSUS   > CCTCGGCGGCGGCACCATCTTGGAGCCTCAGAGCGCTATGCTGGACATGCTCGGTGTGCG 4080

                .........+.........+.........+.........+.........+.........+
```

```
p20hmg12    > AGGATCACATCCAACCAATCCAGGAGACAATGCCCGACGACTTGCACGCATCATTGGTGC 408
p20F        > AGGATCACATCCAACCAATCCAGGAGACAATGCCCGACGACTTGCACGCATCATTGGTGC 408
hmg29       > AGGATCACATCCA                                                378
CONSENSUS   > AGGATCACATCCAACCAATCCAGGAGACAATGCCCGACGACTTGCACGCATCATTGGTGC 4140

                .........+.........+.........+.........+.........+.........+

p20hmg12    > TTCCGTCTTGGCTGGTGAGCTGTCTCTTTGCAGTGCTCTCCAGGCCGGCCACCTAGTCAA 468
p20F        > TTCCGTCTTGGCTGGTGAGCTGTCTCTTTGCAGTGCTCTCCAGGCCGGCCACCTAGTCAA 468
CONSENSUS   > TTCCGTCTTGGCTGGTGAGCTGTCTCTTTGCAGTGCTCTCCAGGCCGGCCACCTAGTCAA 4200

                .........+.........+.........+.........+.........+.........+

p20hmg12    > AGCGCACATGCAGCACAACCGAAGCGCCGCCCCATCAAGAAGCACTACACCCGCTCCTCC 528
p20F        > AGCGCACATGCAGCACAACCGAAGCGCCGCCCCATCAAGAAGCACTACACCCGCTCCTCC 528
CONSENSUS   > AGCGCACATGCAGCACAACCGAAGCGCCGCCCCATCAAGAAGCACTACACCCGCTCCTCC 4260

                .........+.........+.........+.........+.........+.........+

p20hmg9     >                                                       AAAgCGC 7
p20hmg15    >                    tcactggccatGACCATTGcTCAAGACAAGTcGAGCAAAAGCGc 44
p20hmg12    > TCCCATGAcgCCcgTCTCActggcca                                   554
p20F        > TCCCATGACGCCCGTCTCACTGGCCATGACCATTGCTCAAGACAAGTCGAGCAAAAGCGC 588
CONSENSUS   > TCCCATGACGCCCGTCTCACTGGCCATGACCATTGCTCAAGACAAGTCGAGCAAAAGCGC 4320

                .........+.........+.........+.........+.........+.........+

p20hmg9     > TGCGGCCcAACAGCGGTCAAAGCGGTAAACAAAAGTATCAAGATGCCACGCAAGGCAAGG 67
p20hmg15    > TGCGGCCCAACAGCGGTCAAAGCGGTAAACAAAAGTATCAAGATGCCACGCAAGGCAAGG 104
p20F        > TGCGGCCCAAC                                                  599
CONSENSUS   > TGCGGCCCAACAGCGGTCAAAGCGGTAAACAAAAGTATCAAGATGCCACGCAAGGCAAGG 4380

                .........+.........+.........+.........+.........+.........+

p20hmg9     > AAGGGGTGGGAATAAAGATGATAGAAGAGAGGGTGTTGGAGTTTAGATTTCGCATTCCGG 127
p20hmg15    > AAGGGGTGGGAATAAAGATGATAGAAGAGAGGGTGTTGGAGTTTAGATTTCGCATTCCGG 164
CONSENSUS   > AAGGGGTGGGAATAAAGATGATAGAAGAGAGGGTGTTGGAGTTTAGATTTCGCATTCCGG 4440

                .........+.........+.........+.........+.........+.........+

p20hmg9     > GAGTTCATTATTGACTACGCGCAGGGCCTGACATACAATAGCACCTACCTATATAGACGA 187
p20hmg15    > GAGTTCATTATTGACTACGCGCAGGGCCTGACATACAATAGCACCTACCTATATAGACGA 224
CONSENSUS   > GAGTTCATTATTGACTACGCGCAGGGCCTGACATACAATAGCACCTACCTATATAGACGA 4500

                .........+.........+.........+.........+.........+.........+

p20hmg9     > TATACTGCTACATATAAGGCCTGAGGGTGGGTTTCCAGAGATCTTAGACCTCCACGGTCA 247
p20hmg15    > TATACTGCTACATATAAGGCCTGAGGGTGGGTTTCCAGAGATCTTAGACCTCCACGGTCA 284
CONSENSUS   > TATACTGCTACATATAAGGCCTGAGGGTGGGTTTCCAGAGATCTTAGACCTCCACGGTCA 4560

                .........+.........+.........+.........+.........+.........+

p20hmg9     > TTATAATCGTTACTTTGTTTCACTGCTACAATTCTGTCCAGCTTCGATATCTCCTGTACA 307
p20hmg15    > TTATAATCGTTACTTTGTTTCACTGCTACAATTCTGTCCAGCTTCGATATCTCCTGTACA 344
CONSENSUS   > TTATAATCGTTACTTTGTTTCACTGCTACAATTCTGTCCAGCTTCGATATCTCCTGTACA 4620

