Comparative and Functional Analysis of Gene Expression in *Ophiostoma* Species

A thesis

submitted in partial fulfillment

of the requirements for the Degree of

Doctor of Philosophy in Biological Sciences

at

The University of Waikato

by

Lisa Marie Robson

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2008

Abstract

Ophiostoma floccosum and *Ophiostoma piliferum* are polymorphic ascomycete fungi found throughout the world. Both species are important economically as they are known to colonise timber and cause discoloration of wood thus reducing its aesthetic value and subsequently price. Albino variants of the two species, in particular *O. piliferum*, are used as biological control agents to prevent sapstaining and have been used commercially for the past 15 years to reduce pitch/wood extractives in paper manufacturing. Other members of the genus include the plant pathogens *O. novo-ulmi* and *O. clavigerum*, known to have a severe effect on forest health and economy around the world. *O. floccosum* and *O. piliferum* have been demonstrated in the laboratory to be fermented in large volumes and they are particularly suitable as hosts capable of secreting extracellular recombinant proteins.

This research aimed to investigate the transcriptome and molecular functioning of *Ophiostoma floccosum* and compare this to transcriptomic data available for *Ophiostoma piliferum* and other *Ophiostoma* species, *O. novo-ulmi, O. clavigerum* and *O. piceae*. This research contributes to the development of *O. floccosum* and *O. piliferum* as hosts for protein expression and advances the knowledge of gene expression and molecular functioning in this genus.

To gain insight into the molecular functioning of *O. floccosum*, an expressed sequence tag (EST) collection from yeast-like growth (blastospores) was created during early phase growth. A total of 1207 EST sequences with an average length of 713 bp were identified. Clustering and assembly of the high-quality EST data set resulted in the identification of 598 unique putative transcripts (UPTs). Functional classification of these UPTs, using both homology searching and *ab-initio* methods, indicated that the majority of protein transcripts produced were involved in metabolism and cell proliferation. Up-regulation of mitochondrial transcripts involved in respiration and the presence of transcripts homologous to enzymes involved in the tri-carboxylic acid cycle indicated that aerobic respiration was likely the preferred method of ATP production in *O. floccosum* blastospores. However, the putative identification of genes

encoding alcohol dehydrogenases within *O. floccosum* ESTs and the presence of homologues in other *Ophiostoma* species would suggest that these *Ophiostoma* species are also likely to be capable of metabolic functioning under anaerobic conditions.

To identify homologous genes between *Ophiostoma* species, the *O. floccosum* EST data set was compared to 20,783 ESTs from other *Ophiostoma* species including *O. piliferum, O. novo-ulmi, O. clavigerum* and *O. piceae*. All UPTs identified within each of the datasets were aligned resulting in the identification of 347 clusters containing EST sequences from more than one *Ophiostoma* species. Six were identified that had homologues in all of the datasets excluding *O. piceae*. Three of the six homologous UPTs were predicted to function in core metabolism with two of the UPTs identified as encoding enzymes used in the glycolysis pathway and one encoding a 60S ribosomal protein. The other three homologous UPTs were thought to have a functional role in protein fate and were putatively identified as being a superoxide dismutase, heat-shock protein and a structural alpha-B chain tubulin gene. Of the 347 clusters, 86 of these contained transcripts identified in the *O. floccosum* EST datasets, and of these 86, only 10 fragments did not align with any significant homology to other fungal sequences contained in the NCBI non redundant database, indicating that the majority these transcripts are conserved in other fungal species.

Predicted genes within the *Ophiostoma* EST datasets were also investigated to determine codon usage and to identify the presence of genes predicted to encode proteases. Both are important factors in recombinant protein expression. Protease production can severely inhibit the production of recombinant protein in fungal hosts. Based on sequence homology to known proteases, putative proteases were identified in all of the *Ophiostoma* species investigated with the exception of *O. piceae*. Homologues for all six peptidase groups were identified including a possible glutamic acid protease and proportionally high numbers of serine and metallo-protease homologues. This research constitutes the first reported findings of putative peptidases in the aspartic, cysteine, glutamic and threonine peptidase families in *Ophiostoma* species.

Key to the over-expression of recombinant proteins is the optimisation of codons in a cloned gene to better utilise available tRNA species within the recombinant host. No codon bias was apparent between up-regulated and lower frequency transcripts in *O. floccosum, O. piliferum, O. clavigerum* and *O. novo-ulmi*. Codon usage was found to be consistent between these *Ophiostoma* species. However, a large difference between the codon usage in mitochondrially encoded genes compared to nuclear encoded genes in *O. floccosum* was indicated.

To optimise the efficiency of a recombinant expression system, we sought to identify promoters in both *O. floccosum* and *O. piliferum* that may be applied to a vector system. Using EST data, the most up-regulated UPTs identified from *O. floccosum* and *O. piliferum* ESTs were a putative subunit 4 of the NADH-ubiquinone oxidoreductase protein (NADH-UR4) and a possible heat-shock protein (HSP), respectively. A unique hydrolase gene was also identified by molecular probing of *O. floccosum* genomic DNA. This putative 96 kd protein, called PLIP-Lg, was predicted to be a mitochondrial A1 phospholipase based on both nucleotide and predicted amino acid sequence structure and homology. These gene sequences were investigated using genome walking methods to further elucidate nucleotide sequences in the 5' and 3' directions. *In silico* investigation of the 5' promoter region of the genes identified a number of predicted transcription factor binding sites, including possible TATA boxes identified previously in the promoter region of an *O. floccosum* protein. Additionally, RT-PCR methods were used to compare the expression of these transcripts throughout growth in both the mycelial and blastospore forms. All three predicted genes were found to be transcribed throughout growth in both morphological forms and, thus, the use of their promoters in a vector system would not be limited to one morphology. However, the level of expression in blastospores compared to mycelial growth varied by up to 20 fold. Therefore, the morphological form of the fungi did influence the level of expression of these genes and is a factor for consideration for future promoter use.

This PhD thesis research provides the first comprehensive investigation into gene expression and the transcriptome of *O. floccosum* while also providing the first comparative look into similarities between the transcriptomes of several *Ophiostoma*

species. Subsequently, this research adds to the knowledge of metabolic functioning in *Ophiostoma* species and illustrates the usefulness of EST analysis in determining core molecular functioning within this group. Further to addressing these goals, the research will augment future research into various biotechnological applications for the genus, specifically the development of *O. floccosum* and *O. piliferum* as hosts for recombinant protein expression.

Acknowledgements

Firstly, I would like to thank my chief supervisor, Professor Roberta Farrell for giving me the opportunity to undertake this PhD thesis and for all of her guidance, support and friendship throughout this journey. Working with her over the past six years has been an absolute pleasure.

I would also like to thank my second supervisor, Professor Louie Bernier from the Université Laval, Québec. Your comments and suggestions have been invaluable throughout this process. I hope someday we can again discuss science and travel over a good glass of wine.

I am grateful to the University of Waikato for the receipt of a Doctoral Scholarship and the Foundation for Research, Science and Technology along with my supervisor Prof. Roberta Farrell for the research study award and funding to attend "The Ophiostomatoid Fungi: Expanding Frontiers meeting in Brisbane 2006". Also I am grateful to Cambridge Healthtech Institute for the invitation to present and expenses paid to attend 'Protein Expression Europe', in Prague, September 2007. Thank you also to the University of Waikato Biological Sciences Department for the travel award granted to further cover expenses during that trip.

Thank you to all those who have given me help and guidance throughout my research in particular Colin Monk for all of your technical help, Shona Duncan for the years of incessant questioning and Joanna McKenzie for your hours of proofing, all with a smile. To all of my lab mates over the years; Jo, Joel, Ian, Vic, Arvina, John, Emma, Marisa, Matt, Judith and the many more who have passed through the doors, thank you for the laughs and smiles.

Thank you to all of my friends, old and new, especially Kate, Stacey, Corina and Katrina, you have kept me sane throughout this quest.

To my new family, Elyse our beautiful baby girl, although you did not know it you kept me company during many of the long hours it took to put this work together. Brendon, your patience, unfailing love, and support made everything possible, nothing insurmountable and it all worthwhile. Without your encouragement and absolute faith in me I may never have reached the end of this journey. Thank you.

Finally, I wish to thank my family, Mum, Dad, Ainsley and Sarah. It has been a very long road. Never once have you doubted me. You have supported me financially, physically and emotionally through everything I have ever done and without any of you none of this would ever have been imaginable let alone possible.

You cannot stay on the mountain forever so why bother in the first place.
Just this.
What is above, knows what is below
But what is below does not know what is above
One climbs
One sees
One descends
There is an art of conducting ones self in the lowlands
One may no longer see
But one can at least know!

Annon.

Table of Contents

List of Figures

List of Tables

Glossary of Terms / List of Abbreviations

3':	Three prime – the asymmetric end of a DNA strand possessing a terminal hydroxyl group
5':	Five prime – the asymmetric end of a DNA strand possessing a terminal phosphate group
μg:	Micrograms
Appressorium:	The swollen tip of a hypha or germ tube that facilitates attachment and penetration of the host by the fungus
AMP:	Adenosine Monophosphate
ATP:	Adenosine Triphosphate
BLAST:	Basic Local Alignment Search Tool
bp:	Base pairs
CAI:	codon adaptation index
cDNA:	copy DNA
CDS:	Coding Regions
Cluster:	A unique putative transcript derived from a single or group of EST sequences likely to be derived from the same gene or same RNA species
Codon:	The sequence of three nucleotide bases (triplet) in mRNA that specifies an amino acid and specifies the position of that amino acid in a polypeptide chain through complementary base pairing with an anticodon in tRNA
Contig:	A contiguous stretch of DNA sequence assembled from partial, overlapping sequences
C_T:	Cycle threshold. The cycle when sample fluorescence exceeds a chosen threshold above calculated background fluorescence
ddH_2O:	Double distilled water
DNA:	Deoxyribonucleic acid
dsDNA:	Double stranded Deoxyribonucleic acid

EST:	Expressed Sequence Tag
FMN:	Flavin mononucleotide
Fragment:	An EST nucleotide sequence used in a consensus alignment
Genome:	The complete set of genetic information contined in the DNA of an organism
GTP:	Guanidine Triphsosphate
kb:	Kilo base
kDA:	Kilo dalton
L:	Litre
LSU:	Large Subunit (ribosomal DNA)
MDa:	Mega dalton
ml:	Millilitre
Met-$tRNA_i^{Met}$	(eIF2)/GTP/methionyl initiator tRNA
Mycelium:	(pl. mycelia); the entire mass of hyphae that constitutes the vegetative body or thallus of a fungus
mRNA:	messenger RNA
N:	Normal
NAD:	Oxidized form of electron plus H^+ carrier nicotinamide adenine Dinucleotide
NADH:	Reduced form of electron plus H^+ carrier nicotinamide adenine Dinucleotide
NCBI:	National Centre of Biotechnology Information
ORF:	Open Reading Frame
Orphan:	Putative ORFs without any resemblance to previously determined protein coding sequences
PABP:	Poly(A)-binding protein
PCR:	Polymerase Chain Reaction

psi:	Pounds per square inch
PTM:	Posttranslational modifications
qRT-PCR:	Quantitative Real-Time Polymerase Chain Reaction
RNA:	Ribose nucleic acid
rpm:	Revolutions per minute
RT-PCR:	Real-Time Polymerase Chain Reaction
SAGE:	Serial Analysis of Gene Expression
Singlet:	A UPT containing only 1 EST fragment
SSU:	Small Subunit (ribosomal DNA)
TAE:	Tris Acetate EDTA
TESS:	Transcription Element Search Software
Transcriptome:	The portions of genes expressed as mature mRNA
T_m:	Melting temperature
tRNA:	Transfer Ribose Nucleic Acid
UPT:	Unique Putative Transcript
UTR:	Untranscribed region

1 Introduction and Literature Review

1.1 Introduction

The Ophiostomataceae family represent a family within the Ascomycota Phylum that are morphologically similar, share similar niches linked to their biological characteristics and typically show adaptations for their dispersal by insects (Zipfel *et al*., 2006). *Ophiostoma floccosum* and *Ophiostoma piliferum* are non-pathogenic saprophytic polymorphic fungi found throughout the world. Both species are of economic importance as they are known to cause discoloration of wood thus reducing its aesthetic value and subsequently price. Other members of the genus, such as *O. ulmi* and *O. novo-ulmi*, are known to be plant pathogens that have had a severe impact on forest health and economy around the world. Albino variants of *O. floccosum* and *O. piliferum* are used as biological control agents to prevent sapstaining with *O. piliferum* used commercially for the past 15 years to reduce pitch/wood extractives in paper manufacturing (Blanchette *et al*., 1992; Farrell *et al.*, 1992; Wendler *et al.*, 1992). Extracellular xylanases and lipases were characterised from these species (Schirp *et al.*, 2002, 2003a,b) and were demonstrated to have post translational modifications (Abraham and Breuil, 1996; Brush *et al*., 1999). However, in addition to these studies, little has been published about the biochemical and genetic properties of *Ophiostoma* despite the global impact of the genus. In the last five years there has been a significant effort made by different groups to further the understanding of protein production in these species using transcriptomic approaches. Large expressed sequence tag (EST) datasets have been constructed for *Ophiostoma piliferum, Ophiostoma novo-ulmi* and *Ophiostoma clavigerum* (*Grosmannia clavigera*). However, to date (2008) only one publication which described the *O. clavigerum* EST dataset has been published (DiGuistini *et al.*, 2006).

Due to the capability of culturing *Ophiostoma piliferum* and *O. floccosum* in large scale stirred tank fermenters and their biotechnological applications, they have been considered for their potential as host expression systems capable of the production and secretion of extracellular recombinant proteins. Currently, several fungal species and

strains are used for protein/enzyme production either for academic or commercial purposes with the best known being *Saccharomyces cerevisiae*, *Trichoderma reesei* and *Aspergillus nidulans*. These are typically for the most part patented and proprietary to select groups, therefore, not available for general use. Also, they may not be effective expression systems for specific genes and protein expression. New systems are constantly being sought for heterologous protein production.

This PhD research aimed to contribute to the knowledge of *Ophiostoma* species by investigating the transcriptome of *O. floccosum* using EST as well as genome walking and reverse transcription polymerase chain reaction (RT-PCR) methods and compare this with transcriptomic data from other *Ophiostoma* species with particular focus on *O. piliferum*. In addition to trying to further the molecular understanding of this genus, this thesis research will augment future research into various biotechnological applications for the genus. It was part of the University of Waikato New Zealand (NZ) Foundation for Research Science and Technology (FRST) funded project *Defining Quality, Expanding Markets and Related Technologies* (funded 2002-2007*)* with the objective to develop native New Zealand *Ophiostoma* species (sp.) as efficient recombinant systems for expression of heterologous biomolecules. From 2002-2004 the University of Waikato collaborated with Helena Nevalainen, Junior Te'O and Peter Bergquist of Macquarie University, Sydney, Australia, on the objective to develop *Ophiostoma* sp. recombinant expression systems. The collaborative research results were described in the International Patent: Nevalainen, P. L. Bergquist, V.S.J. Te'O and R.L. Farrell (2004). PCT. #1405967 entitled "Fungal host for expression and production of recombinant products". Filed 24 September 2004. Some of the results of this PhD thesis research were presented on a poster at 'The Ophiostomatoid Fungi: Expanding Frontiers' meeting in Brisbane, Australia, 2006 and as an invited oral presentation at the 'Protein Expression Europe' meeting held in Prague, Czech Republic in September 2007.

1.2 Literature Review

1.2.1 Kingdom of Fungi

The kingdom of fungi are a monophyletic group including a diverse range of organisms such as moulds, yeasts, mushrooms, truffles, rusts, mildews and puffballs. There is wide acceptance that some of the "fungi" traditionally studied by mycologists are Protozoa (e.g. myxomycetes) and Chromista (e.g. downy mildews and water moulds) (McKenzie, 2004). Indeed, it is difficult to define the exact limits of the fungal group. Traditionally, biologists have defined fungi as eukaryotic, spore-producing, achlorophyllous organisms with absorptive nutrition that reproduce both sexually and asexually (Alexopoulos *et al*., 1996). Many fungi have filamentous, branched hyphae *(*sing. hypha, from the Greek word *hyphe* = web) surrounded by cell walls containing cellulose and/or chitin (Alexopoulos *et al*., 1996). The mass of hyphae, constituting the thallus (pl. thalli, from the Greek word *thallos* = green shoot) of a fungus, is called the mycelium (pl. Mycelia from the Greek word *mykes* = mushroom). All fungi are heterotrophic organisms devoid of chlorophyll that live either saprobically on dead organic matter or symbiotically in association with living cells in other organisms (Alexopoulos *et al*., 1996; McKenzie, 2004). Such symbiotic relationships include examples of mutualistic (e.g. lichens and mycorrhizas), commensalistic (e.g. endophytes) and parasitic (e.g. species such as *Magnaporthe oryzae* which causes rice blast and *Candida albicans* a human pathogen) species (Talbot, 2003; McKenzie, 2004; Ryan and Ray, 2004). Fungi are found in every ecosystem and are disseminated by spores produced either sexually or asexually, termed meiospores (from the Greek words *meioun* = lessen and *spora* = spore) and mitospores (from the Greek word *mitos* = thread), respectively (Alexopoulos *et al*., 1996; McKenzie, 2004). Fungi are generally divided into five main divisions dependent on their generation of sexual spores (Waites *et al*., 2001). These Phyla are Ascomycota, Basidiomycota, Zygomycota, Deuteromycota (imperfect fungi, lacking a defined 'perfect sexual stage) and Chytridiomycota.

Within New Zealand, approximately 7,400 species of fungi are known. The predicted number of species present globally is thought to be in excess of 1.5 million taxa (Hawksworth 1997, McKenzie, 2004).

1.2.1.1 Ascomycota

Ascomycetes are characterised by the presence of sexually produced spores (ascospores) formed within an ascus. However, this is not always the case as some members classified within Ascomycota do not reproduce sexually or form asci. Rather, their assignment to Ascomycota is based on other factors such as morphological, physical and/or phylogenetic comparisons to other members of the phyla (Ainsworth and Bisby, 2001). Like most fungi they are also able to reproduce asexually by the formation of non-sexual spores termed conidia. Ascomycota are composed of subphyla Pezizomycotina, Saccharomycotina and Taphrinomycotina. Saccharomycotina are commonly referred to as yeasts. Within this group most species typically grow in unicellular form or short chains of cells that reproduce vegetatively by budding rather than by the production of hyphae. Pezizomycotina is the largest subphylum. This subphylum contains all of the Ascomycota that produce ascocarps (fruiting bodies) with the exception of the genus *Neolecta* and are thought to grow mostly in the filamentous form. Pezizomycotina typically degrade biomass via extracellular enzymatic activity to produce free sugars for carbon and energy for example starch from cellulosics. Taphrinomycotina are a group of Ascomycota determined by molecular analysis to be disparate and generally considered to be more primitive than the other subphyla.

1.2.1.2 Ophiostomataceae

The Ophiostomataceae family represent a family within the Ascomycota Phylum that are morphologically similar, share similar niches linked to their biological characteristics and typically show adaptations for their dispersal by insects (Zipfel *et al*., 2006). They include the genera *Ophiostoma* Syd, and P. Syd., *Ceratocystiopsis* Upadh and Kendr. and *Grosmannia* Goid., Boll. Staz. Patol. Veg. (Van Wyk and Wingfield 1992; Seifert and Okada 1993; Wingfield *et al*., 1993; Zipfel *et al*., 2006). Controversy has surrounded the phylogeny of Ophiostomaceae since their discovery in the late 1800's and the group has been subjected to many changes (Halsted, 1890; Bakshi, 1951; de Hoog, 1974; Upadhyay and Kendrick, 1975; Upadhyay, 1981; Harrington, 1981; de Hoog and Scheffer, 1984; Harrington, 1987; Hausner *et al*., 1993a, 1993b; Samuels, 1993; Upadhyay, 1993; Mouton *et al*., 1994; Spatafora and Blackwell, 1994; Hausner and Reid, 2003; Zipfel *et al*., 2006). Much of the confusion

surrounding this group has been due to the methods of classification used. Mycologists have long used a system of classification that allows anamorphs to be named separately from the holomorph of which they form part, often resulting in two different names (Thwaites, 2003). Given the diversity of anamorphs associated with *Ophiostoma*, anamorph morphology has been a preferred characteristic with which to group species in the genus (Melin and Nannfeldt 1934; Hunt; 1956; Davidson, 1958; Mathiesen-Käärik, 1960; Zipfel *et al*., 2006). However, identification of species in anamorph genera is problematic, particularly as a number of *Ophiostoma* species produce combinations of up to three of the four possible anamorph states (Okada *et al*., 1998).). Some examples of stalked spore drops found in *Ophiostoma* species are given in Figure 1.2.1.2A. Teleomorph characters applied in taxonomic studies of *Ophiostoma* include the shape and size of the ascomata and ascospores, and the presence or absence of sheaths surrounding the ascospores (Zipfel *et al*., 2006).