                .........+.........+.........+.........+.........+.........+
```

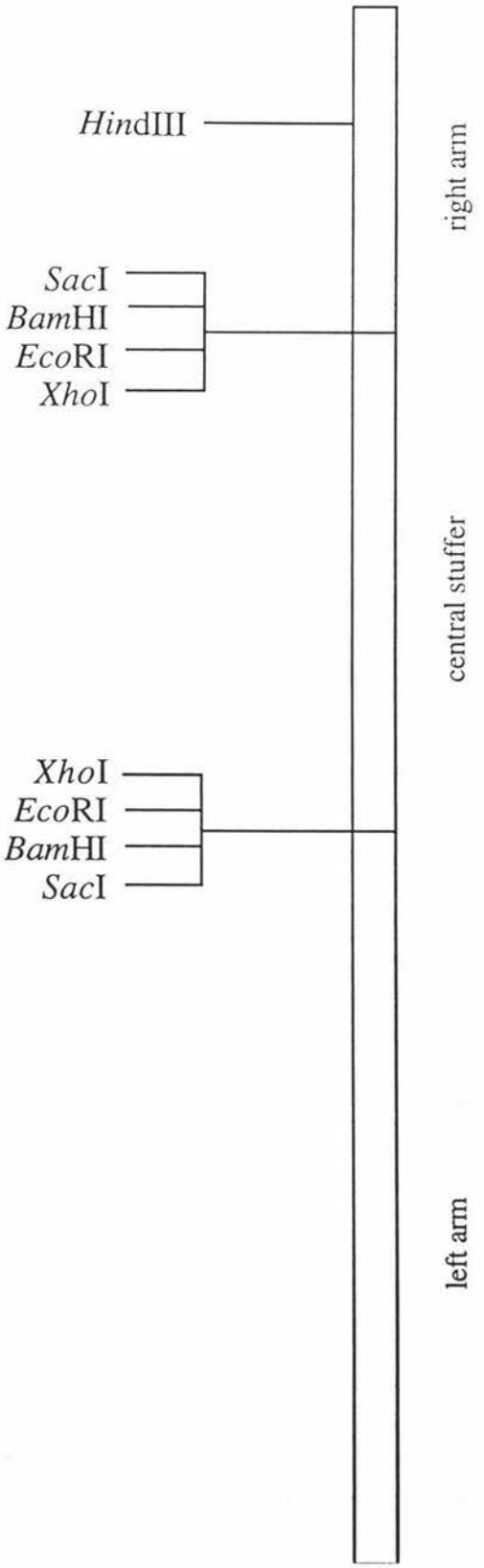
```
p20hmg9     > TGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTC 367
p20hmg15    > TGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTC 404
CONSENSUS   > TGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTC 4680

              .........+.........+.........+.........+.........+.........+

p20hmg9     > TGTCTGTCTGTCCATCTGTAACTGTCCCTCACACCAGCGCCTTGACTCGTGCAATCTCTT 427
p20hmg15    > TGTCTGTCTGTCCATCTGTAACTGTCCCTCACACCAGCGCCTTGACTCGTGCAATCTCTT 464
CONSENSUS   > TGTCTGTCTGTCCATCTGTAACTGTCCCTCACACCAGCGCCTTGACTCGTGCAATCTCTT 4740

              .........+.........+.........+.........+.........+.........+

p20hmg9     > GTCCATCCATGGAAATCATTTTGACACCCTTGTTCATAATCTCTCGTCTGCATTCATCCC 487
p20hmg15    > GTCCATCCATGGAAATCATTTTGACACCCTTGTTCATAATCTCTCGTCTGCATTCATCCC 524
CONSENSUS   > GTCCATCCATGGAAATCATTTTGACACCCTTGTTCATAATCTCTCGTCTGCATTCATCCC 4800

              .........+.........+.........+.........+.........+.........+

p20hmg9     > ACTGCTCCGGCATGACAGGCTTTGTAgCCTCCTCGACAATGTACGACCTAAAACCTTGCT 547
p20hmg15    > ACTGCTCCGGCATGACAGGCTTTGTAGCCTCCTCGACAATGTACGACCTAAAACCTTGCT 584
CONSENSUS   > ACTGCTCCGGCATGACAGGCTTTGTAGCCTCCTCGACAATGTACGACCTAAAACCTTGCT 4860

              .........+.........+.........+.........+.........+.........+

p20hmg9     > CGATGGCATGCTCCGCCGTA                                       567
p20hmg15    > CGATGGCATGCTCCGCCGTAGCTTT                                  609
CONSENSUS   > CGATGGCATGCTCCGCCGTAGCTTT                                  4885
```

A 1.4 BIGPICTURE from *hmg* Contig




Fragments sorted by strand
p20hmg9
p8F
p20F
Hmg29
Hmg1
p6hmg7
p6F
p41hmg11
Hmg26
p41hmg14
p41hmg19
p41hmg20
p41hmg22
CONSENSUS
0 500 1000 1500 2000 2500 3000 3500 4000 4500
p41hmg23
p41hmg21
p41hmg18
p41hmg13
p41hmg10
p41F
p124hmg8
p12hmg8
p12F
Hmg3
p20hmg12
p20hmg15
CONSENSUS
0 500 1000 1500 2000 2500 3000 3500 4000 4500

Appendix 2.0 λGEM-12

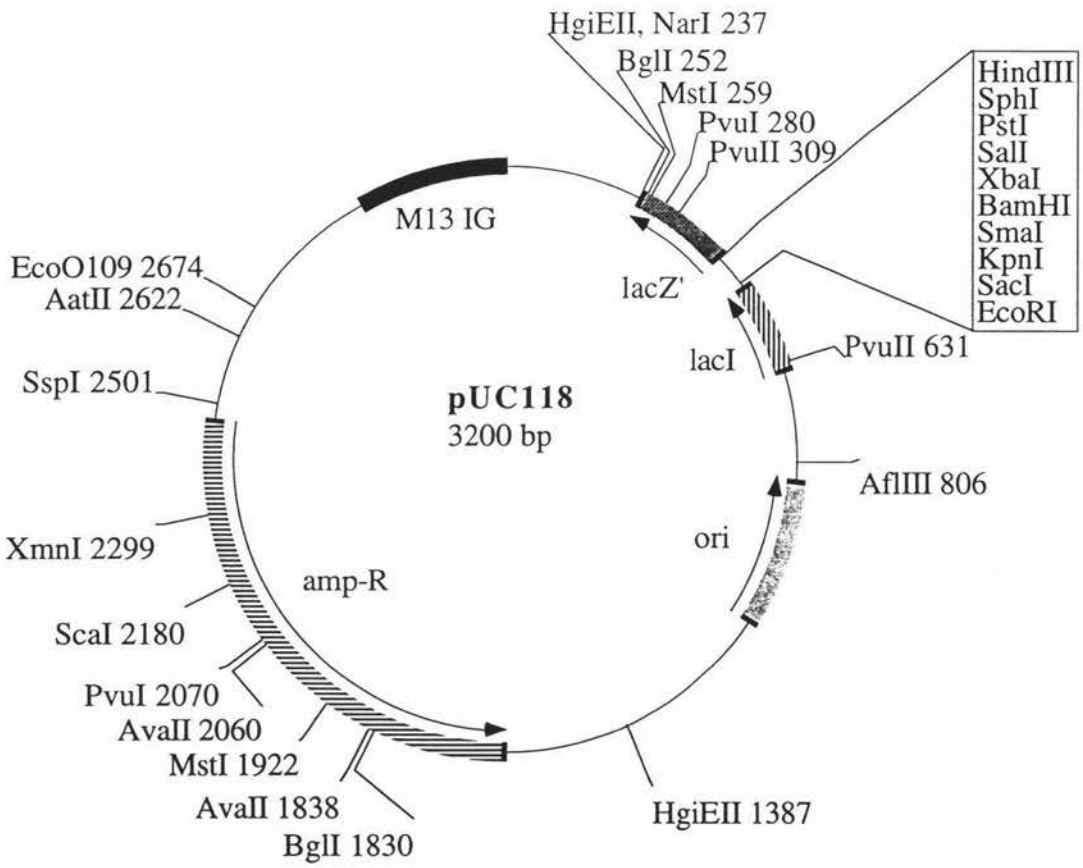


Restriction map showing *Sac*I, *Bam*HI, *Eco*RI and *Xho*I within the multiple cloning region of λGEM-12, and *Hind*III within the right λ arm.

Appendix 3.0 VECTOR MAPS

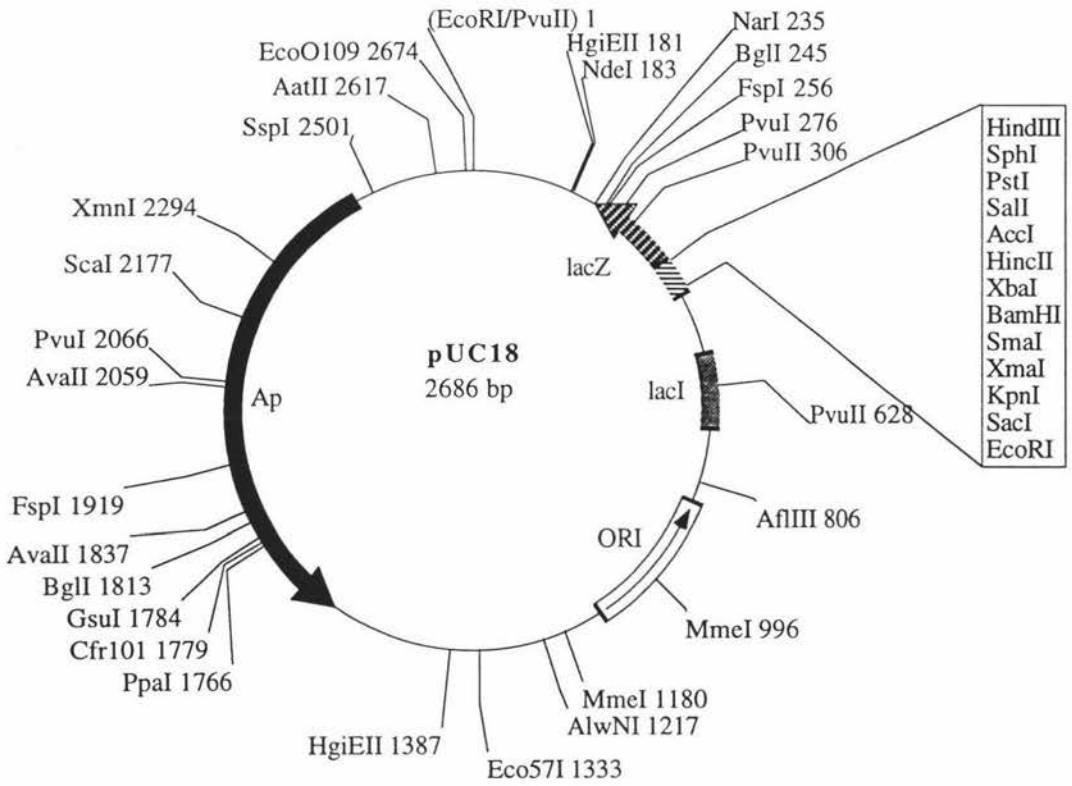
A 3.1 pUC118

Restriction map of pUC118 showing all restriction sites.



A 3.2 pUC18

Restriction map of pUC18 showing all restriction sites.



Appendix 4.0 RT-PCR ANALYSIS OF THE TRANSCRIPTION START SITE

A 4.1 RNA Extraction from Lp19 Cultures

All precautions outlined in Section 2.18 were used throughout. In a pre-cooled mortar approximately 500 mg of frozen Lp19 mycelia was ground to a fine powder under liquid nitrogen and resuspended in 15 ml of extraction buffer (50 mM Tris-HCl [pH 8.0], 300 mM sodium chloride, 5 mM EDTA, 2% [w/v] SDS, 0.5% [w/v] polyvinylpyrrolidone (PVP) and 0.5 mM aurintricarboxylic acid [pH 8.0]). The solution was mixed, incubated at 65°C for 10 minutes, with occasional shaking and centrifuged for 15 minutes at 10,000 g. The supernatant was transferred to 0.7 ml of 3 M potassium acetate (pH 4.8), mixed, incubated on ice for 30 minutes before centrifugation at 10,000 g for 10 minutes at 4°C. To the resulting supernatant 5 ml of 8 M lithium chloride was added the mixture incubated at 4°C overnight . Following this the mixture was re-centrifuged at 12,000 g for 30 minutes at 4°C. The pellet was washed twice with 10 ml of 3 M sodium acetate (pH 5.2) by resuspending the pellet and further centrifugation at 10,000 g for 10 minutes at room temperature. The final pellet was resuspended in 5 ml of water. The RNA was phenol/chloroform extracted as in Section 2.6 except each chloroform extraction also contained an equal volume of isopropanol. The aqueous phase was added to 0.5 ml of 3 M sodium acetate (pH 5.2) and 13 ml absolute ethanol and incubated overnight at -20°C. The mixture was centrifuged at 10,000 g for 10 minutes at 4°C, the pellet washed in 70% ethanol and air dried for 10 minutes. The RNA was resuspended in 50 µl of water and the RNA concentration determined to be approximately 2.5 mg/ml (Section 2.20.1).

A 4.2 DNase Treatment of RNA