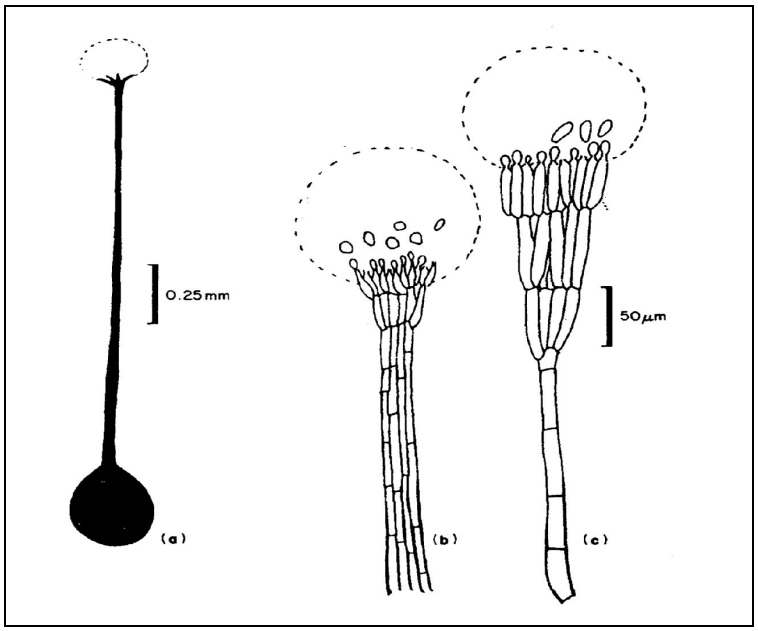


Figure 1.2.1.2A Stalked spore drops of sapstaining fungi (a) *Ceratocystis*; (b) *Graphium*; (c) *Leptographium.* Drawing taken from Hudson (1986)

With the increasing use of molecular methods, the taxonomy and classification of this Family has undergone rapid changes in recent years. Currently (2008), the Family is divided into three groups based on morphological and phylogenetic comparisons as is illustrated in Figure 1.2.1.2B. A recent history of this Family and current classification status is reviewed below.

Earlier descriptions of the Family contained the genus *Ceratocystis* Ellis and Halst. sensu stricto (s.s.). However, molecular characterisation of the genus based on small subunit (SSU), large subunit (LSU) ribosomal DNA sequences and β-tubulin genes, as well as anamorph differences, cell wall chemistry, cycloheximide sensitivity, centrum development and ascosporogenesis, showed that species of *Ceratocystis* s.s were not closely related to those of *Ophiostoma, Grosmannia* and *Ceratocystiopsis* (Harrington, 1981; de Hoog and Scheffer, 1984; Hausner *et al*., 1993a, 1993b; Samuels, 1993; Mouton *et al*., 1994; Spatafora and Blackwell, 1994; Hausner and Reid, 2003; Zipfel *et al*., 2006). *Grosmannia* and *Ceratocystiopsis* were previously thought to be synonyms of *Ophiostoma* and the groups were merged under *Ophiostoma* sensu lato (s.l.). Fungi classified within the *Ophiostoma* s.l. group all contained rhamnose and cellulose in their cell walls, displayed tolerance to cycloheximide, and had anamorphs residing in *Sporothrix, Hyalorhinocladiella, Leptographium* and *Pesotum*. Recently, Zipfel *et al*., (2006) analysed partial LSU and β-tubulin DNA sequence data from fifty *Ophiostoma* s.l. species representing all the major morphological groups. They suggested that the genus *Ophiostoma* consists of at least three groups representing separate genera. The genus *Ceratocystiopsis* has thus since been re-instated to accommodate taxa that have short ascomatal necks, produce falcate ascospores with sheaths and have *Hyalorhinocladiella* and occasionally *Sporothrix*-like anamorphs. The genus *Grosmannia* has been reinstated to accommodate a monophyletic group characterised by intron presence and positioning in the *β-tubulin* gene and a *Leptographium* anamorph.. The proposed classification of the genera *Ophiostoma, Ceratocystiopsis* and *Grosmannia*, including their respective anamorphs, is illustrated in Figure 1.2.1.2B

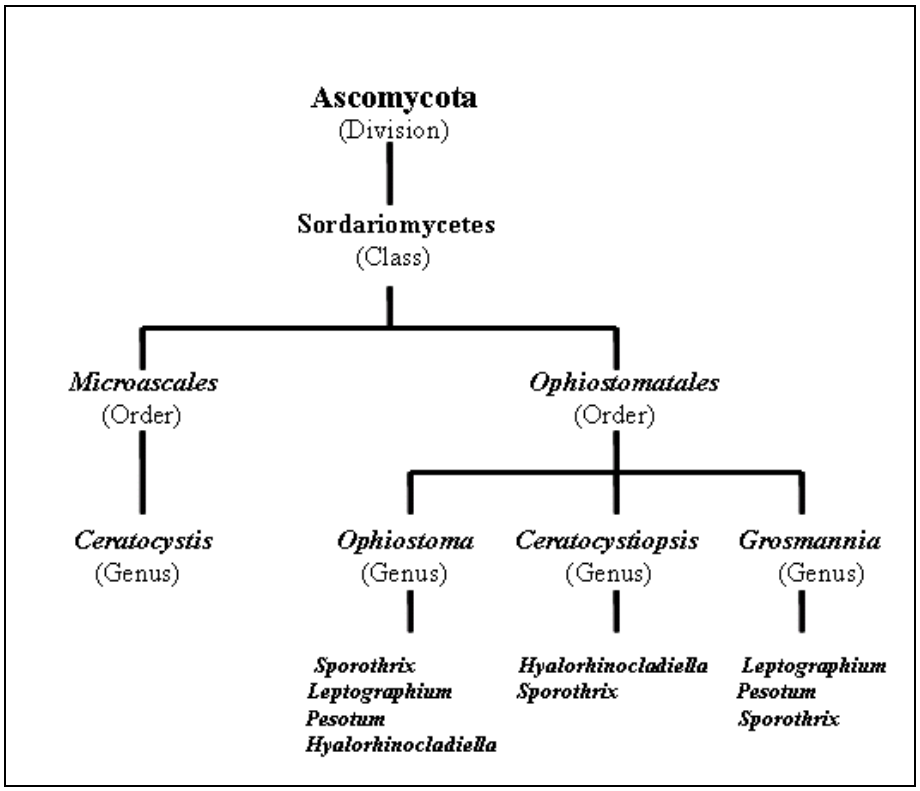


Figure 1.2.1.2B The classification of the genera *Ophiostoma, Ceratocystiopsis* and *Grosmannia* including their respective anamorphs *Sporothrix, Leptographium, Pesotum* and *Hyalorhinocladiella*. This figure was adapted from Thwaites (2003)

1.2.1.2.1 Ophiostomataceae Growth and Nutrient Acquisition

Many Ophiostomatoid fungi are polymorphic and can grow in both a filamentous and yeast-like state. The yeast-like cells are termed blastospores. Blastospores are asexual spores often synonymously called conidia. They are formed by enlargement of a recognizable conidium initial before the initial is delimited by a septum (Hawksworth *et al.*, 1996). It is suggested that the yeast phase plays a role in the dissemination of *Ophiostoma* species. In the tree pathogen *O. novo-ulmi* it is thought that it likely allows dissemination of the pathogen from tree to tree and that it also has a role in the translocation of the infection within the host tree (Pereira *et al.*, 2000). Some *Ophiostoma* species produce conidia on asexual fruiting structures called synnemata (Figure 1.2.1.2.C) or in a sporothrix-stage where conidia are formed on hyphae. A synnema consists of a group of conidiophores cemented together by their stalks.

A. B.

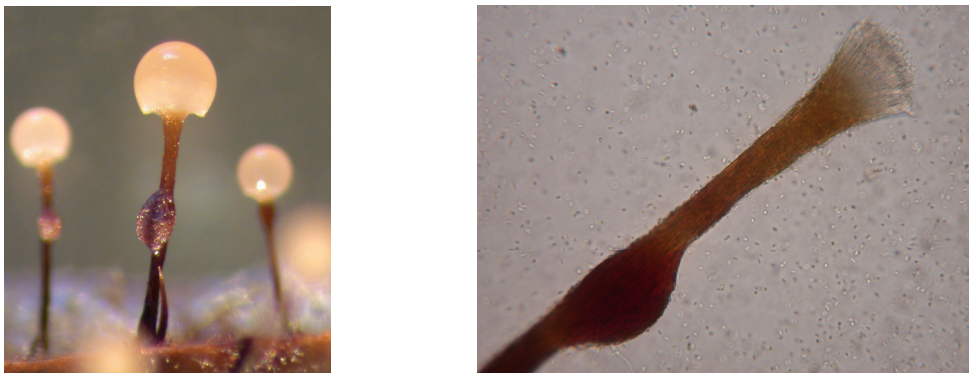


Figure 1.2.1.2.C Synnemata of *O. floccosum* taken from Thwaites (2003) A = 50X magnification; B=400X magnification

The hyphal system of these fungi appears to be uniquely adapted to penetrate, externally digest, absorb and metabolise wood. Hyphae move from one cell to another by growing through the pit membranes or directly through the cell wall using specialised constricted hyphal structures called transpressoria (Thwaites, 2003). These fungi do not affect the structural integrity of wood, rather they impart a discolouration due to melanisation in the hyphae which is observed as stain in the sapwood (Zimmerman *et al.*, 1995; Schirp *et al.*, 2003a). The hyphae may be superficial, or it may penetrate deeply into the sapwood, causing a stain that is blue, brown, red or a combination of colours (Thwaites, 2003). The nutrient, moisture and oxygen content of wood, temperature, relative humidity and wood type was shown to influence the growth of Ophiostomataceae and the level of melanisation which causes the sapstain (Thwaites, 2003).

The majority of nutritive substances in wood, notably conifers, are present in the cytoplasm of parenchyma cells, lumen of tracheids and vessels and in the resin canals of the sapwood. Two types of substances are found; hydrophilic compounds (proteins, amino acids, starch and soluble sugars) and hydrophobic substances including wood extractives and resins. Starch and lipids are the principle nutrition source in the storage tissue of wood (Zabel and Morrell, 1992). Soluble sugars also constitute a major proportion of the total non-structural carbohydrate present (Cranswick *et al.*, 1987). To utilise the available nutrients in wood, Ophiostomataceae fungi typically produce extracellular enzymes that hydrolyse the macromolecules in wood into assimilable nitrogen and carbon (Abraham *et al*., 1998). Many of these fungi express amylases in

order to utilise starch as a carbon source. Mathiesen-Käärik (1960) described the carbon utilisation of a number of sapstaining fungi and found that there was considerable variation in the utilisation of carbohydrates by individual fungi. The types of secretory enzymes utilised by *Ophiostoma* species will be discussed in the following section.

1.2.1.2.2 Extracellular Enzymes Produced By *Ophiostoma* Species

To date, *O. ulmi* and *O. novo-ulmi* are the most well characterised *Ophiostoma* species in terms of genetic makeup and protein production. These species have been reported to produce a number of extracellular enzymes including pectinase, cellulase, β-galactosidase and xylanase (Beckman, 1956; Binz *et al.*, 1996, 1997).

Proteinases are thought to play a vital role in nutrient acquisition from wood for *Ophiostoma* species. Proteinases have been isolated and characterised from *O. piceae* (Abraham *et al.*, 1996, 1998; Gao and Breuil, 1995, 1998) and *O. piliferum* (McNaughton, 1997; Brush *et al.*, 1999; Hofman and Breuil, 2004). Proteinase from *O. piceae* was demonstrated to degrade a variety of proteins of animal and plant origin including proteins isolated from wood (Abraham *et al.*, 1998). Additionally, proteinase activity has been detected in cultures of *O. ainoae* and *O. pupulinum.* Using fluorescence based enzyme substrate assays cysteine, metallo and chymotrypsin like serine proteinase activity were detected in strains of *Ophiostoma floccosum* (Wu *et al.*, 2006). Subtilisin-like proteinase was also identified using N-terminal sequencing of proteins secreted by *O. floccosum* (Wu *et al.*, 2006, 2007).

Wu *et al.*, (2006) observed in *O. floccosum* that only a small range of strongly secreted proteins were present in liquid culture and described an abundantly produced α-amylase in *O. floccosum*. Ray parenchyma cells store food as starch and provide the most favorable path for growth of many *Ophiostoma* species through wood (Ballard *et al.*, 1982; Farrell *et al.*, 1993; Gao and Breuil, 1998; Hoffman and Breuil., 2004). Amylase activity is essential for degradation of this starch into smaller sugars for fungal consumption and therefore, not surprisingly, α-amylase was found to be abundantly secreted by *Ophiostoma* species. An early investigation by Umezurike (1969) reported that amylase investigated in some sapstain species behaved like an

inducible enzyme and was not detected in the cultures after exhaustion of starch. The production of α-amylase in *O. floccosum* at different stages of growth and/ or on different sources of carbon has not been reported.

Lipase activity has been reported for a number of *Ophiostoma* species, as discussed in the next section.

1.2.1.2.3 Lipase Production

Lipases and in particular phospholipases are discussed throughout this thesis.

Esterases and lipases both hydrolyse ester bonds. The classification of the two groups of hydrolases has been somewhat confused in the literature. An in depth review of distinguishing characteristics of both groups was given by Fojan *et al.,* (2000). Lipases are a group of enzymes that act at lipid-water interfaces (known as interfacial activation) and hydrolyse triacylglycerols to diacylglycerols, monoacylglycerols, free fatty acids and glycerol (Thomson *et al.*, 1999). Esterases preferentially break ester bonds of shorter chain fatty acids and show highest activity towards the soluble state of the substrate. In contrast, lipases display a much broader substrate range than the esterases. Lipases also tend to be active towards aggregated substrates as long-chain fatty acids are generally insoluble or have poor solubility. Due to the broad specificity of esterases and lipases, they can not be solely classified by function (Fojan *et al.*, 2000).

The structures of several lipases using X-ray crystallographic as well as NMR studies has been determined in recent years (Jaeger *et al.,* 1999; Fojan *et al*., 2000; van Pouderoyen *et al.*, 2001). Those enzymes examined were found to be members of the α/β hydrolase superfamily that contain a catalytic triad of conserved serine, aspartic/glutamic acid and histidine residues. Many lipases possess an amphipathic loop covering the active site as a lid-like structure which, associated with catalysis, exposes hydrophobic patches involved in docking to the lipid-water interface. Esterases do not display this lid-like structure (Derewenda *et al*., 1992; Fojan *et al.*, 2000). At the amino acid level, the separation of esterases and lipases is again not a simplistic one. Global nucleotide sequence alignment of esterases and lipases shows

no significant similarities. Indeed, even within lipases, despite sharing a common catalytic mechanism and structure, lipases share little overall similarity at the amino acid and nucleotide sequence level (Bradner *et al*., 2003). Some conserved amino acid motifs have been identified in esterases and lipases. The conserved sequence motif GXSXG has been identified around the central active site serine residue in both esterases and lipases (Drabløs and Peterson, 1997; Fojan *et al.*, 2000; Arrese *et al.*, 2006). Also, the ProSite database classification of the active site motif for esterases (F-[GR]-G-x(4)-[LIVM]-x-[LIV]-x-G-x-**S** -[STAG]-G where **S** is the active site residue) is also contained in the lipase motif ([LIV]-x-[LIVFY]-[LIVMST]-G-[HYWV]-**S** -x-G-[GSTAC]) and vice versa (Cygler *et al.*, 1993; Hoffman *et al*., 1999; Fojan *et al.*, 2000).

Lipase activity was previously established in *Ophiostoma* species, specifically *O. piceae* (Gao and Breuil, 1998) and *O. piliferum* (Brush *et al.*, 1999; George *et al*., 1999). Esterase activity was demonstrated in the crude extracellular supernatant of *O. floccosum* cultures by a ρ_1-nitrophenol assay adapted from Janssen *et al.,* (1994) in the MSc thesis of Sandra Wilcocks (2004). Many wood species consist of 2-6% lipids (Abraham *et al*., 1998; Gao and Breuil, 1998). Sapstaining fungi produce extracellular lipases that hydrolyse triglycerides into free fatty acids, resin acids (or whatever R group is attached through oxygen to the glycerol backbone) and glycerol during colonisation (Gao and Breuil, 1998). The triglycerides along with waxes, steryl esters, sterols, free long chain fatty acids and resin acids are part of the complex called 'extractives' that occur in wood (Fengel and Wegener, 1989). Triglycerides are a major compound in pitch; the term given to lipid deposition during mechanical pulping of wood. Hence, *Ophiostoma* species have long been of interest in paper pulp manufacture for their ability to decrease the concentration of wood triglycerides and subsequently reduce pitch (Brush *et al.,* 1994). To date no purified lipase protein or sequence data has been published for *Ophiostoma floccosum*.

A phospholipase is a lipase enzyme that converts phospholipids into fatty acids and other lipophilic substances. Phospholipases are involved in a diverse number of processes including membrane homeostasis and remodeling, nutrient acquisition, generation of bioactive molecules as well as pathogenicity and virulence (Robson *et*

al., 2005; Köhler *et al*., 2006). There are four major classes of phospholipases termed A, B, C and D that are distinguished by the type of reaction they catalyze.

Within the phospholipase A group, A1 phospholipases (phosphatidylcholine 1-acylhydrolase) selectively act on the fatty acid in position 1 (sn-1) in phospholipids to cleave a free fatty acid and form a lysophospholipid. Phospholipase A_1 catalyses the reaction of:

phosphatidylcholine +H_2O => 2-acylglycerophosphocholine + carboxylate (fatty acid).

A_1 phospholipases attack and cleave the fatty acid from the number 1 position (sn-1) of the glycerol backbone of lecithin (so leaving the acyl group remaining on the number 2 position), while phospholipase A_2 attacks the number 2 position (sn-2) (see Figure 1.2.1.2D below) (Köhler *et al.*, 2006).

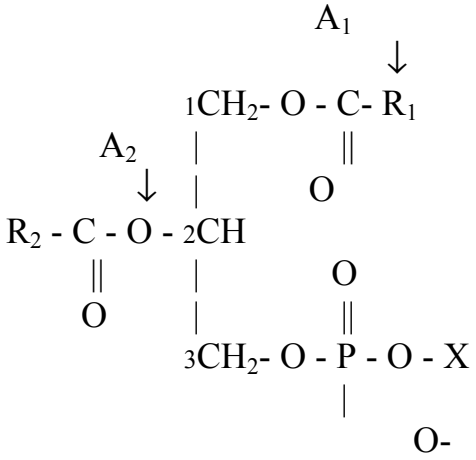


Figure 1.2.1.2D Diagram indicating the two different reaction sites of a phospholipid molecule by the two different phospholipase enzymes, phospholipase A_1 and A_2. Figure taken from Köhler *et al.*, 2006.

Fungal phospholipases cloned and expressed in recombinant fungal systems are used in the food industry, for example, phospholipase A1 from *Fusarium venenatum* was cloned and expressed in *Aspergillus* species for use around the world in cheese production (Pronk *et al*., 2008).

1.2.1.2.4 *Ophiostoma* Species Investigated

Specifically, this PhD thesis research studied various aspects of *O. floccosum, O. piliferum, O. piceae, O. novo-ulmi* and *O. clavigerum (G. clavigera)* with particular focus on *O. floccosum*. The phylogenetic relationship between these species with the exception of *O. clavigerum* was investigated by Schroeder *et al.*, (2001) (Figure 1.2.1.2.4A) and again by Zipfel *et al.*, (2006) (Figure 1.2.1.2.4B). Based on the 18S rRNA gene sequence alignments reported by Schroeder *et al*., (2001) *O. piliferum* was sub-grouped with *O. floccosum*. All *Ophiostoma* species investigated by Schroeder *et al*., (2001) were clearly separated from *N. crassa*. However, 18S rRNA nucleotide sequences from *O. novo-ulmi* and *O. clavigerum* were not included in the analysis.

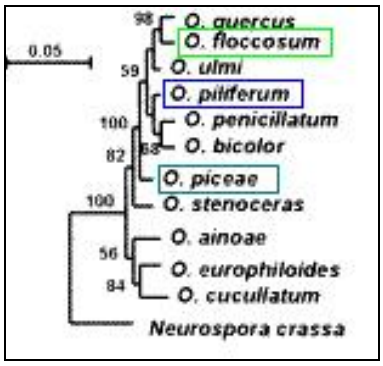


Figure 1.2.1.2.4A Phylogenetic tree generated by neighbor-joining analyses using 18S rRNA gene nucleotide sequences adapted from Schroeder *et al.*, (2001). The scale bar indicates one base change per 100 nucleotide positions. *N. crassa* was used as an outgroup for the alignment.

Zipfel *et al.*, (2006) used LSU and β-tubulin gene sequences from fifty *Ophiostomataceae* species to produce a cladogram by majority rule consensus tree, showing that all *Ophiostoma* species investigated in this research, with the exception of *O. clavigerum*, were grouped within the same sub-clade (highlighted by the red box in Figure 1.2.1.2.4B). This sub-clade contained members of the *O. piceae* complex as well as *O. piliferum* that contained *Pesotum* anamorphs. The sub-clade also included the species *O. alinoae,* and *O. araucariae* with *Pesotum*-like anamorphs as well as *O. destortum, O. flexuosum* and *O. piliferum* with *Sporothrix* anamorphs.