Prior to cDNA synthesis and PCR amplification of the cDNA it was necessary to remove all traces of contaminating DNA from the RNA preparation. To 25 µg of Lp19 RNA the following was added: 1 x DNase buffer (40 mM Tris-HCl [pH 7.5], 6 mM $MgCl_2.6H_2O$ and 2 mM $CaCl_2.2H_2O$), 200 U DNase I (Stratagene), 40 U RNase Block (Stratagene) and water to 50 µl. The mixture was incubated at 37°C for 15 minutes and the DNase reaction stopped by addition of 25 µl DNase stop mix (50 mM EDTA, 1.5 mM sodium acetate and 1% [w/v] SDS) and the RNA extracted with phenol/chloroform/ isoamyl alcohol as in Section A 4.1.

A 4.3 Reverse Transcriptase Synthesis of cDNA

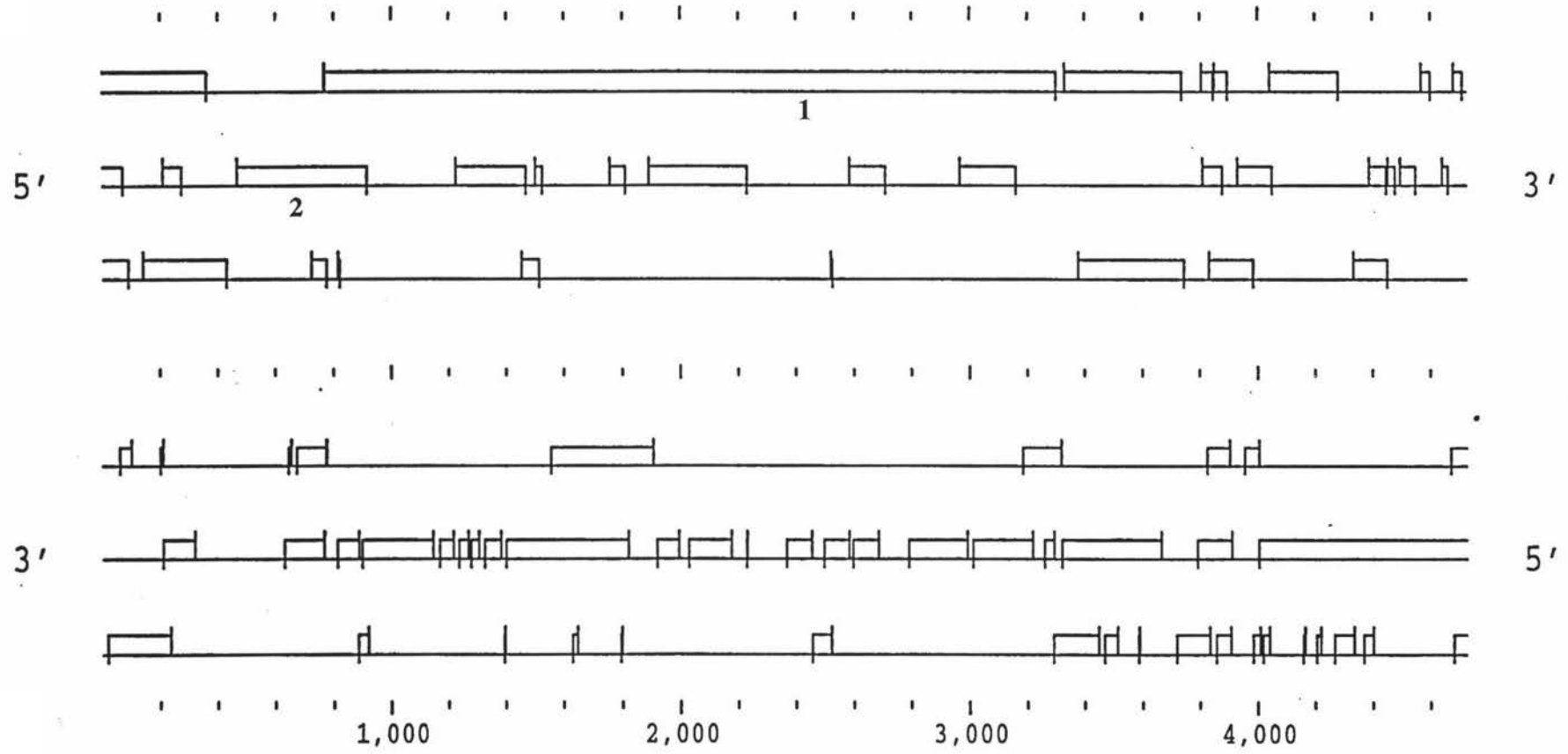
RT-PCR was carried out on Lp19 RNA using the Stratagene RT-PCR kit. In a microcentrifuge tube 300 ng of BRLs random primers and 5 µg of Lp19 RNA in a total volume of 41 µl, were mixed and incubated at 65°C for 5 minutes and slowly cooled to room temperature to allow the primers to anneal to the RNA. To this 1 x first strand buffer, 40 U of RNase Block Ribonuclease Inhibitor, 4 mM of each dNTP, 50 U of MMLV-RT and water to 50 µl was added and the mixture incubated at 37°C for 1 hour and then at 90°C for 5 minutes. The reaction was stored at -20°C.

A 4.4 Amplification of cDNA by PCR

To 2 µl of the cDNA from Section A 4.2, 3 µM of primer hmg 13 and 3 µM of the appropriate primer (hmg 22, H-2, hmg 20, H-3, or H-4), 1 x PCR buffer, 2.5 U Expand High Fidelity (Boehringer Mannheim), 0.5 mM of each dNTP and water to 50 µl were added. The reaction was then placed in a Corbett thermocycler preheated to 94°C and after an initial 2 minute melt at 94°C was subjected to 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 90 seconds. After the 35 cycles were complete the reactions were incubated at 72°C for 7 minutes, then stored at -20°C. The products were then separated by electrophoresis as in Section 2.10.

Appendix 5.0 A 6 FRAME TRANSLATION OF *G. FUJIKUROI HMG* SHOWING ALL OPEN READING FRAMES.

The FRAMES program of the GCG package plots the open reading frames of a nucleic acid sequence as boxes bordered by potential start and stop codons. Potential start codons are shown as short lines that extend above the box and potential stop codons are shown as short lines that extend below the box. By default, only the start and stop codons at the ends of open reading frames are shown in the frame display; if a stop codon has been passed, no stops are shown again until a start codon is passed; if a start codon is passed, no start codons are shown again until a stop codon is passed.


1
5′
2
3′
3′
5′
1,000
2,000
3,000
4,000

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. **Journal of Molecular Biology 215:** 403-410.

Basson, M. E., Thorsness, M., Finer-Moore, J., Stroud, R. M. and Rine, J. (1988). Structural and functional conservation between yeast and human 3-hydroxy-3-methylglutaryl coenzyme A reductases, the rate-limiting enzyme of sterol biosynthesis. **Molecular and Cellular Biology 8:** 3797-3808.

Basson, M. E., Thorsness, M. and Rine, J. (1986). *Saccharomyces cerevisiae* contains two functional genes encoding 3-hydroxy-3-methylglutaryl-coenzyme A reductase. **Proceedings of the National Academy of Sciences of the USA 83:** 5563-5567.

Bischoff, K. M. and Rodwell, V. W. (1997). 3-Hydroxy-3-methylglutaryl-coenzyme A reductase of *Haloferax volcanii*: Role of histidine 398 and attenuation of activity by introduction of negative charge at position 404. **Protein Science 6:** 156-161.

Bochar, D. A., Brown, J. R., Doolittle, W. F., Klenk, H.-P., Lam, W., Schenk, M. E., Stauffacher, C. V. and Rodwell, V. W. (1997). 3-Hydroxy-3-methylglutaryl Coenzyme A reductase of *Sulfolobus solfataricus*: DNA sequence, phylogeny, expression in *Escherichia coli* of the *hmgA* gene, and purification and kinetic characterisation of the gene product. **Journal of Bacteriology 179:** 3632-3638.

Boehringer Mannheim (1995). **The DIG System User's Guide for Filter Hybridisation.** Eds. R. van Miltenburg, B. Rüger, S. Grünewald-Janho, M, Leons and C, Schröder. Germany. Boehringer Mannheim GmbH, Biochemica.

Bruchez, J. J. P., Eberle, J. and Russo, V. A. E. (1993). Regulatory sequences in the transcription of *Neurospora crassa* genes: CAAT box, TATA box, introns, poly(A) tail formation sequences. **Fungal Genetics Newsletter 40:** 89-96.

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

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

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*). **Mycological Research 97:** 1083-1092.

Clarke, L. and Carbon, J. (1976). A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire *E. coli* genome. **Cell 9:** 91-99.

Clay, K. (1990). Fungal endophytes of grasses. **Annual Review of Ecology and Systematics 21:** 275-297.

Collett, M. A. (1994). Cloning, characterisation and evolutionary relationships of two *pyr*4 genes from an *Acremonium* endophyte of perennial ryegrass. Massey.

Croxen, R., Goosey, M. W., Keon, J. P. R. and Hargreaves, J. A. (1994). Isolation of an *Ustilago maydis* gene encoding 3-hydroxy-3-methylglutaryl-coenzyme A reductase and expression of a C-terminal-truncated form in *Escherichia coli*. **Microbiology 140:** 2363-2370.

Darnay, B. G., Wang, Y. and Rodwell, V. W. (1992). Identification of the catalytically important histidine of 3-hydroxy-3-methylglutaryl-coenzyme A reductase. **The Journal of Biological Chemistry 267:** 15064-15070.

Denbow, C. J., Lang, S. and Cramer, C. L. (1996). The N-terminal domain of tomato 3-hydroxy-3-methylglutaryl-CoA reductases: sequence, microsomal targeting and glycosylation. **The Journal of Biological Chemistry 271:** 9710-9715.

Dower, W. J., Miller, J. F. and Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. **Nucleic Acids Research 16:** 6127-6145.

Drew, S. W. and Demain, A. L. (1977). Effect of primary metabolites on secondary metabolism. **Annual Review of Microbiology 31:** 343-356.

Dugaiczyk, A., Boyer, H. W. and Goodman, H. M. (1975). Ligation of *Eco*RI endonuclease-generated DNA fragments in linear and circular structures. **Journal of Molecular Biology 96:** 171-184.

Edelmann, S. E. and Staben, C. (1994). A statistical analysis of sequence features within genes from *Neurospora crassa*. **Experimental Mycology 18:** 70-81.

Enjuto, M., Balcells, L., Campos, N., Caelles, C., Arró, M. and Boronat, A. (1994). *Arabidopsis thaliana* contains two differentially expressed 3-hydroxy-3-methylglutaryl-CoA reductase genes, which encode microsomal forms of the enzyme. **Proceedings of the National Academy of Sciences of the USA 91:** 927-931.