Harrington *et al.,* (2001) described *O. floccosum, O. novo-ulmi* and *O. piceae* as being part of the *Ophiostoma piceae* complex. This monophyletic group comprised of nine described species delimited by synnema morphology. They are insect-dispersed pyrenomycetes with synnemata (*Pesotum)* and micronematous *(Sporothrix)* synanamorphs, while other species outside of the complex that form synnemata lack

the *Sporothrix* state. With the results reported by Zipfel *et al.*, (2006), presented in short in Figure 1.2.1.2.4B, the legitimacy of this grouping by Harrington *et al.*, (2001) was questioned because species without *Pesotum* anamorphs were grouped in between species of the complex. It was the opinion of the authors of the Zipfel paper (2006) that the resolution of the taxonomic lineages produced using this data was poor and that further investigation using other genes, including intron sequence data, was required to resolve that taxonomy of this group.

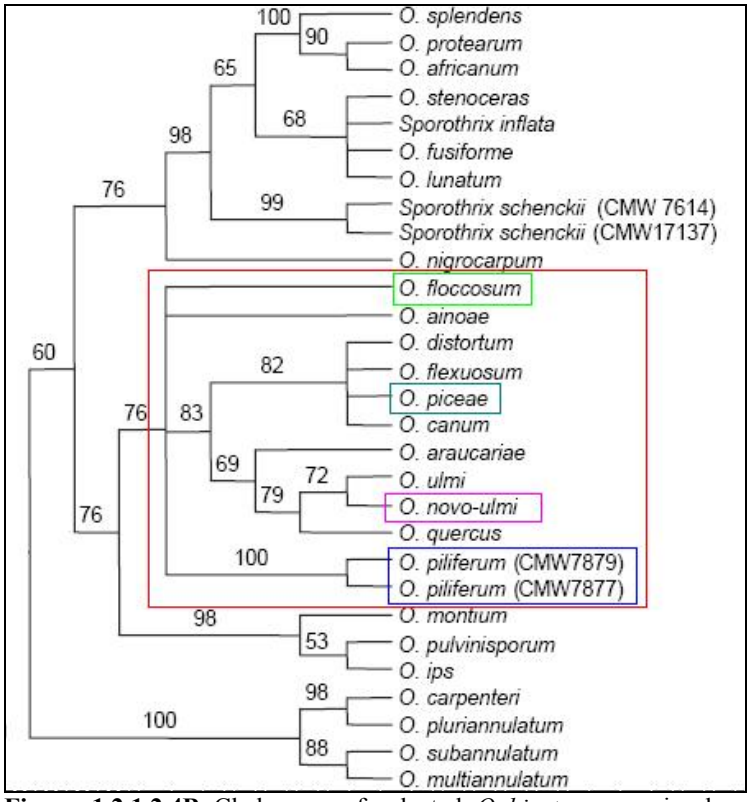


Figure 1.2.1.2.4B Cladogram of selected *Ophiostoma* species based on LSU and β-tubulin gene sequences adapted from Zipfel *et al.*, (2006). Bootstrap support values (1000 replicates) above 50% are indicated at the branches.

With the exception of *O. clavigerum*, which has not been included in any of the phylogenetic analyses, it is evident from the literature that based on the genes investigated, *O. floccosum, O. novo-ulmi, O. piceae* and *O. piliferum* are closely related in terms of nucleotide sequence (Schroeder *et al.,* 2001; Zipfel *et al.,* 2006). These species will be discussed in more detail in the following text.

Ophiostoma floccosum Mathiesen-Käärik, Svenka Botanisk Tidskrift 45:219 (1951).

Kingdom: Fungi
Phylum: Ascomycota

Sub Phylum:	Pezizomycotina
Class:	Sordariomycetes
Order:	*Ophiostomatales*
Family:	Ophiostomataceae
Genus:	*Ophiostoma*
Species:	*floccosum*

Ophiostoma floccosum was first isolated from *Pinus* species in Sweden (Mathiesen-Käärik, 1960) and was well described by Mathiesen in 1951. It was later isolated in North America, Europe, New Zealand, Australia, Asia and Chile (Harrington *et al*., 2001). *O. floccosum* was previously described as being a synonym of *O. piceae* (Griffen, 1968; Przbyl and de Hoog, 1989). Harrington *et al*., (2001) first described it as a distinct species. Since then taxonomic analysis of the LSU and β-tubulin genes showed *O. floccosum* and *O. piceae* as being clearly distinct from one another (Zipfel *et al*., 2006). *O. floccosum* can be easily identified by its characteristic red brown synnema as shown in Figure 1.2.1.2C (Farrell *et al.*, 1998; Thwaites, 2003).

O. floccosum has been isolated throughout New Zealand. Between 1996 and 1998, 1958 samples from branches, twigs, needles, leaves and cones were collected from 869 sites in a New Zealand nationwide survey. In that survey *Ophiostoma floccosum* was the second most abundantly identified *Ophiostoma sp*. with 293 isolates identified. *O. piliferum* was also found to be predominant with 154 isolates (Thwaites *et al.*, 2005).

O*phiostoma piliferum* (Fr.) Syd. & P. Syd., Annales Mycologia 17:43 (1919).

Kingdom:	Fungi
Phylum:	Ascomycota
Sub Phylum:	Pezizomycotina
Class:	Sordariomycetes
Order:	*Ophiostomatales*
Family:	Ophiostomataceae
Genus:	*Ophiostoma*
Species:	*piliferum*

O. piliferum was first described by Fries in 1822. *O. piliferum* was listed as being one of the most common causes of sapstain in the United States (Seifert, 1993). It has been isolated in Canada, United States, New Zealand, Europe and South America. *O. piliferum* can be characterised by mycelium composed of *Sporothrix* conidiophores.

Melanin deficient strains of *O. piliferum* have been produced using controlled sexual crosses. These strains have been used as a commercial product to reduce pitch in wood prior to pulping (Blanchette *et al.*, 1991; Farrell *et al.*, 1993). These melanin deficient strains lack the ability to produce mature perithecia and are therefore restricted in their ability to mate and disseminate viable ascospores (Zimmerman *et al.*, 1995).

Ophiostoma piceae **(**Münch) Syd. & P. Syd., Annales Mycologia 17:43 (1919).

Kingdom:	Fungi
Phylum:	Ascomycota
Sub Phylum:	Pezizomycotina
Class:	Sordariomycetes
Order:	*Ophiostomatales*
Family:	Ophiostomataceae
Genus:	*Ophiostoma*
Species:	*piceae*

O. piceae was first described by Münch in 1907. It has been isolated in North America, U.K., New Zealand, Chile, Japan and Europe (Morlet, 1992; Brasier and Kirk, 1993; Hamlschlager *et al.*, 1994; Pipe *et al*., 1995; Kim *et al.*, 1999; Harrington *et al*, 2001). *O. piceae* grows mostly on coniferous trees (Uzunovic *et al.*, 1999). It was originally described as a hardwood and softwood colonizing species including the sibling species now classified as *O. quercus* (Brasier and Kirk, 1993).

Ophiostoma novo-ulmi (Brasier, Mycopathologia 115: 155 (1991).

Kingdom:	Fungi
Phylum:	Ascomycota
Sub Phylum:	Pezizomycotina
Class:	Sordariomycetes
Order:	*Ophiostomatales*
Family:	Ophiostomataceae
Genus:	*Ophiostoma*
Species:	*novo ulmi*

First described by Brasier (1991), this pathogenic fungus colonises hardwood tree species and is one of the causal agents of Dutch Elm disease (Hubbes, 1989). It has been isolated in the U.K., North America, Europe (Harrington *et al.*, 2001), and also in Australia and New Zealand (MAF – www.maf.govt.nz; AQIS – www.daff.gov.au). It has been suggested that the species may have originated from Asia but this has not been proven (Brasier and Mehrotra, 1995).

Ophiostoma clavigerum (Rob-Jeffr, & R.W. Davidson) T.C. Harr., Mycotaxon 28:41 (1987).

(Grosmannia clavigera) (Rob-Jeffr, & R.W. Davidson) Zipfel, Z.W. deBeer & M.J. Wingfield, Studies in Mycology 55:90 (2006).

Synonym: *Ophiostoma clavigerum. Europhium clavigerum, Grosmannia clavigera*

Kingdom:	Fungi
Phylum:	Ascomycota
Sub Phylum:	Pezizomycotina
Class:	Sordariomycetes
Order:	*Ophiostomatales*
Family:	Ophiostomataceae
Genus:	*Ophiostoma*
Species:	*clavigerum*

O. clavigerum is a pathogenic fungus. It is exclusively associated with the bark beetles *Dendroctonus ponderosae* and *Dendroctonus jeffreyii* (Kim *et al.*, 2005). Like a number of other *Ophiostoma* species, *O. clavigerum* colonises sapwood producing blue/black melanin pigments. It blocks transpiration within the tree resulting in death but does not affect the structural integrity of the wood (DiGuistini *et al.*, 2007).

It was suggested by Zipfel *et al.*, (2006) that *O. clavigerum* be re-classified as the *Grosmannia* genus based on its anamorph and teleomorph characteristics. Within the context of this thesis the terms *O. clavigerum* and *G. clavigerum* will be used synonymously.

1.2.2 Fungal Genomes

Due to their small genome size compared to other eukaryotic organisms, fungi have been particularly suitable for genome analysis. In May 2008, the National Center for Biotechnology Information (NCBI) genome database had records of public sequencing projects of 71 fungal genomes. Of these records, 61 were species of the Phylum Ascomycota and 10 of the Phylum Basidiomycota. Of the fungal genomes investigated, size was shown to vary greatly with between 2.5 megabasepairs (Mb) to greater than 46 Mb in some genomes (Kupfer *et al.*, 1997; Cornell *et al*., 2007). A summary of key descriptive statistics for select fungal genomes is presented in Table 1.2.2; the fungi listed classified within one of two distinct Phyla, Ascomycota and Basidiomycota. The majority of fungal genome sequencing has been conducted on the

Ascomycota. *Ophiostoma* species are grouped within the subphylum Pezizomycotina of which there are several well-known examples listed in the Table. As discussed previously in Section 1.2.1.1, the majority of fungi classified within Pezizomycota grow mostly in filamentous form. The average Pezizomycotina has a genome approximately three times larger than a Saccharomycotina (yeasts). Increased size and complexity of filamentous fungal genomes are thought to be likely due, in part, to their hyphal morphology, elongation and penetration of the complex substrata (Martinez *et al.*, 2004). However, the average number of open reading frames (ORFs) predicted between the two subphylum indicates that Pezizomycotina has only twice the number of predicted ORFs. Predicted ORFs within Pezizomycotina are on average longer and typically contain 2-3 exons compared to Saccharomycotina ORFs of which only 5% have introns. A review of the statistics between the two subphyla was given by Arvas (2007), who pointed out the difference in the percentage of guanines (G) and cytosines (C) (GC%) between the two groups. Saccharomycotina generally had a lower GC% throughout their genomes compared to Pezizomycotina genomes, but wide variation was observed between Pezizomycotina genomes.

Phylum	Subphulum	Genus/ Species	Chromosomes	Genome Length (Mbp)	GC% Genome	GC% Genes	Total tRNAs	Total ORFs	Average ORF Size (bp)
Ascomycota	Pezizomycotina	*Neurospora crassa*	7	38-40	50		424	9,200	1,673
		Magnoporthe grisea	7	39-40	52		316	11,109	1,683
		Trichoderma reesei	7	34	52	-		9,130	1,476
		Aspergillus nidulans	8	30	50		188	9396-9,500	1,868
		Aspergillus fumigatus	8	28	49		179	9,009	1,644
		Aspergillus niger	8	33.9	50	53.4	269	13,238-14,165	1,572
		Aspergillus oryzae	8	37	48			12,074	1,414
		Gibberella zeae	4	36.1				11,640	
	Taphrinomycotina	*Schizosaccharomyces pombe*	3	13.8 -14	36		174	4,824-5,041	1,426
	Saccharomycotina	*Yarrowia lipolytica*	6	21	49		510	6,703	1,428
		Debaryomyces hansenii	7	12	36		205	6,906	1,167
		Kluyveromyces lactis	6	11	39		162	5,329	1,383
		Candida glabrata	13	12	39		207	5,283	1,479
		Saccharomyces cerevisiae	16	12	38		274	5,500 - 5,807	1,455
Basidiomycota	Ustilaginomycetes	*Ustilago maydis*	23	20	54	-		6,902	1,935
		Phanerochaete chrysosporium	7-9	30	57	59	200	11,777 -	
	Hymenomycetes	*Cryptococcus neoformans*	14	19	49		141	6,572	1,900

Table 1.2.2 Descriptive statistics of selected sequence fungal genomes. GC % is the percentage of guanine and cytosine nucleotides. Statistics were gained from the following publications: Bruce Birren, 2003; Mannhaupt *et al*., 2003; Martinez *et al*., 2004; Arvas, 2007; Cuomo *et al.*, 2007; Pel *et al*., 2007.

Contrary to earlier opinions, it is now widely accepted that DNA is dynamic rather than static in that it is constantly being subjected to rearrangements, insertions and deletions (Prak and Kazaian, 2000). Gene duplication within a genome may occur as an error in homologous recombination, retrotransposition or duplication. It can be a

significant advantage in evolution by allowing the acquisition of new functions (Cornell, *et al*., 2008). Gene duplication among the Pezizomycotina, in general, appears to be slightly more frequent than among the Saccharomycotina (Cornell *et al.*, 2008). Transposons are mobile genetic elements that can move from one location in the genome and reinsert at another site (Prak and Kazaian, 2000; Arvas, 2007). While some eukaryotic genomes consist of up to 50% transposon sequences (Prak and Kazaian, 2000; Yu *et al*., 2002), fungi by comparison have very low levels or no transposons at all, as is the case for *Ashbya gossypii* (Dietrich *et al.*, 2004). Within *Aspergillus* species, 1 and 3 % of the genomes consist of transposons (Galagan *et al.*, 2005). In Pezizomycotina species, a repeat induced point (RIP) mutation has been identified for the removal of duplicated sequences. In *Neurospora crassa* these mutations are typically G:C → A (Adenine):T (Thymine) transitions (Mannhaupt *et al.*, 2003). Within the *Neurospora crassa* genome, despite 10% of the genome containing repetitive DNA, no intact mobile elements can be found. This is likely due to the inactivation of these sequences during or prior to meiosis by RIP. By preventing gene duplication, RIP effectively stabilizes genomes. Differing levels of RIP activity have been established in Pezizomycotina species. Interestingly, some examples of fungi identified as undergoing large scale gene duplication are *Rhizopus oryzae*, *Phanerochaete chrysosporium* and *Magnaporthe grisea*. Given the pathogenic nature of these fungi perhaps faster genetic evolution is a more favorable trait.

To date, relatively little is known about *Ophiostoma* genetics and protein production. Various enzymes have been isolated and studied from the genus (Abraham *et al*., 1996; Beckman, 1956; Binz *et al.*, 1996, 1997; Breuil *et al*., 1994; Celero-Rueda *et al*., 2002; Gao and Breuil, 1998; Mc Naughton, 1997; Schirp *et al*., 2003). However, there are no representative *Ophiostoma* species that have had their genomes fully sequenced.

1.2.3 Fungal Protein Production

A protein is defined as a polymer consisting of amino acid units linked between the carboxyl group of one amino acid and the amino group of a second amino acid. Protein synthesis is the process whereby messenger ribonucleic acid (mRNA) is transcribed from the genome and synthesised into a protein in a process called translation. Transcription is described in more detail in Section 1.2.4.

The steady-state levels of proteins in eukaryotic cells are strongly dependent on translational regulatory mechanisms (Rhoades *et al.*, 2006). The translational efficiencies of mRNAs and the overall rate of protein synthesis is regulated in response to endogenous and exogenous environmental conditions such as nutrition, cellular stress, developmental signals and hormonal levels (Sonenberg *et al.,* 2000). The translation of eukaryotic mRNAs is a highly competitive and tightly regulated step in gene expression. Control is thought to be most commonly exerted at the rate limiting initiation phase (Preiss and Hentze, 1999). A simplified diagram of the complexes formed and factors participating in the initiation of protein synthesis in eukaryotes taken from Rhoads *et al.*, (2006), is presented in Figure 1.2.3.

The complex and elaborate process of protein synthesis has been studied in some depth in a number of different organisms (Naranda *et al.*, 1996; Preiss and Hentze, 1999; He *et al.*, 2003; Rhoads *et al.*, 2006). Firstly, before protein synthesis begins, the amino acids must be activated by reacting with adenosine triphosphate (ATP) to form amino acid adenosine monophosphate (AMP) complexes and pyrophosphates. This process is catalyzed by aminoacyl-transfer ribonucleic acid (tRNA) synthetase in the presence of magnesium. Aminoacyl-tRNA synthetase also catalyzes the charging of tRNA. In this step, the amino acid AMP-enzyme complex joins with the amino acid binding site of its specific tRNA releasing AMP and enzyme.

The assembly of amino acids into a polypeptide chain involves three events; initiation, elongation and termination of the polypeptide chain. Initiation of protein synthesis in eukaryotic cells involves activation of the ribosome. Firstly, the small (40S) ribosomal subunit binds to the eukaryotic initiation factor 2 (elongation Initiation Factor (eIF)2)/Guanidine triphosphate (GTP)/methionyl initiator tRNA (Methionine (Met)-$tRNA_i^{Met}$) ternary complex to form the 43S preinitiation complex. mRNA is then bound with eIF4F bound to its m^7G cap, and poly Adenine (A)-binding protein (PABP) bound to its poly(A) tail to produce the 48S pre-initiation complex. eIF4F is a hetero-multimeric complex composed of the cap-binding protein eIF4E and the RNA-dependent ATPase eIF4A. eIF4B is a RNA-binding and RNA annealing protein that acts with eIF4H to stimulate eIF4A helicase activity and the modular factor eIF4G that

specifically binds to and co-localizes all of the other proteins involved in mRNA recruitment. The multi-subunit factor eIF3 stimulates recruitment of Met-$tRNA_i^{Met}$ and mRNA to the 40S ribosome. Factors eIF1 and eIF1A monitor the base pairing between the anticodon of Met-$tRNA_i^{Met}$ and triplet sequences in the mRNA as the 48S complex migrates from the 5' end of the mRNA. This lateral movement of the 48S complex along the mRNA is a process called scanning. Correct base pairing between the Met-$tRNA_i^{Met}$ anticodon and the first Adenine Uracil Guanine (AUG) start codon triggers the hydrolysis of GTP bound to eIF2. The preassembled eIFs then dissociate and an initiation complex containing the AUG-anticodon base pair in the ribosomal P-site is formed. eIF5B then facilitates the joining of the 40S initiation complex with the 60S subunit to form the 80S initiation complex, the direct precursor for the elongation of polypeptide chains on the ribosome (Naranda *et al.*, 1996; Preiss and Hentze, 1999; He *et al.*, 2003; Rhoads *et al.*, 2006).

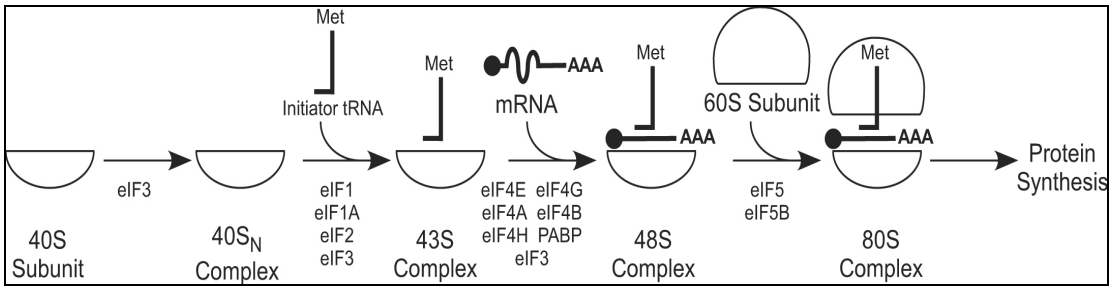


Figure 1.2.3 Complexes formed and factors participating in the initiation of protein synthesis in eukaryotes taken from Rhoads *et al.*, (2006).

The elongation phase of translation involves the decoding of mRNA and synthesis or the corresponding polypeptide chain. In fungi elongation requires three elongation factors, eEF-1, eEF-2 and eEF-3. eEF-1 is a multimer that mediates the binding of the cognate aminoacyl-tRNA to the A site of the ribosome while eEF-2 is a monomer that catalyses the movement of the ribosome relative to the mRNA (Proud, 1994, Anand *et al*., 2003). eEF-3 is an ATPase that is required for the binding of the aminoacyl-tRNA-eEF1A-GTP ternary complex to the ribosomal A-site (Andersen *et al.*, 2005). Both eEF-1 and eEF-2 are highly conserved in all eukaryotes whereas eEF-3 is only required for fungal elongation (Andersen *et al.*, 2005).