Friesen, J. A. and Rodwell, V. W. (1997). Identification of elements critical for phosphorylation of 3-hydroxy-3-methylglutaryl coenzyme A reductase by adenosine monophosphate-activated protein kinase: Protein engineering of the naturally nonphosphorylatable 3-hydroxy-3-methylglutaryl coenzyme A reductase from *Pseudomonas mevalonii*. **Biochemistry 36:** 1157-1162.

Frimpong, K. and Rodwell, V. W. (1994). Catalysis by Syrian hamster 3-hydroxy-3-methylglutaryl-coenzyme A reductase: proposed roles of histidine 865, glutamate 558, and aspartate 766. **The Journal of Biological Chemistry 269:** 11478-11483.

Gallagher, R. T., Campbell, A. G., Hawkes, A. D., Holland, P. T., McGaveston, D. A. and Pansier, E. A. (1982). Ryegrass Staggers: the presence of lolitrem neurotoxins in perennial ryegrass seed. **New Zealand Veterinary Journal 30:** 183-184.

Gallagher, R. T., White, E. P. and Mortimer, P. H. (1981). Ryegrass Staggers: isolation of potent neurotoxins lolitrem A and lolitrem B from staggers-producing pastures. **New Zealand Veterinary Journal 29:** 189-190.

Gant, D. B., Cole, R. J., Valdes, J. J., Eldefrawi, M. E. and Eldefrawi, A. T. (1987). Action of tremorgenic mycotoxins on $GABA_A$ receptor. **Life Sciences 41:** 2207-2214.

Gil, G., Faust, J. R., Chin, D. J., Goldstein, J. L. and Brown, M. S. (1985). Membrane-bound domain of HMG CoA reductase is required for sterol-enhanced degradation of the enzyme. **Cell 41:** 249-258.

Goldstein, J. L. and Brown, M. S. (1990). Regulation of the mevalonate pathway. **Nature 343:** 425-430.

Gurr, S. J., Unkles, S. E. and Kinghorn, J. R. (1987). The structure and organization of nuclear genes of filamentous fungi. In: **Gene Structure in Eukaryotic Microbes.** Eds. J. K. Kinghorn. Oxford, IRL Press. 93-139.

Hampton, R., Dimster-Denk, D. and Rine, J. (1996). The biology of HMG-CoA reductase: the pros of contra-regulation. **Trends in Biochemical Sciences 21:** 140-145.

Herd, S., Christensen, M. J., Saunders, K., Scott, D. B. and Schmid, J. (1997). Quantitative assessment of *in planta* distribution of metabolic activity and gene expression of an endophytic fungus. **Microbiology 143:** 267-275.

Hodge, M. R. and Cumsky, M. G. (1989). Splicing of a yeast intron containing an unusual 5' junction sequence. **Molecular and Cellular Biology 9:** 2765-2770.

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

Innis, M. A. and Gelfand, D. H. (1990). Optimisation of PCRs. In: **PCR Protocols: A Guide to Methods and Applications.** Eds. M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White. San Diego, Academic Press, Inc.

Itakura, K., Rossi, J. J. and Wallace, R. B. (1984). Synthesis and use of synthetic oligonucleotides. **Annual Review of Biochemistry 53:** 323-356.

Itoh, Y., Johnson, R. and Scott, B. (1994). Integrative transformation of the mycotoxin producing fungus, *Penicillium paxilli.* **Current Genetics 25:** 508-513.