1.2.3.1 Protein Modification – Post Translational Modifications

Many proteins require 'processing' or modification to the linear amino acid polymeric structure to attain an active conformation. For example, the human genome codes for around 25,000 genes which work in concert to produce in the order of 1,000,000 distinct proteins. This diversity is because a given gene can serve as the basis for many possible versions of a protein through varying types of post translational modifications (PTM); there are estimated to be more than 300 PTM currently known (Bhadauria *et al.*, 2007). This discussion will highlight some of the most commonly studied PTMs.

The majority of proteins secreted by fungi are glycosylated (Peberdy, 1999). Glycosylation involves the covalent attachment of carbohydrate residues to specific amino acids and can either be linked through nitrogen (N-linked) or oxygen (O-linked); N-linked has the sugar covalently attached to the nitrogen on the asparagine acid and O-linked through the oxygen on the lysine, serine, threonine or proline acids. Fungal glycans are highly branched and commonly rich in mannose (Archer and Peberdy, 1997). In research where fungal glycoproteins were deglycosylated by treatment with endo- and exo- glycosidases, it was concluded that one function of glycosylation in fungi was to aid protein folding and stabilisation, in particular thermal stability, of the mature glycoprotein (Lang and Looman, 1995; Wang *et al.*, 1996).

Proteolytic cleavage is a common method of post-translational modification in fungal proteins particularly for secreted proteins. Research into the signal sequences that are removed by proteolytic cleavage is quite advanced and, as a result, many bioinformatic programs exist that predict cellular localization based on predicted cleavage sites and homology to known signal sequences. Amino acid residues may be removed and side chains chemically altered. Proteolytic cleavage may be required for activation of an enzyme that is synthesised in an inactive form.

Reversible phosphorylation may regulate protein activity. Phosphoryl groups are attached via the hydroxyl groups of specific serine, threonine or tyrosine residues by protein kinases. Other modifications include disulphide bond formation, hydroxylation, acetylation, sulfation and attachment of lipids or membrane anchors.

1.2.3.2 Protein Secretion

The secretory pathway functions in protein folding, glycosylation, processing and targeting, either to extracellular or to intracellular sites (Archer and Peberdy, 1997). The process begins with transcription in the nucleus. These transcripts are translated and the protein translocated to the endoplasmic reticulum. The folding of secreted proteins is supported by chaperones, glycosylation enzymes and oxidoreductases (Arvas, 2007). Proteins that are correctly folded are then transported to the Golgi body for further modification to modulate function and stability. Translocation of a protein across a biological membrane is achieved with the aid of the proteinaceous translocation machinery composed of pore-forming and force-generating systems (Yoda and Noda, 2001). Transmembrane and secreted proteins are synthesised with a signal sequence at the Amino-terminus which directs entry of the protein into the lumen of the endoplasmic reticulum (Archer and Peberdy, 1997). In most prokaryotic and eukaryotic secretory proteins, the Amino-terminus sequence spans about 15 to 30 amino acid residues and contains three domains; a hydrophobic region, a sequence of basic amino acids at the amino-terminal, and a peptidase cleavage site distinguished by a pair of small, uncharged amino acids separated by a single intervening amino acid. The secretion signal is removed by cleaving the signal peptidase from the folded and processed protein during the translocation process (Yoda and Noda, 2001).

In Pezizomycotina, secretion often occurs at the hyphal tip where the fungal cell wall is much thinner (Gordon *et al.*, 2000). In Saccharomycotina, proteins are secreted via fusion of Golgi complex-derived secretory vesicles within the plasma membrane.

Genome analysis of various Saccharomycotina and Pezizomycotina species has revealed very few differences in the ORF content related to secretion machinery (Arvas, 2007). Perhaps this similarity is an indication that some secretion mechanisms have been missed by current homology based annotation or, as would be more advantageous from a biotechnological perspective, the mechanism in many yeast systems could be shaped into producing the levels of secretion achieved in Pezizomycotina with only minor modifications.

1.2.3.3 Recombinant Fungal Protein Production Systems

Recombinant filamentous fungal strains have been developed for the production of various kinds of homologous and heterologous proteins (Gouka *et al*., 2000). However, while the production of homologous proteins can be high in terms of milligrams of recombinant protein produced (Punt *et al.,* 2002), the production of heterologous proteins is often limited (Iwashita, 2002). Many factors may limit heterologous protein production including the nature of the protein, the GC% of the gene or its regulatory DNA elements, codon preference for translation, the requirement of post-translational modifications, the quantity of protein needed and the cost of production. One of the most significant factors thought to lower heterologous protein yields is the abundant production of secreted proteases (Punt *et al.*, 2002).

Research investigating protein production in various *Aspergillus* and *Trichoderma* species has indicated that production of extracellular proteins is much greater in the mycelial form of growth (Carlsen *et al.,* 1996). Both of these genera, although capable of producing spores, do not grow in a single cell, yeast-like state. There are many examples in literature that have demonstrated the capability of yeast to secrete mature homologous and heterologous proteins (Eckart and Bussineau, 1996). Two key advantages of yeast-like production of extracellular proteins include significantly reduced production costs due to the ease of growth in a fermentation process and relative ease of purification reducing further downstream processing.

1.2.3.4 Protein Coding Gene Content of Fungal Genomes

The differences between biological niche, nutrient acquisition and growth forms are reflected in the protein coding gene content of Pezizomycotina and Saccharomycotina genomes. As has been previously mentioned, *Ophiostoma* species are classified within the Pezizomycotina. However, as many of this species are dimorphic in that they can grow in both a yeast-like and filamentous state, their protein coding gene content and protein production may share many similarities with both Pezizomycotina and Saccharomycotina species.

The Pezizomycotina species contain more ORFs related to the degradation of biomass such as glycoside hydrolases and proteinases, enzymes typically important in nutrient acquisition in these species (Arvas, 2007). Also a greater number of ORFs related to the production of secondary metabolites can be identified in Pezizomycotina genomes compared to Saccharomycotina Payne *et al*., 2006; Arvas, 2007). Cornell *et al*., (2008) investigated the genetic basis of yeast and filamentous growth based on the identification of conserved motifs within the genomes of 34 fungal species. They identified many expanded groups of motifs within Pezizomycotina that indicated greater metabolic flexibility and genes associated to responses in nutrient availability and environmental stresses as well as greater numbers and more diversification of ORFs associated with carbon utilisation. For example, some of the protein families cited as being expanded in Pezizomycotina include major facilitator superfamily transporters involved in the transport of small solutes, such as sugar uptake and drug efflux, sugar transporters, ABC transporters, alcohol dehydrogenases and cytochrome P450 proteins, often identified as pathogenicity factors in fungi (Cornell *et al.,* 2008). By contrast, greater numbers of ORFs related to cell wall synthesis, specifically glucan synthases, mannosytransterases and PIR proteins thought to be important for cell wall anchoring (Castillo *et al.*, 2003) were identified in Saccharomycotina compared to Pezizomycotina genomes. Cornell *et al.*, hypothesised this difference was likely due to the difference in the chemical compositions of the cell walls between the two subphyla. The walls of many filamentous ascomycetes such as *Neurospora* and *Aspergillus* species contain greater quantities of chitin compared to yeasts, where chitin is only found in bud scars (Bowman and Free, 2006; Cornell *et al.*, 2008). Cluster analysis of *S. cerevisiae* ORFs involved with metabolism by Arvas (2007), indicated that over 90% were conserved in the 33 fungal genomes investigated. In contrast, Arvas did identify large differences in the predicted mitochondrial genes between Saccharomycotina and Pezizomycotina genomes. He hypothesised that the differences were likely due to the difference in glucose metabolism between the subphyla and the rapid growth and glucose consumption of yeast species.

1.2.4 Fungal Gene Expression

In this PhD thesis research, gene expression was defined as the transcription of information encoded in a gene into ribonucleic acid (RNA). The term “gene

expression" is often used synonymously with "transcription". However, protein expression levels are dependent on transcription as well as control mechanisms such as nuclear export and mRNA localization, transcript stability, translational regulation and protein degradation (Pradet-Balade *et al*., 2001). Chromatin organisation and its regulation also have a major impact in determining the observed levels of mRNA and derived protein. Nucleosome positioning can either restrict or provide access of transcription factors to their recognition sequences in promoters and is therefore a powerful element of control in gene expression (Archer and Peberdy, 1997).

Transcriptional control has a major part in determining the yields of protein produced by fungi (Archer and Peberdy, 1997). Gene regulation in eukaryotes begins with the binding of sequence-specific activator and repressor proteins to enhancers, operators and silencers. All eukaryotic mRNAs contain a contiguous sequence of nucleotides coding for protein synthesis (Min *et al*., 2005). A mature eukaryotic mRNA molecule, starting from the 5' end usually consists of a 5' untranslated region (5'-UTR), a protein-coding region (ORF) consisting of spliced exon DNA, and a 3'-UTR followed by a poly(A) tail. An ORF extends from a start codon AUG (ATG in cDNA) and continues until the reading frame is terminated by one of three translation stop codons, UAG, UAA or UAG. The 5'-UTR region may consist of various elements that affect translational control such as a 7-methyl-guanosine cap; hairpin-like secondary structures; interacting proteins; upstream ORFs and an internal ribosome entry site. The 3'-UTR region may consist of various elements such as antisense RNA and cytoplasmic polyadenylation elements that among other factors may affect subcellular localization, protein stability and may impart a role on translational control (Mignone *et al.,* 2002). A diagram of the basic structure of a eukaryotic gene, from DNA sequence to mRNA transcript is given in Figure 1.2.4.

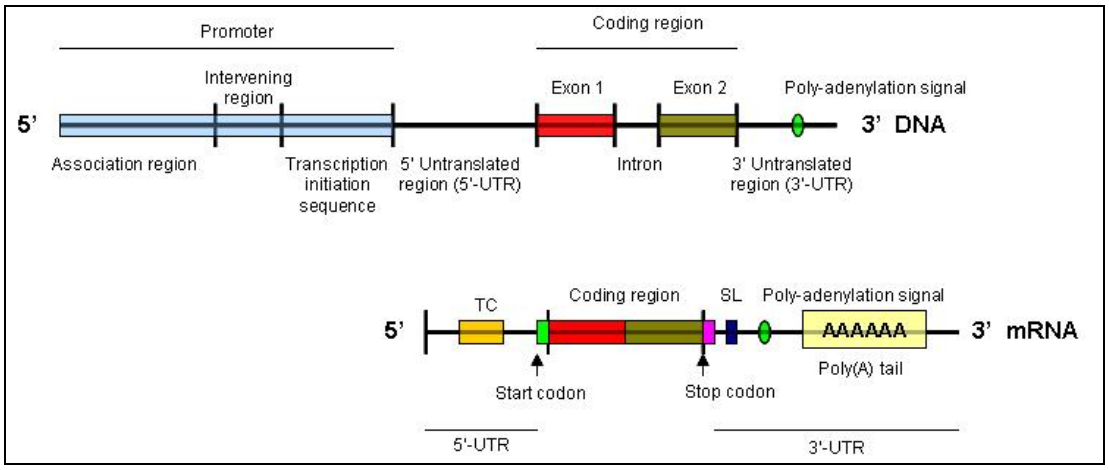


Figure 1.2.4 The basic structure of a eukaryotic gene adapted from Lynch (2006). TC = Aspects of translational control such as 7-methyl-guanosine cap; hairpin-like secondary structures; interacting proteins; upstream ORFs; internal ribosome entry site. SL = Aspects of subcellular localization such as antisense RNA and cytoplasmic polyadenylation elements.

Transcription can be bi-directional. However, frequently regions show more transcripts from one strand compared to the other (Kapranov *et al.*, 2007; Wilhelm *et al.*, 2008).

1.2.4.1 Promoter identification

Gene promoters are stretches of DNA that drive and regulate gene expression. It is thought that the promoter regions of eukaryotic protein can be divided into two arbitrary sections, as follows:

(1) A core promoter region adjacent to the transcription start site approximately 50 nucleotides in length. This region contains the initiator element and is hypothesised to direct the correct assembly of the ribonucleic acid (RNA) polymerase complex (I, II or III) at the right position to direct a basal level of transcription.

(2) A more distant enhancer region thought to contain those elements that regulate the spatio-temporal expression (Nikolov *et al*., 1996; Novina and Roy, 1996; Berk, 1999; Fessele *et al,*. 2002). Cis-acting regulatory elements are short conserved motifs typically 5-20 nucleotides in length usually found in the vicinity of a promoter. Transcription factors bind to the DNA of specific cis-acting regulatory elements and orchestrate the initiation or transcription.

Some promoters are constitutive in that they allow for continual transcription of associated gene(s) whereas many others are active or inactive depending on the developmental stage and the physical environment (growth medium) of the organism. Some of the strongest microbial promoters known are induced by complex carbon

sources that require hydrolysis to smaller units for the organism to be able to use them for food and energy production (reviewed in Nevalainen *et al*., 2004). There are various methods used to isolate effective gene promoters. Strategies described for promoter isolation generally use a nucleic acid based approach. Genes and their promoters can be isolated from DNA libraries using hybridisation to specific probes (*e.g.* Vanhanen *et al*., 1989), using sequence based genome walking methods, or by applying differential hybridisation to separate between genes that are functional or shut off under particular cultivation conditions (*e.g.* Teeri *et al*., 1983). Proteomic methods have also been used to elucidate gene promoters with the use of proteomic displays (Nevalainen *et al*., 2004,). This approach starts from strongly expressed proteins and works 'backwards' to the corresponding genes and their regulatory sequences including gene promoters. The founding assumption is that proteins that are strongly expressed at any particular time or environment are synthesised from genes possessing strong promoters.

Wu *et al.*, (2006) elucidated the sequence both upstream and downstream of an alpha amylase gene in *Ophiostoma floccosum.* Within the 2.6 kilobase (kb) nucleotide sequence upstream of the α-amylase gene, Wu identified various transcriptional elements including promoter regions. The 5' promoter region contained a TATA-like box (TTAATAAA), five putative CAAT boxes and one CCAAT box believed to be the binding sites of transcription factors involved in the regulation of the promoter activity. The promoter regions of the *Ophiostoma ulmi* hydrophobic cerato-ulmin protein (CU) and *O. novo-ulmi COL1* gene have also been investigated. (Bowden *et al*., 1994; Pereira, *et al*., 2000) Putative TATA boxes and CAAT boxes were also transcriptional factors identified within these promoter regions.

1.2.4.2 Comparisons Between Yeast and Mycelial Gene Expression

For all *Ophiostoma* species, including *O. piliferum* and *O. floccosum*, different somatic states may be observed. In liquid culture medium, yeast-like forms (so-called blastospores) can be obtained under shaking conditions, while the mycelial form is obtained when grown in solid or liquid medium under static conditions (Dogra and Breuil, 2004). A number of other fungal species are capable of this morphological dimorphism including many important pathogens such as *Candida albicans*, *Ustilago*

maydis, and *Cryptocossus neoformans*, in which the capability for dimorphic growth is thought to be a virulence factor (Sánchez-Martinez, and Pérez-Martin, 2001). Within *Ophiostoma* species, there has been little work done to identify differences in the genetic state and in particular, differences in gene expression between the two somatic forms. A paper published by Dogra and Breuil (2004) examined the difference between the two forms using molecular methods. They utilised suppressive subtractive hybridisation PCR to isolate transcripts specifically up-regulated in either the blastospore or mycelial forms of *O. piceae,* in order to identify and compare differentially expressed genes in the two forms. This study isolated 260 yeast specific and 151 mycelial specific clones. However, only 50 of these (35 yeast and 15 mycelial) clones were randomly selected for sequence analysis. Of these 50, only 33 were submitted to National Center for Biotechnology Information (NCBI) GenBank. The majority of the 33 sequences did not align with any significance to any known proteins.

In *O. novo-ulmi* the COL1 gene, predicted to encode a RNA binding protein, was identified as being an important factor in mycelial growth (Pereira *et al.*, 2000). RNA binding proteins have functional roles involved in pre-mRNA splicing, RNA processing, translation, polyadenylation and the control of transcriptional regulation (Becht *et al.*, 2005). They are also important post–transcriptional control factors related to somatic development in *C. elegans* (Loria *et al.*, 2003). When this gene was disrupted in *O. novo-ulmi* using insertional mutagenesis, filamentous growth was severely disrupted but yeast-like growth in the mutant was comparative to the wild type strain (Pereira *et al.*, 2000). Using heterologous hybridisation, the group also identified genes with sequence similarity to COL1 in *S. schenkii* and *H. capsulatum*, both of which are dimorphic fungi.

In *S. cerevisiae*, certain diploid laboratory strains have the ability to switch from a unicellular yeast form to a pseudohyphal form when grown on media selective for nitrogen starvation. Within these strains the cyclic adenosine monophosphate (cAMP) pathway, MAP kinase pathway and RIM genes have all been demonstrated to be involved in the transition from yeast to filamentous forms (Madhani and Fink, 1998; Pereira *et al.*, 2000; Sánchez-Martinez, and Pérez-Martin, 2001). Homologues of *S.*

cerevisiae genes involved in the MAP kinase pathway were also identified as regulating factors in filamentous growth in *C. albicans* (Lo *et al.*, 1997). In a publication by Niimi (1996), the addition of cAMP was reported to induce hyphal growth suggesting that *C. albicans* and *S. cerevisiae* share similar genetic mechanisms for controlling dimorphic growth. In contrast to both *C. albicans* and *S. cerevisiae*, low levels of cAMP have been reported to stimulate the transition from yeast-like cells to filamentous growth in *Ustilago maydis* (Gold *et al.*, 1994; Sánchez-Martinez, and Pérez-Martin, 2001). From the few dimorphic fungi investigated, it is evident that although different exogenous and endogenous factors contribute to filamentous and yeast development, some of the genetic strategies for controlling this cell differentiation are broadly conserved among fungi.

1.2.4.3 mRNA Expression in Relation to Protein Production

Abundantly transcribed genes for the most part are indicative of a strong promoter. A criticism of using mRNA expression profiling in characterizing cellular phenotypes is the assumption that the transcriptome faithfully represents the proteome (i.e. the complete set of proteins encoded by the genome) (Pradet-Balade *et al*., 2001). Indeed, protein expression levels are dependent not only on transcription but also on post transcriptional control mechanisms such as nuclear export and mRNA localization, transcript stability, translational regulation and protein degradation (Pradet-Balade *et al*., 2001). There is conflicting evidence supporting a direct correlation between mRNA abundance and protein expression, (Futcher *et al*., 1999; Ghaemmaghami *et al*., 2003; Greenbaum *et al*., 2003) and disputing a direct and predictive correlation (Gygi *et al*., 1999). *Saccharomyces cerevisiae* was used as the test organism for the studies sited, as comprehensive data including the entire genome sequence (Goffeau *et al.*, 1996) and the number of mRNA molecules for each expressed gene were available. mRNA levels were determined using a number of methods such as expressed sequence tag (EST) data, serial analysis of gene expression (SAGE), quantitative PCR (qPCR), Affymetrix chip hybridisation and others (Velculescu *et al.*, 1997; Wodicka *et al*., 1997). Total protein was identified using two dimensional (2D) proteomic methods and differing methods of spot quantification (Futcher *et al*., 1999; Gygi *et al*., 1999).

Direct predictive correlations were determined between mRNA levels and protein abundance using a variety of statistical analysis methods (Futcher *et al*., 1999; Greenbaum *et al*., 2003). As one would expect, the calculated correlations were not in exact correlation in these studies, with at least a 10-fold range in protein abundance. In the Greenbaum *et al.*, (2003) investigation, yeast genes identified as being involved in cellular organization, protein synthesis and energy production were typically enriched from the transcriptome to the proteome. The level of correlation was also found to differ with the level of expressed gene i.e. up-regulated proteins appeared to correlate with greater significance than proteins of less abundance (Futcher *et al*., 1999; Gygi *et al.*, 1999). This was likely due to post transcriptional regulation and errors in the methods used to determine mRNA level and total protein, these errors being more significant in terms of detection with lower levels of transcript and protein (Futcher *et al*., 1999).

In the Futcher *et al.,* study (1999), it was calculated that in abundant yeast proteins, for every mRNA molecule, approximately 4,000 molecules of cognate protein were produced. By taking into account an average mRNA half life of 30 minutes and a cell doubling time of 120 minutes, they predicted that an individual molecule of mRNA may be translated roughly 1,000 times. This implied that translation within *S. cerevisiae* is initiated on an mRNA every 2 seconds and by deduction indicated translation of over 20 amino acids per second. This figure was surprisingly high given that the rates estimated from mammals were 3 to 8 amino acids per second (Mathews *et al.,* 1996). Undoubtedly, these calculations are subjective as the stability of mRNA and cell doubling times are known to vary considerably both within an organisms due to environmental parameters and between different organisms.

Recently, investigations of eukaryotic transcriptomes revealed that much more of the genome is transcribed than expected indicating that genomes are not collinear, rather there is extensive overlap resulting in interleaved and modular transcription (Kapranov *et al.*, 2007). In *S. pombe*, between 94 and 99 % of the mitochondrial and nuclear genomes were reported to be transcribed (Wilhelm *et al.*, 2008). Also, many of the functional properties of transcripts were found to be not necessarily based on coding sequences but on regulatory sequences in untranslated regions of non-coding RNAs.

Thus, some genomic sequences may be multifunctional in that they may act as both regulatory regions and independently regulated transcripts. This regulatory function of some regions may directly affect subsequent mRNA stability. For example, in the Wilhelm *et al*., study (2008), they reported that the length of 5' UTR directly affected the stability of mRNA transcript in *S. pombe* with more stable transcripts having shorter 5' UTRs. In *S. cerevisiae*, genes encoding membrane and kinase proteins had long 5'UTRs. In contrast, genes encoding proteins involved in ribosome biogenesis had short 5'UTRs in both *S. pombe* and *S. cerevisiae* (David *et al.*, 2006; Wilhelm *et al.*, 2008). These newly identified levels of complexity in transcriptional control and mRNA stability are still little understood and the subject of much current research. However, consideration of their implications is vital when conducting transcriptomic analysis as they may offer further explanation as to why the transcriptome may not faithfully represent the proteome.

In general, despite recent evidence to suggest a greater level of complexity in the transcriptome than originally thought, based on direct comparisons between mRNA abundance and cognate total protein, the transcriptome and proteome appeared to be similarly represented (Futcher *et al*., 1999; Ghaemmaghami *et al*., 2003Greenbaum *et al*., 2003). It is poignant to consider, however, that most opinion has been formed on the in-depth analysis of only a few organisms. There is still insufficient transcriptional and proteomic data for the vast majority of organisms to deduce any definite conclusions.

1.2.5 Gene Expression Profiling - Transcriptomics

A general goal of molecular biology is to relate genotype to phenotype in order to further understand and/or manipulate the functioning of an organism. However, the relationships between genotype and phenotype are complex and our understanding of these associations is poor. By investigating the intermediates between them, such as mRNA transcripts, protein and lipid production, as well as the production of metabolites, we can further decipher these fundamental interactive relationships. The ease of manipulation and stability of RNA has meant that transcriptomic methods (the measurement of mRNA transcription) of expression profiling have been used extensively.

A genome is context independent because for the most part every cell within an organism contains the same DNA sequence. As the mRNA complement and level within a cell varies with physiology, pathology and/or development, the transcriptome is considered context dependent (Bustin and Nolan, 2004). Therefore, an expression profile is effectively a snapshot of cellular functioning at a particular time.

One transcriptomic method includes express sequence tag analysis (ESTs). This method is particularly suited to the investigation of organisms whose genome has not been sequenced as it requires no prior knowledge of genetic makeup which is often needed for the design of probes used in some hybridisation methods.

1.2.5.1 EST Molecular Approach

An EST is a single pass sequence taken from a randomly selected complementary (c)DNA clone. Through random selection of cDNA libraries, EST analysis is useful for large-scale identification of fungal active genes and can be an effective method of providing quantitative data on gene transcription frequency (Ebbole *et al.*, 2004, Suárez *et al.,* 2007). Given recent improvements in sequencing methods and the subsequent dramatic reduction of cost, the method is currently widely employed for gene discovery and genome characterisation (Min *et al*., 2005).

Although a commonly used and highly successful method, there are a number of pitfalls to EST sequence analysis that should be considered. EST data is comprised of a single read only sequence (which means the DNA sequence is a non-consensus sequence, read only once in a single direction). Therefore, the average quality of sequences is low and frame-shift errors due to insertions, deletions as well as artifactual stop codons are inherent with single pass sequencing. In non-normalised libraries, genes that are expressed at low levels may not be represented. Normalisation (usually by self hybridisation) is used to reduce the representation of highly expressed genes, thereby enhancing the probability of selecting clones derived from rarer mRNAs (Jongeneel, 2000). These normalised libraries may increase the diversity of mRNA transcript in a cDNA library, however, due to the reduction of abundant

transcript they cannot be assessed for frequency. Genes with long 3' untranslated regions may contain little or no coding region information. Also, chimeras resulting from artifactual ligation of unrelated cDNAs may be present (Jongeneel, 2000). EST sequences may be present in alternate forms, either due to differences in transcription initiation or termination sites or as alternatively spliced forms (Fukumura, 2003; Ebbole *et al.*, 2004). When constructing a cDNA library, partially spliced RNA species which are typically undistinguishable from splice variants are likely to be present (Jongeneel, 2000). Within a constructed cDNA library used for EST sequence analysis there may be possible contamination from genomic DNA, cDNA from other species and/or vector DNA.

As mentioned above, aside from being an effective method of transcriptome profiling, EST analysis can be used to investigate rare transcript or transcript that is different between two samples. A common method of pre-selection or 'normalisation' of a cDNA library is by cDNA subtraction. In this method, the cDNA from a tester sample is hybridised to excess cDNA from up-regulated transcript in a driver sample. The non-hybridised tester molecules are then separated from hybridised molecules and other driver sequences. This method of hybridisation has long been used with great success, however, it can produce high levels of false positives. A more efficient method is suppressive subtractive hybridisation (Diatchenko *et al.*, 1996). In this method adaptors with terminal inverted repeats are added to tester cDNA fragments. These inverted repeats result in secondary structures that are highly stable and subsequently prevent PCR amplification using oligo DT primers (Qin *et al.*, 2003; Dogra and Breuil, 2004).

Over recent years there has been a significant drive to further understand gene expression and molecular functioning in *Ophiostoma* species using EST methods. At the start of this research in August of 2004 there were only 34 entries encoding *Ophiostoma sp*. ESTs deposited in NCBI Expressed Sequence Tag Database (dbEST). As of May, 2008 there were 17,499 entries for *Ophiostoma* species. At this time no EST sequences for *Ophiostoma floccosum* species were deposited in the NCBI dbEST. Within the dataset, 5975 EST sequences were produced from four *O. clavigerum* cDNA libraries (DiGuistini *et al.*, 2006). The cDNA libraries were constructed from mycelial cultures grown on 4 different types of media designed to simulate laboratory

growth and growth on lodgepole pine in order to assess and characterise the molecular interaction between this pathogen and its host. Within the EST dataset, they identified 2620 unique putative transcripts (UPTs) of which 67 % were significantly homologous to known or predicted proteins in the public databases. They assessed up-regulated transcripts within the subsets. However, as normalisation methods were used to produce the cDNA libraries transcript frequency would not be accurately portrayed. Although little in the way of biological functioning was implied from the molecular data, some interesting results were reported by DiGuistini *et al.* (2006). Many of the up-regulated transcripts were associated with metabolism and growth. In the ESTs produced from cultures grown on nutrient limited lodgepole pine sawdust, the predicted gene ontology of UPTs had many similarities with nutrient limited yeast cells (Gasch and Werner-Washburne, 2002). In cultures grown on media containing oleoresin, a chemical produced as part of the lodgepole pine defense mechanism, an increase in the number of oxidoreductases was reported. Oxidoreductases in many plant pathogens function in detoxification and host chemical perception and are therefore important pathogenicity factors (Idnurn and Howlett, 2001). The DiGuistini *et al.* paper (2006) represented the first report of large-scale EST analysis in an *Ophiostoma* species.

Although the EST method is considered fairly reliable in terms of mRNA quantification, the expression profiles of select genes for further investigation are usually verified by Northern hybridisations and or RT-PCR.

1.2.5.2 Real Time Polymerase Chain Reaction

Real time polymerase chain reaction (RT-PCR) is one of the most sensitive and commonly used methods in use for characterising or confirming gene expression patterns and comparing mRNA levels in different sample populations (Orlando *et al*., 1999; Bustin, 2002; Liu and Saint, 2002). As a research tool it can have considerable benefit, particularly in the detection of rare mRNA targets. However, there are a number of pitfalls that require consideration when utilising RT-PCR methodology.

RNA isolation is the primary step in a RT-PCR assay. The reproducibility and biological relevance of a RT-PCR assay is highly dependent on the quality of RNA

used (Bustin and Nolan, 2004). Indeed, variability may simply be introduced by operator and reagent variability. Care must be taken to ensure that samples are free of DNA that will skew results. Also all reagents used in the RNA isolation must be free of nucleases that will degrade RNA samples during storage. It is important to try to avoid the co-purification of inhibitors. A number of PCR inhibitors have been reported in the literature. These include culture media, various components found in reagents used in nucleic extraction, components found in laboratory consumables such as powdered gloves etc (Rossen *et al*, 1992; Bustin and Nolan, 2004). Variability in the presence of inhibitors may cause variation in reverse transcription efficiency and assay amplification efficiency. High levels of co-purified RNA may also result in the failure of a PCR assay (Pikaart and Velleponteau, 1993). The efficiency of reverse transcription in the construction of cDNA inherent with different types of reverse transcriptase and different priming methods (oligo-dT, random or target-specific) will introduce variability into the template (Zhang and Byrne, 1999; Lekanne Deprez *et al*., 2002). There is evidence that cDNA yield from sequences near the 5' end of partially degraded mRNAs is significantly less than that from sequences near the polyA tail (Swift *et al*., 2000; Bustin, 2002). Due to the context-dependent nature of transcriptomic analysis, the presence of non-target cell types and variability in environmental influences, culture physiology and development, will all influence the reproducibility and validity of RT-PCR results.

There are two main categories of chemistries available for use in RT-PCR, as follows:

1) Non-specific DNA binding dyes for example SYBR Green;
2) Specific detection with dye-labeled, sequence-specific oligonucleotide primers or probes for example TaqMan, LUX etc.

This literature review has focused solely on non-specific chemistries for RT-PCR. Non-specific chemistries offer many advantages over specific methods: the assays are generally easier to design and simpler to establish, they are significantly cheaper and they allow melt-curve analysis to check the specificity of the amplification reaction. There is at least one report in the literature that suggests the SYBR Green I detection is more precise and produces a more linear decay plot than TaqMan detection (Schmittgen *et al*., 2000; Bustin and Nolan, 2004).

Non-specific methods utilise intercalating dyes that bind to any double-stranded DNA generated during the PCR reaction and emit fluorescence. SYBR Green I is the most commonly used DNA-binding dye for RT-PCR (Real-Time PCR Applications Guide, BIORAD). SYBR Green I has little fluorescence when it is free in solution, when bound to double stranded (ds)DNA its fluorescence increases by up to 1,000 fold. There are a number of disadvantages to non-specific chemistries. Due to their indiscriminate binding and multiplicity of bound dye molecules to dsDNA, fluorescence may be influenced by competing side reactions such as primer dimers. (Chou *et al*., 1992; Bustin, 2002). Fluorescence is dependent on mass. Therefore, if it was assumed that all RT-PCR reactions being compared had equivalent amplification efficiencies, amplification of a longer product would generate more signal than a shorter one (Bustin and Nolan, 2004). The need to undergo melt-curve analysis may also be seen by some as adding to the complexity of data analysis and therefore a hindrance.

1.2.5.3 Bioinformatic Approach – Gene Prediction and Cluster Analysis

In order to reduce redundancy and error within EST data-sets, it is necessary to cluster EST sequences into groups that have likely been derived from the same gene (Jongeneel, 2000). This allows the transcript frequency to be calculated and is an effective method of determining the diversity of genes being transcribed by identifying how many UPTs are present. Basic Local Alignment Search Tool (BLAST) is a commonly used set of software programs to identify homologous transcript using alignment methods to detect similarities to known protein sequences or the detection of conserved domains. As with any method, there are limitations to the BLAST algorithm that should be considered. The algorithm requires the query and database sequence to share two small regions of significant similarity before it will calculate a longer alignment i.e. it detects only local alignments not global. Therefore, sequences whose similarity extends over a relatively long region without islands of stronger homology will not be identified by BLAST (Jongeneel, 2000). BLAST limits the combination of scoring matrices and gap penalties that can be used. The BLASTx, tBLASTn and tBLASTx programs, perform a six-frame translation of the query and or the database search one frame at a time, so sequences containing frameshift errors (common in ESTs) may be missed (Jongeneel, 2000). Other errors may occur during

cluster analysis for example when multiple transcripts are present for a gene more UPTs will be identified than is truly reflective of the EST dataset. This is the case for the *Drosophila melanogaster Adh* gene The gene has different transcripts during larval and adult phases of growth while the protein-coding part of the exon is identical (Lockett and Ashburner, 1989; Amador *et al*., 2001). Also, genes may code for multiple proteins specified by alternative splice sites and sequencing errors, which are inherent in methods such as EST sequencing, may create stop codons.

The primary tool for locating protein-coding exons in genomic DNA is by the detection of an open reading frame (ORF). This method utilises conserved start and stop codons to detect gene transcription sites. However, the detection of ORFs is problematic as there are a number of difficulties in determining if an ORF is actually used by an organism to code for a protein. Due to nature of detecting ORFs short exons, that often lack sufficient identifying features, may be difficult to distinguish from intron DNA. One must also consider when identifying possible ORFs that the non-coding strand of an exon may contain statistically significant ORFs. Some of these problems can be resolved if there is conservation of the gene and information from related species or with the acquirement of additional information such as promoters, cap sites, translation signals, splice signals and transcription terminators to make a model of the gene that can be used to reject unlikely ORFs.

1.2.5.4 Codon Usage

The characterisation of codon usage patterns within an organism is important for several reasons. Within an organism, understanding the biological basis of codon usage can add to the understanding of its molecular biology and evolution such as the mutational biases present in the genome, and the extent to which different genes are under selection for efficient translation (Lloyd and Sharp, 1992). More practically, codon optimisation remains an important factor in the production and over-expression of recombinant heterologous proteins. It is therefore crucial to establish codon preference within a host expression system.

Within the genetic code, 64 triplet combinations of nucleotides (codons) code for 22 amino acids or peptides. It has been established that in a variety of organisms, highly

expressed genes exhibit preference for some of the codons encoding the same amino acids (Grantham *et al*., 1980; Wada *et al*., 1990; Jansen *et al.,* 2003; Carbone *et al*., 2004). This is called codon preference or codon bias. These 'biased' codons are codons recognised by the most abundant isoacceptor tRNA species (Bennetzen and Hall, 1982; Ikemura, 1982; Berg and Kurland, 1997). There is some evidence to suggest that this codon bias can contribute to the rapidity and accuracy of ribosomal decoding events in some but not all organisms investigated (Curan and Yarus, 1989). Currently, it is the consensus view in the literature that codon bias is an adaptation that enhances the fidelity, kinetic, and starvation responses of the translation system (dos Reis *et al*., 2004; Buchan *et al.*, 2006). However, despite the large amount of research conducted in this area, the underlying causes of this translational selection are still little understood.

Many early investigations into codon usage were conducted on genes identified from *Saccharomyces cerevisiae*. A strong bias toward a limited number of codons was identified in genes with high mRNA levels (Bennetzen and Hall, 1982; Ikemura, 1985; Sharp and Li, 1986). More recently, investigations by Futcher *et al*., (1999) and Jansen *et al*., (2003) using whole genome analysis and 2D proteomic methods compared with mRNA transcript frequency to identify abundantly expressed proteins established that codon bias was significant in abundant and even moderately expressed genes within *S. cerevisiae*. Codon bias has also been established in many Pezizomycotina species (Lloyd and Sharp, 1991; Nelson *et al.,* 2004) such as *Neurospora crassa*, *Aspergillus niger Aspergillus awamori* and *Aspergillus nidulans*. A significant amount of research into fungal codon usage and bacterial codon usage has resulted from the need to optimise and subsequently increase heterologous protein expression. Indeed for prokaryotes, codon usage has been identified as the single most important factor in gene expression (Lithwick, and Marghalit, 2003; Gustafsson *et al.*, 2004). The existence of slightly different codes in different organisms is a significant barrier to heterologous gene expression. A common strategy to improve expression is to alter the rare codon in a target gene so that they more closely reflect the codon usage of the host without modifying the amino acid sequence of the encoded protein (Gustafsson *et al.*, 2004).

Due to the difference in codon usage between up-regulated and low level mRNA in a cell, various models have been constructed that predict the expression level of a gene based on its codon sequence (Karlin *et al*., 1998; Sharp and Li, 1987). These expression level indicators have been used to assess the likelihood of transcription and translation of an ORF into a protein thereby aiding in genome annotation. The codon based-expression indicators have also been used to enhance heterologous protein expression by predicting codon sequences that are likely to be highly expressed (Jansen *et al.,* 2003). These models were originally based on low numbers of highly expressed genes and the heavy assumption that some functional classes of genes are highly expressed in fast growing bacteria (Jansen *et al.,* 2003). Recent re-analysis of the codon adaptation index (CAI) and codon usage index by Jansen *et al.*, (2003) using genome-wide expression data in yeast revealed that the models are still relatively robust. However, although claims have been made that these models are quantitative with respect to predicted expression levels (Bennetzen and Hall, 1982; Karlin *et al*., 1998; Sharp and Li, 1987) they are unable to account for physiological changes such as such as the availability or lack of certain nutrients (Coghlan and Wolfe, 2000) immune response, and other possible exogenous influences, therefore, they are unreliable predictors of mRNA or protein concentration. One example presented by Holstege *et al*., (1998) was the gene *EN01* which indicated high codon bias but was repressed in the presence of glucose.

It has been hypothesised that codon bias correlates with protein length in a number of species including *Escherichia coli, Saccharomyces cerevisiae,* and *Caenorhabditis elegans* (Coghlan and Wolfe, 2000; Duret and Mouchiroud, 1999; Eyre-Walker, 1996; Moriyame and Powell, 1998). Coghlan and Wolfe (2000) found that when the effect of mRNA concentration was eliminated, protein length showed a weak positive partial correlation with codon bias. This supported a previous hypothesis that long proteins are energetically more expensive to produce, so translation selection for codons which minimize missense errors during translation are more effective in long genes (Eyre-Walker, 1996). Not all literature reviewed supported this hypothesis. A strong negative correlation between codon usage and protein length was noted for Eukaryotes (Duret and Mouchiroud, 1999). Whether this was an error due to the statistics used (Coghlan and Wolfe, 2000) or indeed a factor influencing translational codon bias, it is

recognised as having an effect on calculating codon bias and should be considered (Carbone *et al*., 2003).

Factors other than translational bias may also contribute to biased codon usage. These include GC content, the preference for codons with G or C at the third nucleotide position (wobble position), a leading strand richer in G+T than a lagging strand and horizontal gene transfer which may induce chromosome segments of unusual base composition (Carbone *et al*., 2003; Lafay *et al*., 1999; Moszer *et al*., 1999; Sharp and Li, 1987). In a recent study of 16 prokaryotic and eukaryotic genomes, it was established that codon pair frequencies were also non-random within all of the genomes investigated. This phenomenon is also thought to be another mechanism of translational optimisation.

1.3 Hypothesis, Aims and Objectives

This PhD thesis research combined molecular and transcriptomic methodologies to advance the knowledge of *Ophiostoma* gene expression and molecular functioning. The overall goal of this PhD research was to further the understanding of protein transcription and transcriptional regulation in *Ophiostoma* species with particular focus on the species *O. floccosum* and *O. piliferum*. These two species were chosen because of the following reasons;

- *O. floccosum* is the second most abundant *Ophiostoma* species found in New Zealand and was first described by The University of Waikato laboratory in which this PhD thesis research was conducted and as described in Harrington *et al*., (2001).
- *O. piliferum* was developed as a biocontrol agent in the early 1990's, and has been studied throughout the world by both academic institutions and industry in both fundamental and applied studies.

As part of this goal, the thesis research aimed to identify and elucidate the gene sequences of up-regulated transcripts and investigate possible transcriptional elements upstream of their transcription start sites in *O. floccosum* and *O. piliferum*. Future

research building upon the thesis research results could augment the requirements for the use of *Ophiostoma* species as recombinant expression hosts

1.3.1 Hypotheses

Up-regulated mRNA transcript can be identified within an *O. floccosum* culture using EST methods.

Levels of unique mRNA transcripts are different between blastospore (yeast-like single cell) and mycelial morphologies.

Levels of unique mRNA transcripts change during different stages of growth.

1.3.2 Specific Goals and Objectives of Thesis

Objective one: To determine the growth characteristics, specifically the rate and morphology of growth of *O. floccosum* and *O. piliferum* under the specified laboratory conditions.

Objective two: To profile transcription in *O. floccosum* in early log phase blastospores using EST methods. Sequence data will be clustered into UPTs and the frequency of ESTs within each cluster determined in order to identify abundant transcripts and putatively coded genes of interest. Up-regulated transcript in *O. piliferum* will also be identified using EST data available online. All EST sequences available for *Ophiostoma* species will be compared to identify conserved genes important within the genus for molecular functioning.

Objective three: To identify and determine the DNA sequences of putative genes of interest from both *O. piliferum* and *O. floccosum*. From the DNA sequences the protein sequence and possible transcriptional elements both upstream and downstream will be identified *in silico*. Specific focus will be to identify possible strong promoters and regulatory units for potential use in recombinant expression vector construction.

1.4 Ethical and Regulatory Requirements

Ophiostoma floccosum is found throughout New Zealand and therefore no permits were required to work with New Zealand isolates in the laboratory.

Ministry of Agriculture and Fisheries (MAF) permits for the importation of *Ophiostoma piliferum* isolates into MAF certified Microbiology lab (C2:10), University of Waikato, Hamilton (reference number 759) were granted, the most current of which was MAF permit 2008033250.