Kawasaki, E. S. (1990). Amplification of RNA. In: **PCR Protocols: A Guide to Methods and Applications.** Eds. M. A. Innis, J. J. Sninsky and T. J. White. San Diego, Academic Press, Inc. 21-27.

Keogh, R. G., Tapper, B. A. and Fletcher, R. H. (1996). Distributions of the fungal endophyte *Acremonium lolii*, and of the alkaloids lolitrem B and peramine, within perennial ryegrass. **New Zealand Journal of Agricultural Research 39:** 121-127.

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.

Kozak, M. (1981). Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. **Nucleic Acids Research 9:** 5233-5252.

Kozak, M. (1988). Leader length and secondary structure modulate mRNA function under conditions of stress. **Molecular and Cellular Biology 8:** 2737-2744.

Kumagai, H., Chun, K. T. and Simoni, R. D. (1995). Molecular dissection of the role of the membrane domain in the regulated degradation of 3-hydroxy-3-methylglutaryl coenzyme A reductase. **The Journal of Biological Chemistry 270:** 19107-19113.

Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. **Journal of Molecular Biology 157:** 105-132.

Latch, G. C. M., Christensen, M. J. and Samuels, G. J. (1984). Five endophytes of Lolium and Festuca in New Zealand. **Mycotaxon 20:** 535-550.

Lawrence, C. M., Rodwell, V. M. and Stauffacher, C. V. (1995). Crystal structure of *Pseudomonas mevalonii* HMG-CoA reductase at 3.0 Angstrom resolution. **Science 268:** 1758-1762.

Lewin, B. (1990). **Genes IV**. Cell Press, Cambridge. Oxford University Press, Oxford.

Liscum, L., Finer-Moore, J., Stroud, R. M., Luskey, K. L., Brown, M. S. and Goldstein, J. L. (1985). Domain structure of 3-hydroxy-3-methylglutaryl Coenzyme A reductase, a glycoprotein of the endoplasmic reticulum. **The Journal of Biological Chemistry 260:** 522-530.

Lum, P. Y., Edwards, S. and Wright, R. (1996). Molecular, functional and evolutionary characterisation of the gene encoding HMG-CoA reductase in the fission yeast, *Schizosaccharomyces pombe*. **Yeast 12:** 1107-1124.

Luskey, K. L. and Stevens, B. (1985). Human 3-hydroxy-3-methylglutaryl coenzyme A reductase: conserved domains responsible for catalytic activity and sterol-regulated degradation. **The Journal of Biological Chemistry 260:** 10271-10277.

Miles, C. O., Wilkens, A. L., Gallagher, R. T., Hawkes, A. D., Munday, S. C. and Towers, N. R. (1992). Synthesis and tremorgenicity of paxitriols and lolitriol: possible biosynthetic precursors of lolitrem B. **Journal of Agriculture and Food Chemistry 40:** 234-238.

Miller, J. H. (1972). Experiments in molecular genetics. **Cold Spring Harbour Laboratory, Cold Spring Harbour, NY.**

Munday-Finch, S. C., Wilkins, A. L. and Miles, C. O. (1996). Isolation of Paspaline B, an indole-diterpenoid from *Penicillium paxilli.* **Phytochemistry 41:** 327-332.

Omkumar, R. V. and Rodwell, V. W. (1994). Phosphorylation of Ser 871 impairs the function of His 865 of Syrian hamster 3-hydroxy-3-methylglutaryl-CoA reductase. **Journal of Biological Chemistry 269:** 16862-16866.

Roitelman, J., Olender, E. H., Bar-Nun, S., Dunn Jr, W. A. and Simoni, R. D. (1992). Immunological evidence for eight spans in the membrane domain of 3-hydroxy-3-methylglutaryl coenzyme A reductase: implications for enzyme degradation in the endoplasmic reticulum. **The Journal of Cell Biology 117:** 959-973.