Approval to develop in containment a project of low risk genetically modified organisms involving the use of *Escherichia coli, Ophiostoma floccosum, Ophiostoma piceae, Ophiostoma piliferum, Ophiostoma pluriannulatum, Ophiostoma quercus* and *Saccharomyces cerevisiae* was granted by The University of Waikato IBSC committee. The corresponding ERMA approval codes for this permit were; GMD003538, GMD003539, GMD003540, GMD003541, GMD003542, GMD003543 and GMD003544.

2 Materials and Methods

2.1 Culture Isolation, Maintenance and Growth

2.1.1 Fungal Isolates

All fungal isolates used in this study with the exception of *O. piliferum* (Cartapip™) 97 strain, were maintained and belong to the mycological culture collection at the Department of Biological Sciences, University of Waikato, Hamilton (Table 2.1.1). Cartapip™ 97, an albino *O. piliferum* developed in 1992 using consecutive rounds of sexual reproduction (Blanchette *et al.*, 1992) was a commercial product and is available from Parrac Ltd, Hamilton, New Zealand. The isolate originated from the U.S.A. and is described in Zimmerman *et al*., 1995.

Strain Number	Species	Origin (*P. radiata* D. Don)	Site in New Zealand	Date Isolated
J20206	*O. floccosum*	Felled log	Kinleith Forest, North Island	Aug-01
J2122	*O. floccosum*	Felled log	Kinleith Forest, North Island	Nov-01
J1619	*O. piliferum*	Felled log	Kinleith Forest, North Island	Dec-00

Table 2.1.1 List of fungal isolates used in this PhD research.

2.1.2 General List of Materials

2.1.2.1 General List of Chemicals and Reagents

All chemicals used, such as salts, buffers, organic solvents (including chloroform, phenol, TRIZOL), alcohols (i.e. ethanol, isopropyl alcohol) were obtained from Sigma (St Louis, MO, U. S. A.) unless noted in the text. All reagents used in molecular reactions such as oligonucleotides, restriction enzymes, reverse transcriptase and polymerases were obtained from Invitrogen (Carlsbad, CA, U. S. A.).

2.1.2.2 General List of Media

Media and agarose used throughout the thesis research were obtained from Becton Dickinson and Company (Sparks, MD, U.S.A.).

The media used in this research are listed below:

Luria-Bertani (LB) Medium

Per litre (L):

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

Volume made up to 1 L with deionised H_2O

Approximately 0.2 millilitre (ml) of 5 normal (N) sodium hydroxide (NaOH) was added to adjust the pH to 7.0. Media was sterilised by autoclaving. If antibiotic was required the autoclaved medium was left to cool to 55°C and an appropriate volume of antibiotic stock added. In the case of ampicillin antibiotic this equaled a final concentration of 100 micrograms per milliliter (µg/ ml).

LB Agar Plates

LB medium was prepared as above but before autoclaving 15 grams (g) /L of agar was added. The medium was then autoclaved for 20 minutes.

Malt Agar

Per litre:

Malt Extract (1.5%)	15 g
Agar (2%)	20 g

Yeast Medium

Per litre:

Malt Extract (1.5%)	15 g
Yeast Extract (0.2%)	2 g

Media 6

Per litre:

Malt Extract (1.5 %)	15 g
Yeast Extract (0.2 %)	2 g
Agar (2 %)	20 g
Chloramphenicol (200 micrograms (ug) /L)	200 ug

Volume made up to 1 L with deionised H_2O. The medium was autoclaved for 20 minutes to sterilise. The following antibiotics were added once the sterile media had cooled to 55 °C.

Streptomycin sulphate (100 ug/L)	100 ug
Cycloheximide (400 ug/L)	400 ug

2.1.3 Quantitation of Fungal Cells

Blastospores were counted using a haemocytometer under a light microscope (Olympus BH-2 phase contrast microscope) at 40X magnification. Approximately 10 µl of cell suspension was transferred onto the haemocytometer and a cover slip placed on top. All fungal cells in five sub-squares of the total 25 squares were counted. The optimum dilution for counting is 10-20 cells per sub-square. The total number of spores per ml was obtained using the following equation:

(Number of cells) x 5 x 10^4 x (dilution factor)

2.1.4 Determination of Fungal Biomass

Cultures of known volume were filtered under vacuum through pre-dried and weighed filter papers (0.2 µm cellulose ester, Advantec MFS Inc). The filter papers were subsequently dried in an oven at 50 ºC until a constant weight was obtained.

2.1.5 Determination of Fungal Viability

The number of colony forming units (CFUs) was determined by plate assay. A serial dilution of the culture was plated. After 48 hours of growth at 25 ºC the numbers of colonies on a plate were counted. The viability was determined by the number of colonies (best counted per plate between 30 and 300) multiplied by the dilution factor.

2.1.6 Storage of Fungal Isolates

Short term storage of fungal isolates was conducted by aseptically transferring a slab (typically 1 cm by 3 cm) of culture grown on agar to a sterile 1.5 ml microcentrifuge tube and storing at 4 ºC for a period of up to 3 months.
Fungal isolates were stored indefinitely in 20% glycerol at -70 ºC.

2.1.7 Culture Conditions

Ophiostoma floccosum strain J2122 and *O. piliferum* strain Cartapip 97 which were in constant use throughout this thesis research were maintained on agar plates, either 1.5% malt extract or Media 6 to minimize contamination. All liquid cultures were

prepared from actively growing cultures on agar plates. Unless otherwise stated, all cultures were grown at 25 ºC.

2.1.8 Growth Analysis

An accurate determination of growth characteristics for *Ophiostoma* sp was required. Growth curves were constructed in YM media in an unadjusted starting pH of 5.8 at 25ºC using the following three conditions.

1. shaken cultures: 200ml cultures in 800 ml conical flasks were maintained in a flask at 180 revolutions per minute (rpm) in a refrigerating rotary shaker (Bioline, Edwards Instrument Company, Australia)
2. bench-top stirred tank fermenter runs: cultivation was performed using a LH Series 210, 10 liter benchtop bioreactor with 8 liters of working volume. 2 litres of head space were maintained to avoid the use of antifoaming agent. Aeration of 1000 cm^3/min was used with 150 rpm and an output of 20% maintained. Both temperature and pH were monitored throughout growth. A 10 mL sample was collected every 6 hours until a steady state was observed in terms of blastospore number, fungal biomass and viability. All data was calculated from duplicate fermenter runs.
3. static cultures: 200ml cultures in 800 ml conical flasks were maintained in a flask without any movement.

All cultures were inoculated with a concentration of 2×10^5 blastospores per 100ml of medium from seed cultures produced in sterile YM media with a non adjusted pH of approximately 5.8. The seed cultures were inoculated with a $1cm^2$ plug of fungal mycelium growing on Media 6 or M agar medium and cultivated for approximately 3 days at 180 rpm and 25 ^oC before being used as inoculum.

All photographs were taken at 40X and 100X magnification using a Zeiss Axiostarplus microscope and Nikon E4500 Coolpix digital camera.

2.2 Molecular Methodologies

2.2.1 Isolation of DNA

Genomic DNA was extracted using the PSC-B method described by Miller *et al*., 1999, employing a FastPrep® FP120 Cell Disrupter (BIO101 Thermo Savant, USA). Samples were shaken for 2 consecutive rounds of 30 seconds at 4.0 m/s. All samples were assessed to determine the concentration and degree of degradation by gel electrophoresis.

2.2.2 Isolation of RNA

It was particularly important to avoid nuclease contamination in the preparation of RNA template. To ensure minimal contamination from exogenous DNA, RNA and nucleases, the following laboratory precautions were always adhered to.

- All workstations, pipettes and equipment used were wiped down with RNase AWAY™ and DNA AWAY™ (Molecular Bio Products, Inc. San Diego, C.A.) to remove RNase, DNA and DNAase contamination.
- Only DNAase and RNAase free aerosol-barrier pipette tips were used.
- Diethyl pyrocarbonate (DEPC) treated water and reagents were used wherever appropriate.

Isolation of total RNA from blastospores

Cells grown in YM media were harvested by centrifugation at 10000 x g for 20 minutes. The majority of the supernatant was discarded with the exception of approximately 5 ml. The cells and remaining 5 ml were mixed and filtered through three layers of sterilised cheesecloth. Cheesecloth was sterilised by autoclaving at 121 ºC and 2.4 kilograms per square centimetre (kg/cm^2) for 20 minutes. The flow-through material was collected into a sterile 50 ml falcon tube and the biomass (blastospores) was pelleted by centrifugation at 4000 rpm for 10 minutes.

To 500 microlitres (µl) of concentrated blastospore material in a 2 ml round bottomed screw capped polypropylene tube, 0.5 g of 0.1 millimetre (mm) and 0.5 g of 2.5 mm zirconium-silica beads and 1 ml of TRIZOL™ (Invitrogen) were added. The cells were

disrupted by bead-beating in a FastPrep® FP120 Cell Disrupter (BIO101 Thermo Savant, USA) at 4.0 m/s for 30 seconds, left to cool for 1 minute. This step was repeated twice. To settle the mixture, the sample was centrifuged for 3 minutes at 13,000 rpm. The time taken to bead-beat and cool samples allowed for the dissociation of nucleoprotein complexes by the TRIZOL solution. The supernatant was then transferred to a sterile 1.5 ml microcentrifuge tube and to this 0.4 volumes of chloroform was added. This step was to precipitate polysaccharides, extracellular matrix (ECM) and high molecular weight DNA. The sample was then shaken vigorously for 15 seconds by hand and left for 5 minutes at room temperature.

Upon centrifuging at 13000 rpm for 5 minutes the mixture separated into a lower RED (phenol-chloroform phase), an interphase and a colourless upper aqueous phase. RNA was forced exclusively into the aqueous phase whereas the DNA and the protein partition were in the interphase and the lower phenol phase. The aqueous phase was collected and the RNA precipitated by the addition of 0.5 ml of isopropyl alcohol. After incubation at room temperature for 10 minutes the RNA material was pelleted by centrifugation at 13000 rpm for 10 minutes. The pellet was washed with 1 ml of 70% ethanol, centrifuged at 13000 rpm for 3 minutes, then the ethanol was removed and the pellet air dried to remove residual alcohol.

To ensure the removal of inhibitors of translation or cDNA synthesis from the RNA preparation, a lithium chloride (LiCl) precipitation was performed. LiCl is frequently used in RNA isolation as it does not efficiently precipitate DNA, protein or carbohydrate (Barlow *et al*., 1963). The pelleted RNA was dissolved in 500 µl of DEPC treated H_2O. 1 volume of LiCl buffer was added and the sample was left to precipitate for at least 1 hr at -20 ^{o}C. The sample was centrifuged at 13000 rpm at 4 ºC for 30 minutes. Many protocols state that the sample can be left overnight at this stage but in the experience of this thesis research, this leads to greatly increased degradation of the RNA and therefore a same day completion of the protocol was employed throughout the thesis research experiments. The pellet was washed twice with 70% ethanol to remove any salt as these act as inhibitors of reverse transcriptase. The pellet was then dried to remove any residual ethanol and the dry RNA pellet was the

dissolved in a small volume of DEPC treated ddH_2O (the volume was dependent on the size of the pellet and ranged between 20 µl – 200 µl.)
The RNA was quantified using a NanoDrop® ND-1000 Spectrophotometer (Analytical Technologies, BIOLAB).

If the RNA sample was prepared for subsequent RT-PCR, then cDNA was immediately constructed. If the sample needed to be stored it was immediately frozen using liquid nitrogen and stored at -70 ºC for no more than 6 months.

Preparation of Filamentous Fungi

The protocol used for the isolation of RNA from filamentous fungi was essentially the same as used for blastospores with the following modifications:

-The mycelia were collected by filtering the liquid culture through 3 layers of sterile cheesecloth. The filamentous material was then rinsed with DEPC treated ddH_2O and the flow through discarded.

-Approximately 0.5 g of filamentous material was placed in a 2 ml round bottomed screw capped polypropylene tube containing 0.5 g of 0.1 mm and 0.5 g of 2.5 mm zirconium-silica beads to which 1 ml of TRIZOL was then added. Filamentous material analysed in this research were disrupted by bead-beating in a FastPrep® FP120 Cell Disrupter (BIO101 Thermo Savant, USA) at 4.0 m/s for 30 seconds, left to cool for 1 minute then repeated.

DEPC Treatment of H_2O

DEPC is an acylating agent that reacts with primary amines and sulfhydryl groups. All DEPC utilised was first added to an equal volume of 100% ethanol and inverted to mix before adding to the H_2O. This enabled the DEPC, which is very insoluble in water, to disperse in the solution. 0.1% (1 ml in 1 L) of DEPC was then added to MQ H_2O, incubated at 37 ^oC for at least 12 hours and subsequently heat treated at 100 ^oC for at least 15 minutes to deactivate DEPC.

2.2.3 mRNA Isolation

All mRNA was isolated using the FastTrack®MAG Micro mRNA Isolation Kit (Invitrogen) according to the manufacturer's instructions. If the mRNA sample was prepared for use in constructing cDNA then it was used immediately. If the sample needed to be stored it was immediately frozen using liquid nitrogen and stored at -70 ºC for no more than 6 months.

2.2.4 Gel Electrophoresis

Genomic DNA, PCR products and digested vectors were analyzed by agarose gel electrophoresis through 1% (wt/vol) SeaKem® agarose gel (BioWhittaker Molecular Applications, USA) or Agarose Low EEO® (AppliChem, Germany) prepared with 1X TAE buffer (4.84 g Tris base, 1.14 ml glacial acetic acid, 2 ml 0.5 moles EDTA pH 8.0).

Genetic material (PCR product, digested vector, RNA etc) was visualised using two different methods, as follows:

1) Ethidium Bromide Method - Genetic material (5 µl) product was mixed with 1 µl of 6X loading buffer (0.25% bromophenol blue, 40% glycerol in water) and applied to the well of an agarose gel, as described above, electrophoresed for 30 minutes, or until sufficiently migrated towards the bottom of the gel, at 100 volts. Gels were stained in 1X TAE with 0.5 µg/ml ethidium bromide for 10-15 minutes, and destained in MQ-H_2O for 5-10 minutes. The gels were visualised using TFX-35M GIBCO BRL UV Transilluminator and viewed with Scion Image, Beta 4.0.2, Scion Corporation©.

2) SYBR® Safe Method - For visualising genetic material using SYBR® Safe DNA Gel Stain (Invitrogen), a 1:10,000 concentration of SYBR® Safe 10,000X to TAE buffer was prepared. The buffer containing the SYBR® Safe stain was then mixed with powdered agarose to prepare a 1% agarose gel. 5 µl of genetic material was mixed with 1 µl of 6X loading buffer (0.25% bromophenol blue, 40% glycerol in water) and applied to the well of an agarose gel, and electrophoresed for 30 minutes, or until sufficiently migrated towards the bottom of the gel, at 100 volts. The gels were

visualised using a Safe Imager™ blue-light transilluminater from Molecular Probes (Invitrogen).

2.2.5 cDNA Synthesis

cDNA was prepared using two different methods, depending on whether it was used for EST analysis or RT-PCR, as follows:

cDNA Synthesis For Use In EST Analysis

For high quality, double stranded cDNA libraries used in EST analysis, the OrientExpress™ cDNA system (Novagen) was used. cDNA was constructed from isolated mRNA according to the manufacturer's instructions with the following modifications;

- cDNA synthesis was primed using oligo (dT) primers.
- Double the amount of cDNA was produced to increase the cDNA yield. At each step the reagents used were doubled (except glycogen and ETOH washing steps).
- Only one resin column was used to fractionate the cDNA to remove excess linkers and small cDNA products (<300 base pairs (bp)).

cDNA greater than 300 bp in size were pooled and unidirectionally ligated to λSCREEN phage vector arms for in vitro packaging into the λSCREEN vector system (Novagen).

cDNA Synthesis For Use In RT-PCR

Single stranded cDNA used in RT-PCR was constructed from isolated RNA using a SuperScript™ III Platinum® SYBR® Green Two-Step qRT-PCR Kit with ROX (Invitrogen). 1 µg of RNA was used for every 20 µl of cDNA produced.

Assessment Of cDNA Purity

All fungal cDNA used for molecular analysis was first screened by PCR methods using universal bacterial specific primers to identify the presence of contaminating bacterial cDNA and DNA. Two sets of primers (listed in Table 2.2.5) were used to amplify a 181 bp region (519RC and 338F) and a 1.5 kb region (Eub A (R) and Eub B

(F)) of the 16S gene. The presence of amplicons these sizes were indicative that bacterial contamination of the sample had occurred.

Primer name	Nucleotide Sequence	Reference
519RC	5' ATT ACC GCG GCT GCT GG 3'	(Muyzer *et al.,* 1993)
338F	5' TCC TAC GGG AGG CAG CAG 3'	(Muyzer *et al*., 1993)
Eub A (R)	5' AAG GAG GTG ATC CA(ACGT) CC(AG) CA 3'	(Medlin *et al*., 1988)
Eub B (F)	5' AGA GTT TGA TC(AC) TGG CTC AG 3'	(Medlin *et al.,* 1988)

Table 2.2.5 Universal bacterial specific primers targeting the 16S gene taken from the literature (Muyzer *et al.*, 1993; Medlin *et al.*, 1988)

PCR Reaction Mix

Volume	*Reagent*
1 µl	Gene Specific Primer (F) (10 µM)
1 µl	Gene Specific Primer (R) (10 µM)
2 µl	dNTP (10 mM)
1 µl	$MgCl_2$ (50 mM)
10 µl	10X Buffer (Invitrogen)
0.4 µl	Platinum Taq DNA Polymerase (Invitrogen)
1.0 µl	cDNA Template
Up to 50 µl	MQ-H_2O

PCR Cycling Parameters

- 94°C 2 min
- 20 cycles:
  - 94 °C 30 sec
  - 65 °C 30 sec
  - 72 °C 1 min
- 8 cycles:
  - 94°C 30 sec
  - 56 °C 30 sec
  - 72 °C 1 min
- 72 °C 5 min

2.3 PCR Based Method of Identifying Novel Lipases in *O. floccosum*

Degenerate primers were used to target two conserved regions including a serine active site and an oxyanion hole in fungi (see Table 2.3) (Herggard *et al*., 2000; Bell *et al*., 2002; Bradner *et al*., 2003).

The primers were designed to produce an amplification product with an expected size of around 220 bp.

Function	Primer Name	Sequence (5' - 3')	Primer Attachment Points
Forward Oxyanion hole prospecting primers	FoxF1	atc gtt ctg gYn KtN MgN gg	
	FoxF2	att gtc ctt KcN KtN MgN gg	
	FoxF3	att tac att KYN ttN MgN gg	
	FoxF4	atc ggc atc RSN ttN MgN gg	542 - 561
Reverse Active site prospecting primers	FacR1	tgc ccc tcc NaK Nga Rtg NSc	
	FacR2	tgc gcc Ncc NaK Rct Rtg NSc	766 - 746

Table 2.3 Degenerate primers targeting two conserved regions, a serine active site and an oxyanion hole in fungal lipases taken from the literature (Herggard *et al.*, 2000, Bell *et al*., 2002; Bradner *et al*., 2003)

PCR Reaction Mix

Volume	*Reagent*
1 µl	Gene Specific Primer (F) (10 µM)
1 µl	Gene Specific Primer (R) (10 µM)
2 µl	dNTP (10 mM)
1 µl	$MgCl_2$ (50 mM)
10 µl	10X Buffer (Invitrogen)
0.4 µl	Platinum Taq DNA Polymerase (Invitrogen)
1.0 µl	DNA Template
Up to 50 µl	MQ-H_2O

PCR Cycling Parameters

- 95°C 1.5 min
- 39 cycles:
  - 95 °C 1 min
  - 55 °C 1 min
  - 72 °C 2 min
- 72 °C 5 min

2.4 Genome Walking

Upstream and downstream sequences of all target genes were isolated using Universal GenomeWalker™ kit designed by Clontech (USA). The protocol was followed according to the manufacturer's instructions.

2.4.1 Oligonucleotide Primer Design

The following parameters were targeted when designing gene specific primers for genome walking:

- Oligonucleotide of 26-30 bp in length
- GC% content (40-60%)
- Complementarity between primer sequences (self dimer formation). A 3' end self dimer with a thermodynamic value (ΔG) of no larger than -5 kcal/mol calculated using the software program Vector NTI (Invitrogen) and an internal cross dimer with a ΔG of no larger than -6 kcal/mol was tolerated.
- A predicted melting temperature of greater than 69 ºC.
- Self-complementarity of primer sequences (hairpin loops); A 3' end hairpin with a ΔG of no larger than -2 kcal/mol and an internal hairpin with a ΔG of no larger than -3 kcal/mol was tolerated.
- No more than three G's and C's in the last six positions at the 3' end of the primer.
- Sequences at the 3' end of primers were not able to anneal to the 3' end of the adaptor primers.
- Nested PCR primer annealed beyond the 3' end of the primary PCR primer

Gene	Primer name	Nucleotide Sequence
Adaptor	AP1	5' GTAATACGACTCACTATAGGGC 3'
Nested Adaptor	AP2	5' ACTATAGGGCACGCGTGGT 3'
phospholipase	PL 1	5' - AGATGATCCAGTCGCTAAGTTCGAGAAG -3'
(*O. floccosum*)	PL 2	5'- GGTCTATCGCCTTTGGTGCCAGAACAA -3'
	PL 3	5'- GCTCTTGGTGTCAAACTCGAAAAAGTCG -3'
	PL 4	5'- GGAGATGGAGGTTGGCTGCTTGGACA -3'
	PL 5	5'- AGGACCTAGACGAGACCCCCATTGGTT -3'
	PL 6	5'- GACTTGATTCGGATGCTCTGCGTTGAG -3'
	PL 7	5'- GCACAACAAGCACAAGATCGGTAACCT -3'
	PL 8	5'- GTAACCTGGTGGCATTTCTTTTTCGAC -3
	PL 9	5'- CTTGCGGTAGTGAAGCGATGGGAATG -3'
	PL 10	5'- AGAAGGGAGCCAATCGGGGAGTCTAC -3'
	PL 11	5'- GGCATCTTGGCTGGGTCTATC -3'
NADH	NAD 1	5'- ACCACTAGAAACAAATCCGTGAGCTAAACC -3'
(*O. floccosum*)	NAD 2	5'- TCCGTGAGCTAAACCTAAAATTATACTTCC -3'
	NAD 3	5'- AGGTTTAGCTCACGGATTTGTTTCTAGTGG -3'
	NAD 4	5'- TTCTAGTGGTTTATTTATATGTGCAGGTGG -3'
Heat Shock	HS 1	5'- CAGAAGCAGCACGATACTGGCAAGACC -3'
(*O. piliferum*)	HS 2	5'- CAGCCGCTCAATTTGCAGTGGACATATC -3'
	HS 3	5'- GGTTCGCCGCATTGTCGTCTTTTAAGC -3'
	HS 4	5'- TCGACAGCTTGACCACCACTTGACTTTC -3'

Table 2.4.1: Primers used in genome walking.

2.4.2 Construction of GenomeWalker™ Libraries

2.4.2.1 *Ophiostoma* Genomic DNA Extraction

Genomic DNA was isolated from mixed mycelial and blastospore cultures of *O. floccosum* and *O. piliferum* shaken for 72 hours at 25 ºC and 180 rpm. DNA was isolated as described in Section 2.2.1. The DNA was visualised by gel electrophoresis using ethidium bromide fluorescence. High quality DNA was indicated by the degree of shearing. DNA that was more degraded produced a greater degree of smearing due to lower molecular weight pieces traveling further in the agarose gel.

2.4.2.2 Blunt Digestion of *Ophiostoma* Genomic DNA

Four subsets of genomic DNA were digested using the restriction endonucleases *Pvu* II, *Eco*RV, *Stu* I and *Dra* I according to the manufacturer's instructions. All restriction enzymes were supplied in the GenomeWalker™ kit (Clonetech USA).

2.4.2.3 GenomeWalker™ Adaptor-Blunt *Ophiostoma* Genomic DNA Ligation

All ligation reagents and adaptors were supplied in the GenomeWalker™ kit (Clonetech USA). The reaction mix was prepared according to the manufacturers instructions and incubated at 16 ºC overnight.

2.4.3 GenomeWalker™ DNA Walking PCR Amplification

The digested libraries containing the blunt adaptors were amplified by nested PCR using adaptor specific primers AP1 and AP2 in conjunction with primary and nested gene specific primers (refer to Table 2.4.1).

2.4.3.1 Primary PCR Reaction Mix

The following reagents were used in a 50 µl Primary PCR reaction:

Volume	*Reagent*
1 µl	AP1 Adaptor Primer (10 µM)
1 µl	Gene Specific Primer (10 µM)

1 µl	dNTP (10 mM)
10 µl	Buffer B (Invitrogen)
2 µl	ELONGase® Enzyme Mix (Invitrogen)
1.5 µl	DNA Template
Up to 50 µl	MQ-H_2O

2.4.3.2 Nested PCR Reaction Mix

Volume	*Reagent*
1 µl	AP2 Adaptor Primer
1 µl	Gene Specific Primer (10 µM)
1 µl	dNTP (10 mM)
10 µl	Buffer B (Invitrogen)
2 µl	ELONGase® Enzyme Mix (Invitrogen)
1.5 µl	DNA Template
Up to 50 µl	MQ-H_2O

2.4.3.3 Primary PCR Cycling Parameters

The DNA Engine® Peltier Thermal Cycler (PTC-200) manufactured by MJ Research Incorporation was used to carry out all PCR reactions. The cycling parameters for the primary PCR reaction were as follows:

- 7 cycles:
  - 94ºC 25 sec
  - 72 ºC 3 min
- 32 cycles:
  - 94 ºC 25 sec
  - 67 ºC 3 min
- 67 ºC for an additional 7 min after the final cycle

2.4.3.4 Nested PCR Cycling Parameters

For the nested PCR amplification, the cycling parameters were as follows:

- 5 cycles:
  - 94 ºC 25 sec
  - 72 ºC 3 min
- 20 cycles:

94 ºC 25 sec

67 ºC 3 min

- 67 ºC for an additional 7 min after the final cycle.

2.5 Cloning and Sequence Analysis

2.5.1 Purification Of DNA Used In Cloning Reactions

PCR products and digested plasmids were purified using E-Gel® CloneWell precast agarose gels with the E-Gel® iBase™ Power System (Invitrogen) according to the manufacturer's instructions.

2.5.2 Vectors and Ligation Reactions

Two cloning vectors were used in this thesis research, the non-patented $pP_{RO}EX$ HTb (Invitrogen) shown in Figure 2.5.2, and PCR4 Topo (Invitrogen). The latter vector was supplied as a component in TOPO TA Cloning® Kit for Sequencing and was used according to the manufacturer's instructions for sequence analysis of select DNA transcripts.

Preparation of $pP_{RO}EX$ HTb Vector

The vector was amplified in *E.coli* DH5α competent cells grown in LB liquid media with ampicillin antibiotic.

Plasmids were isolated using a Qiagen mini prep system. The presence of the correct sized plasmid vector was confirmed by electrophoresis and SYBR® Safe visualisation as described in Section 2.2.4.

The vector was digested with *BamH* 1 and *Pst* I restriction enzymes. The digested plasmids were purified using the CloneWell system (Invitrogen).

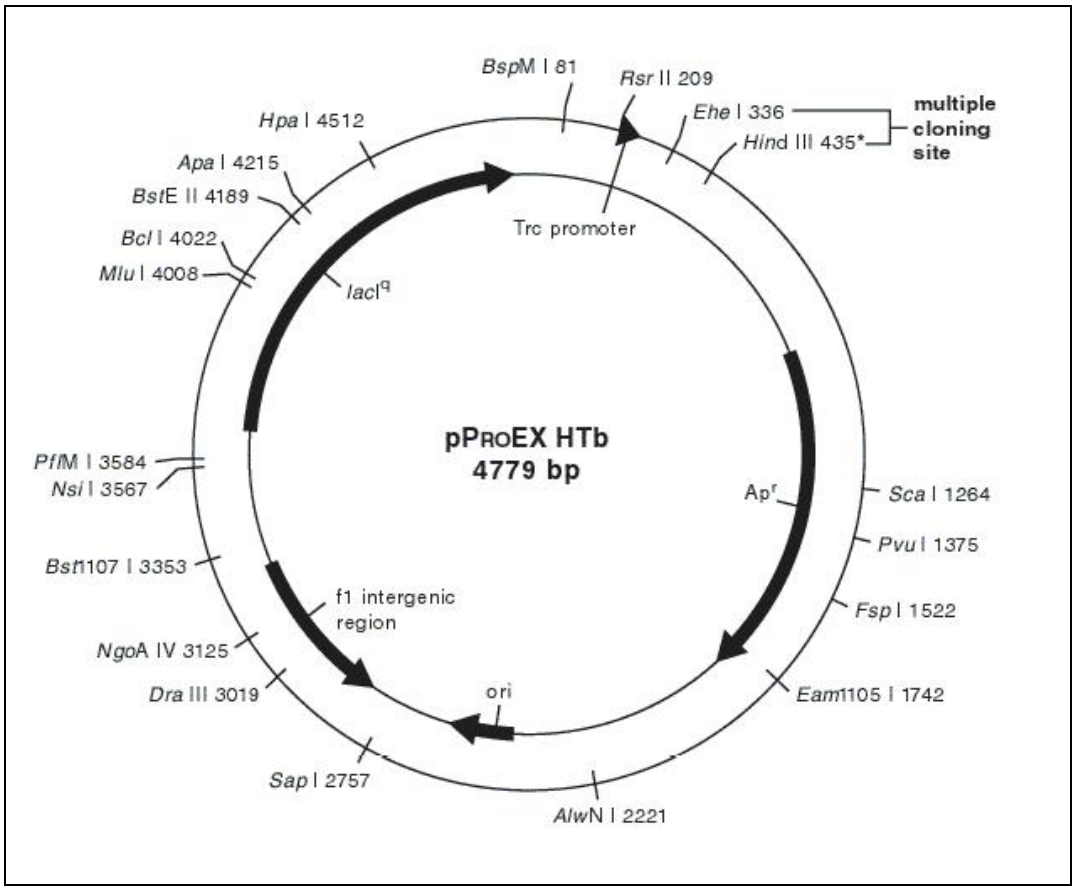


Figure 2.5.2 Map of the non-patented $pP_{RO}EX$ HTb vector (Invitrogen)

Ligation Reaction

A molar ratio of 3 parts insert to 1 part vector was used for the rapid ligation of DNA inserts to vectors to produce circular recombinant molecules.

1. To an autoclaved, 1.5ml microcentrifuge tube, the following was added:

5X Ligase Reaction Buffer	4µl
Vector DNA	3 to 30 fmol
Insert DNA	9 to 90 fmol
T4 DNA Ligase (units)	1 unit (in 1 µl)
Autoclaved distilled water	to 20 µl

2. The samples were mixed gently, centrifuged briefly to bring the contents to the bottom of the tube and incubated at 16 ºC for 4 hours, or overnight at 4 ºC.

2.5.3 Transformation Using Chemically Competent Method

The TOPO TA Cloning® system (Invitrogen), using Top10 chemically competent cells, was used for chemically competent cloning of genetic material for amplification and sequencing. The protocol was followed according to the manufacturer's instructions.

2.5.4 Transformation Using Electrocompetent Methods

All $pP_{RO}EX$ vectored transformations and the majority of TOPO TA Cloning® (Invitrogen) used for amplification and sequencing were transformed using electrocompetent methods. This typically produced much greater transformation efficiencies than the chemical methods. All TOPO TA Cloning® (Invitrogen) were performed according to the manufacturer's instructions.

2.5.4.1 Electroporation

All electroporation was conducted in the University of Waikato Transitional/ Containment Facility (reference number 759), using a BIORAD Gene Pulser™, and a BIORAD Pulse Controller. The machine was set to 250 micro fluorescein isothiocyanate–labeled dextran (µFD) capacitance, and 200 OHMS. The Pulser was then set to 1.8 for the 0.1cm and 0.2 cm electrodes used. One to two microlitres of ligation reaction were transferred to 50 µl of electrocompetent cells that had been thawed slowly on ice, then placed in a 0.1 cm Gene Pulser® Cuvette (BIORAD) that had been chilled on ice. The cuvette was dried and the electrical pulse applied. 950 µl of room temperature SOC medium was added to the transformed cells. The transformed cells were then transferred to a sterile 50 mL Falcon tube and incubated with for at least 1 hour at 37 ^{o}C shaking at 200 rpm. A 200 µl volume of the transformed cells were then plated onto pre-warmed LB medium plates containing 100 ug/mg of ampicillin or 50 ug/mg Kanamycin. In some cases X-Gal (20 mg/ml of X-Gal dissolved into dimethylformamide) was added to the plates at least 1 hour prior to inoculation as suggested in the reference used (Promega Technical Manual No. 042) for use in blue/white screening.

2.5.5 Oligonucleotide Primer Design

The following parameters were targeted when designing gene specific primers for genome walking:

- Primers in frame of target gene
- GC% content (40-60%)
- Complementarity between primer sequences (self dimer formation). A 3' end self dimer with a ΔG of no larger than -5 kcal/mol and an internal cross dimer with a ΔG of no larger than -6 kcal/mol was tolerated.
- Self-complementarity of primer sequences (hairpin loops). A 3' end hairpin with a ΔG of no larger than -2 kcal/mol and an internal hairpin with a ΔG of no larger than -3 kcal/mol was tolerated.
- No more than three G's and C's in the last six positions at the 3'end of the primer

Primer name	Description	Nucleotide Sequence
M13 F	Invitrogen Topo TA Cloning	5' –GTAAAACGACGGCCAG- 3'
M13 R	Invitrogen Topo TA Cloning	5' –CAGGAAACAGCTATGAC- 3'
SP6 Promoter	Orient Express System, Novagen	5'- GATTTAGGTGACACTATAG -3'
T7 Terminator	Orient Express System, Novagen	5'- GCTAGTTATTGCTCAGCGG -3'
Lip_lng pst1	cloning phospholipase large ORF in pPROEX	5'- CTGCAGCTACGCTTGTACCATTCCTGGAAT -3'
Lip_sht Hind R	cloning short phospholipase ORF in pPROEX	5'- GATAAGCTTCTACGCTTGTACCATTCCTGGAAT -3'
Lip_sht Bam F	cloning short phospholipase ORF in pPROEX	5'- GCAGGATCCATTGAGGTATTGTCCAAG -3'
Lip_lng Bam F	cloning long phospholipase ORF in pPROEX	5'- GTCGGATCCAAGCCCATCTACTGGTCC -3'

Table 2.5.5 Primers used in cloning and sequencing of cloning reactions.

2.5.6 DNA Sequencing

Samples sequenced for EST analysis were sent to Macrogen Inc, Seoul, South Korea. Samples were sent as unpurified PCR product with a minimum volume of 20 µl.

All other sequencing was conducted using a MegaBASE DNA Analysis System (Amersham Biosciences) at the University of Waikato DNA Sequencing Facitlity (Hamilton, New Zealand). All DNA and PCR product sent for sequencing was purified using GENECLEAN® kit (Q.BIOgene, U.S.A) according to the manufacturer's instructions. This glassmilk method of purifying DNA resulted in higher yields during the thesis research when compared to column filter methods of purification.

2.5.6.1 Sequence Editing

Sequence editing consisted of 3 steps. Firstly, nucleotide sequences were edited to remove contaminating vector sequence using NCBI VecScreen. Strong and moderate contaminating sequence identified was then manually removed. Secondly, poor quality sequence determined by electropherogram analysis was removed. Thirdly, all sequences were edited to ensure they were in FASTA format with an adequate identifier present within the first 10 characters.

In addition to this analysis all EST sequences analysed with fewer than 300bp were removed.

2.5.6.2 Consensus Sequence Alignment

DNA fragments were aligned into consensus (synonyms: contiguous sequence or contig) using the program ContigExpress (Vector NTI, Invitrogen). Input sequences (fragments) were cut and pasted as chromatogram files or, if no chromatogram files were available, text files in FASTA format. Default settings for clipping and overlap values were used. For greater than 500 fragments being analysed lite contigs were assembled. Forward and reverse sequences were denoted with and x and y, respectively, and forward and reverse constraints were used. Individual UPTs identified as contigs and singlets were exported as text files in FASTA format for further analysis.

2.5.6.3 Prediction of Open Reading Frames

Open reading frames (ORFs) within DNA nucleotide sequences were predicted using two different programs.

ORFinder : The NCBI Open Reading Frame Finder (ORF Finder) was a graphical analysis tool used to find all open reading frames of a minimum size of 100 bp in the select nucleotide sequences. Input sequences were added individually as text files in FASTA format. ORFs were identified using the standard genetic codes. Output sequences were saved as both amino acid and nucleotide sequences in FASTA format.

OrfPredictor: OrfPredictor (Min *et al*., 2005) which is employed via the website https://fungalgenome.concordia.ca/tools/OrfPredictor.html, identified potential ORFs using frames in sequences by predicting the most probable coding regions based on the intrinsic signal of the sequences. Input sequences were added in batches as text files in FASTA format. Output sequences were amino acid sequences in FASTA format with a definition line including the nucleotide identifier, the translation reading frame and the nucleotide positions where the coding region began and ended.

2.5.6.4 Sequence Annotation – Determination of Function

The putative function of predicted ORFs and UPTs was determined in three ways, as follows:

1) Prediction based on sequence homology and annotation to Genbank non-redundant protein sequences (nr) using BLASTx and BLASTp algorithms;
2) Annotation using TargetIdentifier (Min *et al.,* 2005); and
3) *Ab initio* predictions of protein function from amino acid sequence using ProtFun 2.2 server, (Jensen *et al.,* 2002; 2003).

BLASTx: Putative function was determined by BLASTx identified best hit of an annotated sequence. Note the best hit annotated sequence was not always the most homologous sequence identified. Input sequences were text files in fasta format.

All programs using BLAST algorithms were subject to the following qualifiers to indicate the level of similarity between the queried sequence and the annotated hit sequence.

***E*-value**	**Qualifier**
E <=1e-50	Homologous to
E <=1e-30	Highly similar to
E <=1e-10	Similar to
E <=1e-5	Weakly similar to
E <=0.1	Very weakly similar to
E>0.1	Very similar to

Table 2.5.6.4 Qualifying thresholds for BLASTx homology assessment

For annotation using BLASTx and the prediction of ORFs used for codon analysis, a minimum BLAST *E*-value of <=1-15 and a score of <100 was required for the hit to be significant and included in the analysis.

TargetIdentifier: UPTs identified in OF48 were annotated using TargetIdentifier (Fungal Genomics Project, Concordia University). TargetIdentifier used the frames predicted in the pre-run BLASTX (NCBI BLASTall package) results to identify full-length EST derived sequences according to the predicted ORF length and completeness. Predictions of ORF length were categorized according to the following definitions (Min *et al.,* 2005):

Full length: The sequence was considered to be full length if it had at either one of both of the following: (1) A 5' stop codon followed by a start codon or (2) No 5' stop codon but an in-frame start codon present prior to the 10th codon of the subject sequence.

Short full-length: the sequence was considered to be less (short) then the full length if it had an in-frame start codon that was aligned to a position between the 10th to 100th codon downstream from the start codon of the subject sequence.

Ambiguous: the sequence was considered to be ambiguous if it had a 5' stop codon but did not contain a predicted start codon. This type of anomaly probably arose because a sequencing error introduced a frame shift. This can occur in EST sequences generated by single-pass sequencing.

Partial: A sequence that was not assigned to one of the above categories.

Input sequences were added in batches of up to 1000 text files in FASTA format. Output data was formatted into MS Excel spreadsheets with the following fields separated into individual columns: Query identifier, provisional function, prediction of sequence length, predicted start codon position and sequence status as to whether the OED has been completely sequenced.

ProtFun : http://www.cbs.dtu.dk/services/ProtFun/. The ProtFun method queried a large number of other feature prediction servers to obtain information on various post-translational and localisational aspects of the protein as well as isoelectric point and composition of the polypeptide chain, which were integrated into final predictions of the cellular role, enzyme class (if any), and selected Gene Ontology categories of the submitted sequence.

Input sequences consisted of predicted ORF amino acid sequences for all UPTs in OF48 gained using OrfPredictor as described in Section 2.5.5.3. Prior to input, all amino acid sequences in FASTA format were screened to remove symbols other than A C D E F G H I K L M N P Q R S T V W. Sequences shorter than 15 and longer than 4000 amino acids were not compatible with the software.

Output scores consisting of two numbers were given for each prediction. Scores with the highest information content (marked with an arrow by the program) were used to determine predicted cellular role for the putative ORFs. If the highest information score had odds lower than 1, no prediction was made. Classification of cellular role was divided into 12 functional categories based on the scheme developed by Monica Riley for *E. coli* (1993). These categories included:

- Amino acid biosynthesis
- Biosynthesis of cofactors
- Cell envelope
- Cellular processes
- Central intermediary metabolism
- Energy metabolism
- Fatty acid metabolism
- Purines and pyrimidines
- Regulatory functions
- Replication and transcription

2.5.6.5 Identification of Putative Peptidases

Putative peptidases were identified from UPTs using BLASTx algorithms to determine nucleotide and amino acid sequence homology to sequences contained within the database. The *MEROPS* database used hierarchical, structure-based schemes for the

classification of the peptidases. Each was assigned to a Family on the basis of statistically significant similarities in amino acid sequence, and Families that are thought to be homologous are grouped together in a Clan. Input sequences were batches of text files (>5000 files) in FASTA format. Only alignments with an *E*-value <=E-04 was considered significant. Significant matches were presented in report format as submitted sequence identifier, the MEROPS family name, the range of the peptidase or inhibitor unit, active site residues, ligands for catalytic metal ions, the MERNUM of the closest homologue and the E-value for the match.