Rowan, D. D. and Gaynor, D. L. (1986). Isolation of feeding deterrents against Argentine Stem Weevil from ryegrass infected with the endophyte *Acremonium loliae.* **Journal of Chemical Ecology 12:** 647-658.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). **Molecular Cloning: A Laboratory Manual**. New York, Cold Spring Harbour Laboratory Press.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA Sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences of the USA 74:** 5463-5467.

Schardl, C. L. and An, Z. (1993). Molecular biology and genetics of protective fungal endophytes of grasses. In: **Genetic Engineering.** Eds. J. K. Setlow. New York, Plenum Press. 191-212.

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.

Schlötter, C. and Pemberton, J. (1994). The use of microsatellites for genetic analysis of natural populations. In: **Molecular Ecology and Evolution: Approaches and Applications.** Eds. B. Schierwater, B. Streit, G. P. Wagner and R. DeSalle. Switzerland, Birkhäuser Verlag Basel. 203-214.

Scott, B. and Schardl, C. (1993). Fungal symbionts of grasses: evolutionary insights and agricultural potential. **Trends in Microbiology 1:** 196-200.

Senstag, C., Stirling, C., Schekman, R. and Rine, J. (1990). Genetic and biochemical evaluation of eucaryotic membrane protein topology: multiple transmembrane domains of *S. cerevisiae* 3-hydroxy-3-methylglutaryl coenzyme A reductase. **Molecular and Cellular Biology 10:** 672-680.

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. **Journal of Chemical Ecology 16:** 3301-3315.

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

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. **Journal of Molecular Biology 98:** 503-517.

Stermer, B. A., Bianchini, G. M. and Korth, K. L. (1994). Regulation of HMG-CoA reductase activity in plants. **Journal of Lipid Research 35:** 1133-1140.

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

van Heeswijck, R. and McDonald, G. (1992). *Acremonium* endophytes in perennial ryegrass and other pasture grasses in Australia and New Zealand. **Australian Journal of Agricultural Research 43:** 1683-1709.

von Heijne, G. (1992). Membrane protein structure prediction, hydrophobicity analysis and the positive-inside rule. **Journal of Molecular Biology 225:** 487-494.

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

Wilkensen, H. H., Bush, L. P., Siegel, M. R. and Schardl, C. (1997). **Lolines are fungal metabolites**. International *Neotyphodium*/Grass Interactions Symposium, Athens, Georgia,

Woitek, S., Unkles, S. E., Kinghorn, J. R. and Tudzynski, B. (1997). 3-Hydroxy-3-methylglutaryl-CoA reductase gene of *Gibberella fujikuroi*: isolation and characterization. **Current Genetics 31:** 38-47.

Wolfe, K. H. and Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. **Nature 387:** 708-713.

Yoder, O. C. (1988). *Cochliobolus heterostrophus*, cause of southern corn leaf blight. **Advances in Plant Pathology 6:** 93-112.