2.5.6.6 Identification of Conserved Domains Within Putative Proteins

Conserved domains were detected by searching predicted amino acid sequences individually against the Pfam Database (Finn *et al.,* 2006). The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs). Predictions were made to include matches to both Pfam-A and Pfam-B entries.

2.5.6.7 Identification of Transcriptional Elements

Transcriptional elements were predicted by the identification of conserved motifs listed in the literature and using Transcription Element Search Software (TESS) (Schug, J. 2003). TESS is a web tool that identifies binding sites using site or consensus strings and positional weight matrices from the TRANSFAC, JASPSR, IMD, and CBIL-GibbsMat databases. Default settings were used for all searches. Default settings selected for string matches using only core positions for TRANSFAC strings, a maximum allowable string mismatch % (tmm) of 10, a minimum log-likelihood ratio score (ts-a) of 12, and a minimum string length (tw) of 6. The predicted transcriptional elements identified were then manually screened to find homologues of fungal transcriptional elements.

2.6 Real-Time Polymerase Chain Reaction

2.6.1 Sample Preparation

Two sets of nineteen 400 ml conical flasks containing 100 ml of YM media were inoculated with 2 x 10^5 blastospores of *O. floccosum* strain J2122 and O. *piliferum* Cartapip™ 97, respectively. The samples were shaken at 140 rpm in an orbital incubator at 25 ºC. Duplicate samples were destructively sampled every 24 hours for 8 days with the exception of 168 hours when no sample was taken for either culture. Samples were analysed to determine blastospore count, biomass and viability to establish if cultures were representative of previously determined growth curves for both species.

2.6.2 RNA Isolation

40 ml of each culture sampled was filtered through 3 layers of sterile cheesecloth to separate mycelial and blastospore forms of growth. RNA was isolated from the two somatic forms according to the protocol listed in Section 2.2.2.

2.6.3 cDNA Construction

cDNA was constructed from RNA isolated in duplicate from separate cultures. This was done immediately following RNA isolation to ensure minimum degradation of the samples. All cDNA was constructed according the protocol listed in Section 2.2.5. Samples were frozen immediately following construction using liquid nitrogen and stored at -70 ºC until further use.

2.6.4 RT-PCR Primer Design

The following parameters were targeted when designing gene specific primers for RT-PCR:

- Amplicon size of between 75 and 200 bp.
- Complementarity between primer sequences (self dimer formation). A 3' end self dimer with a ΔG of no larger than -5 kcal/mol and an internal cross dimer with a ΔG of no larger than -6 kcal/mol was tolerated.

- Self-complementarity of primer sequences (hairpin loops). A 3' end hairpin with a ΔG of no larger than -2 kcal/mol and an internal hairpin with a ΔG of no larger than -3 kcal/mol was tolerated.
- A melting temperature of between 50 and 65ºC
- No more than three G's and C's in the last six positions at the 3'end of the primer
- GC% content (40-60%)
- Avoided templates with >4 repeats of single bases
- Avoided repeats of Gs or Cs longer than three bases

Ideally, a primer set designed to span an exon splice site is preferable. This enables the detection of contaminating genomic DNA. However, for the genes selected in this thesis no introns were detected.

Gene	Primer	Nucleotide Sequence	Reference
β-tubulin	Tub 2	5'- CCAGAGGCCTCGTTGAAGTA -3'	Tanguay *et al*., 2006
β-tubulin	Tub 3	5'- CCTTGACAGCAATGGCGT -3'	Tanguay *et al*., 2006
NADH	NADH F	5'- GCTCACGGATTTGTTTCTAG -3'	
NADH	NADH R	5'- GAAAATAATGGCATAATTTGAG -3'	
Phospholipase	Plip 1	5'- GATAGCCAGGATACCCGAGTAG -3'	
Phospholipase	Plip 2	5'- CCGATATTGGCTGAAGGCAC -3'	
Phospholipase	Plip 3	5'- GATAGCCAGGATACCCGAGTAG -3'	
Phospholipase	Plip 4	5'- GTCTTCTTTGGCAAGGATGTTG -3'	
Heat shock	HSF	5'- GAGTTCTTCAACAGCCTCCAG -3'	
Heat shock	HSR	5'- CGTAGGTGTTCTCAGTCTCG -3'	

Table 2.6.4 Primers used in RT-PCR. Where applicable a reference has been given. For all other sequences listed the primers were designed by the author.

2.6.5 RT-PCR Reaction Mix

Each reaction contained the following components:

Platinum® SYBR® Green qPCR SuperMix-UDG with ROX	10 ul
Forward primer, 10 pmol	0.4 ul
Reverse primer, 10 pmol	0.4 ul
Template (cDNA generated from 1ug of total RNA)	≥2 ul
DEPC-treated H_2O	to 20 ul

2.6.6 RT- PCR Cycling Parameters

RT-PCR was conducted using ABI 7300 Real-Time PCR System (Applied Biosystems). Reactions were run according to the following program:

50 ºC for 2 minutes hold (UDG incubation)

95 ºC for 2 minutes hold

40 cycles of

95 ºC, 15 seconds

60 ºC, 60 seconds

Melt curve analysis was performed during every reaction to identify the presence of primer dimers and to analyze the specificity of the reaction.

2.7 Cloning of an *O. floccosum* Phospholipase Gene in *E. coli*

2.7.1 PCR Amplification of The Phospholipase Gene

Phospholipase (PLIP)-Lg and PLIP-St ORFs were amplified by PCR from *O. floccosum* genomic DNA using PLIP-Lg and PLIP-St specific primers (listed in Table 2.7.1) containing BamH1, Pstl and Hind III restriction sites using the protocols listed below.

Primer name	Description	Nucleotide Sequence
PLIP-Lg BamH1	cloning long phospholipase ORF in pPROEX	5'- GTAGGATCCATGGTCTCCTTGGGCAAAGT -3'
PLIP-Lg Pstl	cloning long phospholipase ORF in pPROEX	5'- GTGACTGCAGATACATATCTACGCTTGTACCAT -3'
PLIP_St Hind R	cloning short phospholipase ORF in pPROEX	5'- GATAAGCTTCTACGCTTGTACCATTCCTGGAAT -3'
PLIP_St Bam F	cloning short phospholipase ORF in pPROEX	5'- GCAGGATCCATTGAGGTATTGTCCAAG -3'

Table 2.7.1 Primers used to amplify PLIP-Lg and PLIP-St ORFs in *O. floccosum*

PCR Reaction Mix

Volume	*Reagent*
1 µl	Gene Specific Primer - forward (10 µM)
1 µl	Gene Specific Primer - reverse (10 µM)
1 µl	dNTP (10 mM)
5 µl	Buffer A (Invitrogen)
5 µl	Buffer B (Invitrogen)
2 µl	ELONGase® Enzyme Mix (Invitrogen)
1.5 µl	DNA Template
Up to 50 µl	MQ-H_2O

PCR Cycling Parameters

The DNA Engine® Peltier Thermal Cycler (PTC-200) manufactured by MJ Research Incorporation was used to carry out all PCR reactions.

- 36 cycles:

  94ºC 30 sec

  72 ºC 2 min

- 67 ºC for an additional 7 min after the final cycle

2.7.2 Ligation and Cloning Of Phospholipase Genes Into $pP_{RO}EX$ HTb

Amplified product of the correct size, determined by gel electrophoresis, was ligated into $pP_{RO}EX$ HTb vector containing a $(his)_6$ spacer region and TEV protease cleavage site as described in Section 2.5. The recombinant ligate was transformed using electrocompetent methods into *E.* coli Top10 cells (Invitrogen). A negative control containing empty $pP_{RO}EX$ HTb vector was also cloned into *E.* coli Top10 cells in a separate reaction.

Recombinant transformants were cultivated overnight on ampicillin antibiotic LB medium containing 1.5% agar. Single colonies were selected and cultivated in 500 ml of LB media containing ampicillin at a pH of 7.0 in a 2 L conical flask shaken at 180 rpm at 37 ºC.

2.7.3 Isolation of Recombinant Protein

Overnight cultures containing recombinant *E. coli* were centrifuged at 13000 rpm for 10 minutes at room temperature to pellet the cells. The supernatant was removed and to the pelleted cells 8 ml of 50 mM phosphate buffer containing 150 mM NaCl at pH 8.8 was added. The resuspended cells were once again centrifuged at 13000 rpm for 10 minutes at room temperature to pellet the cells and 6 ml of phosphate buffer was added resulting in a resuspended culture with 7 ml of total volume. To this approximately 26 mg of Complete mini EDTA protease inhibitor cocktail tablets (Roche, Germany) was added. To disrupt the cells the culture was sonicated (Misonix Inc, NY, U. S. A.) in 30 second bursts for 3.5 minutes at level 4. The disrupted cells were then centrifuged at 13000 rpm for 10 minutes at 4 °C and the supernatant collected for further use.

Purification of Protein Using The Nickel Pull-Down Method

To 50 µl of nickel sepharose beads (GE healthcare, Sweden) in a sterile 1.5 ml microcentrifuge tube, 1 ml of phosphate buffer was added. The beads were left to settle for 2-5 minutes before the samples were spun at low speed (2000 rpm) for 30 seconds. The supernatant was removed and to the beads 200 µl of protein was added. The samples were shaken at room temperature for 10 – 15 minutes and the left once more to settle for 2- 5 minutes and spun at low speed (2000 rpm) for 30 seconds. The supernatant was again removed and the beads washed with 1 ml of phosphate buffer as before and spun at low speed (2000 rpm) for 30 seconds before removing the supernatant. The wash was repeated twice. To visualise any bound protein an equal volume of 4X Quench was added to the beads. The samples were then electrophoresed using SDS-PAGE methods as described below.

2.7.3.1 SDS- Polyacrylamide Gel Electrophoresis

Protein samples were analysed using discontinuous SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE). Stacking gels consisted of 4% acrylamide (BIORAD, 30% acrylamide solution, 37.5:1 acrylamide:bis acrylamide) in stacking gel buffer (1.0 M Tris pH 6.8) and were used to focus the protein band. Polymerisation was initiated by the addition of 0.4% w/v APS and 0.05% v/v TEMED. The resolving gel consisted of 10% acrylamide (BIORAD, 30% acrylamide solution, 37.5:1 acrylamide:bis acrylamide) in resolving gel buffer (1.5 M Tris pH 8.8) to separate protein by molecular weight. Polymerisation was initiated by the addition of 0.1% w/v APS and 0.07% v/v TEMED.

Prior to loading, 4X Quench was mixed with protein samples to a final 1X concentration. Samples were then heated at 90 °C for 5 minutes and centrifuged briefly before loading into wells in the stacking gel. Gels were run at 15 mA until the dye front entered the resolving gel and then at 20 mA until it reached the end of the gel. A water-cooled Hoefer Mighty Small II gel box and EPS-600 Electrophoresis power supply (Amersham Pharmacia Biotech) were used. Protein molecular weight was estimated by comparison with the broad rang Precision Unstained Protein Standards (BIORAD).

Coomassie Blue Staining of PAGE Gels

SDS-PAGE gels were stained using a 'Fast' coomassie blue staining method. To the SDS-PAGE gel, 100 ml of Fairbanks A staining solution was added (0.05% coomassie blue, 25% isopropanol, 10% acetic acid). The gel in solution was then heated in a microwave until boiling point was reached (approximately 2 minutes) and then allowed to cool at room temperature for approximately 5 minutes while shaking. The Fairbanks solution A was then discarded and the gel rinsed briefly with distilled water to remove residual solution. To the gel 100ml of Fairbanks B staining solution was added (0.005% coomassie blue, 10% isopropanol, 10% acetic acid). The gel in solution B was then heated, left to cool and rinsed with distilled water as before. 100 ml of Fairbanks C staining solution was then added (0.002% coomassie blue, 10% acetic acid). The gel in solution C was heated, left to cool and rinsed with distilled water. The final step involved the addition of 100 ml of Fairbanks D staining solution (10% acetic acid). Again the gel was heated, left to cool and rinsed before being ready for analysis.

2.7.4 Assay of Esterase Activity

Esterase activity was measured using a discontinuous p-nitrophenyl (*p*NP) assay according to the following protocol (Janssen *et al*., 1994; Winkler and Stuckmann, 1979):

> The para-nitrophenol (*p*NP) substrates, *p*NP-palmitate and *p*NP-laurate, were dissolved in 1 ml ethanol at 1 mg/ml and then disbursed in 8ml of 0.1 M MOPS buffer (0.1 M 3-[N-Morpholino] propanesulfonic acid, 5 mM $CaCl_2$, 0.1% Triton X-100, pH 7.2) by sonicating continuously for 1 min. To 900 ul of substrate emulsion, 100 ul of sample was added. The solution was mixed thoroughly and incubated at 37 ^{o}C for 2.5 hours. As a result of lipolysis (*p*NP) from p-nitrophenyl esters of fatty acids was released producing a yellow colour (Winkler and Stuckmann, 1979). The carbon chain length of the esters used was dependent on the lipase specificity. The reaction was stopped by adding 500 ul of stop reagent. Tubes were chilled on ice for 10 min then centrifuged at 10,000 x g for 5 min The absorbance was read at 400nm. 1 unit of lipase was defined as the release of 1 µmole *p*NP per min under the assay conditions.

Stop reagent was produced by adding 20 ml of 2.8% $CuCl_2.H_2O$ slowly to 40 ml 6.85% $Na_3PO_4.12H_2O$ with continuous stirring. The solution was centrifuged at 3,000 x g for 5 min at room temperature. The pellet was then washed twice with 100 ml 1.91% $Na_2B_4O_7.10H_2O$ and resuspended in 100 ml of 1.91% $Na_2B_4O_7.10H_2O$ containing 6 g NaCl. Note: The stop reagent was stored for at least 2 days and mixed well before use.

3 *In vitro* Growth of *Ophiostoma floccosum* and *Ophiostoma piliferum*

3.1 Introduction

Ophiostoma floccosum and *Ophiostoma piliferum* are polymorphic fungi and can grow in both a filamentous and a yeast-like somatic state. The nutritional requirements of these fungi in their natural environment was discussed in Section 1.2.1.2.1 *Ophiostoma* species are readily grown on a variety of different growth media under laboratory conditions. *O. floccosum* and *O. piliferum* have previously been grown in a stirred tank fermenter (Haryati, 2001). An albino isolate of *O. piliferum* known as Cartapip™ 97, was produced on a commercial scale from fermentation at >10,000 L^3 (Farrell *et al*., 2005).

Within a species, different strains may perform differently in terms of growth rate, as well as other factors including enzyme production and secretion, and the degree of melanisation. Therefore, it was important within the scope of this PhD research to determine the standard growth of *O. floccosum* strain J2122 and *O. piliferum* strain 97 (commercialised as the biocontrol and pitch reducing product Cartapip™97 as described in Wendler *et al*., 1992; Blanchette *et al*., 1992; Hoffmann *et al*., 1992 and as reviewed by Farrell *et al*., 1997) under the specific *in vitro* conditions used for molecular analysis. By understanding the growth of these strains under these conditions we can better relate their morphology to their molecular functioning.

3.1.1 Specific Aims and Objectives For *In vitro* Growth of *Ophiostoma floccosum* and *Ophiostoma piliferum*

The specific objective for this chapter was to profile the somatic forms of *O. floccosum* J2122 and *O. piliferum* 97 strains during *in vitro* growth under the laboratory conditions used to produce the cells used for molecular analysis and to compare this with two other typical laboratory conditions.

3.2 Profile Of *O. floccosum* and *O. piliferum* Growth Under Specific *In vitro* Experimental Conditions

The growth of *O. floccosum* strain J2122 and *O. piliferum* strain 97 in sterile YM media, with an unadjusted pH of 5.8, and at a temperature of 25°C, was profiled under three different physical conditions.

Condition 1) Shaken liquid culture

Condition 2) 8L stirred tank bioreactor (fermenter)

Condition 3) Stationary liquid culture

Profiling here is defined as cell density under conditions of growth, as determined by blastospores per ml, as well as the dry-weight of filtered biomass per ml. Viability of the cultures was also assessed by colony forming units (CFU).

Condition 1 represented the standard laboratory conditions for *Ophiostoma* growth used for DNA and RNA isolation and related experiments throughout this PhD thesis as described in Sections 2.2.1 and 2.2.2 respectively. Conditions 2 and 3 were very different commonly used methods of cultivating liquid fungal cultures in the laboratory. Therefore, it was expected that growth data produced under these three conditions would provide a broad scope of the level and type of growth the two strains could achieve *in vitro*.

3.2.1 *O. floccosum* Growth Under Shaking Conditions in Liquid Culture (Condition 1)

Measurements for the construction of a standard growth curve (Figure 3.2.1A) for *O. floccosum* strain J2122 were made from replicate samples ranging from 3 – 16 cultures. Cell density was determined by the number of blastospores of the culture in terms of cells per ml as well as the dry-weight of filtered biomass produced in 10 ml of culture as described in Sections 2.1.3 and 2.1.4, respectively. Viability was determined, as described in Section 2.1.5. Growth was described as early, mid and late phase growth as opposed to the more conventional early, log and stationary stages. The key difference between the two descriptive methods is that early phase described in this research incorporated some of the exponential phase of growth. This adapted method was used as determining cell density was difficult to achieve in pre-exponential growth unless a large inoculum was used.

During late phase growth (> 96 hours), *O. floccosum* J2122 strain typically produced between 10^8 and 10^9 blastospores per ml. The number of CFUs produced from the culture during this phase was approximately 10 fold greater than the level of blastospores observed, between 10^9 and 10^{10} CFUs/ml. The highest level of biomass was achieved at 144 hours with an average of 3.96 g/L.

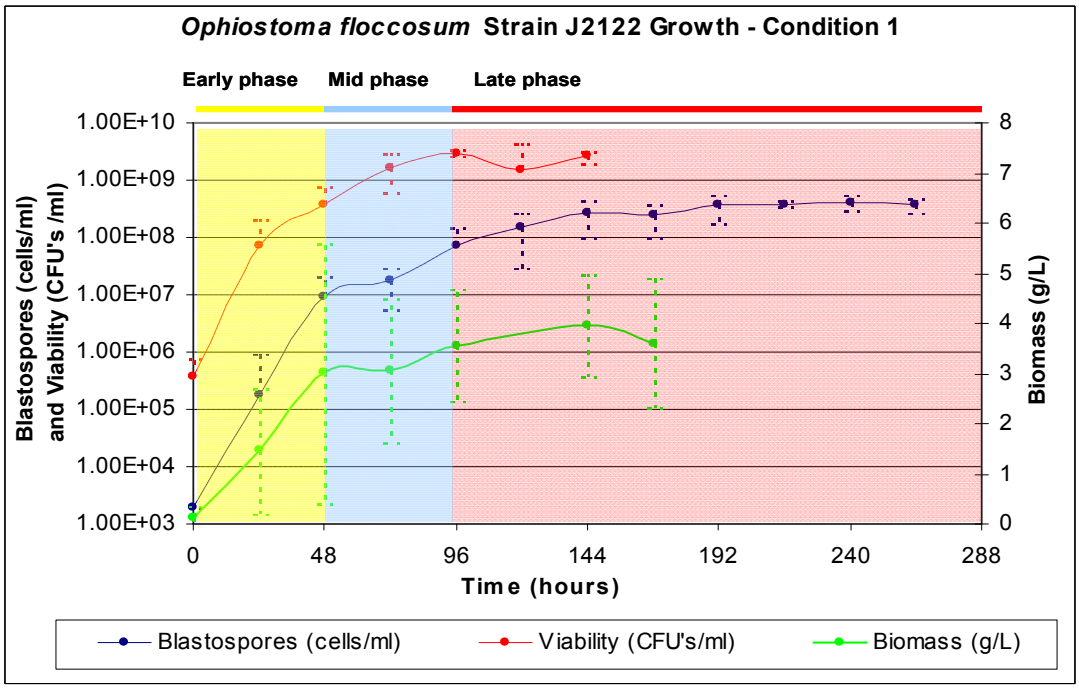


Figure 3.2.1.A Standard growth curve of *O. floccosum* strain J2122.

Cultures were examined microscopically throughout growth. Morphologically many distinguishing features were identified that were characteristic of each growth phase. A key point to consider when interpreting these growth results was that fungal growth is highly variable even within a strain. Mycologists have amply confirmed the incessant tendency of filamentous fungi toward spontaneous change and this mutability has become recognized as one of the outstanding traits of this group (Foster, 1951). Within a strain, despite asexual reproduction, variation in growth and morphology can be observed as a result of differences in external stimuli such as temperature, nutrient availability and pH as well as other factors such as growth stage and the age of mother cells.

Early phase of growth (0-48 hours):

The cultures were typically comprised of small fragments including mycelial fragments (Figure 3.2.1B(A)), blastospores (Figure 3.2.1B(B)) and 'detritus' material

(Figure 3.2.1B(C)). It was hypothesised that 'detritus' material consisted of cellular debris as well as crystalline and coagulated polysaccharide material contained within the media. The media on visual examination appeared to be clear with turbidity increasing as growth progressed. Mycelial growth was evident during microscopic examination. These mycelial fragments were typically not clumped. Some blastospores were evident microscopically, these were usually elongated with only a single bud. Many of these blastospores appeared more like budding small mycelial fragments and were often difficult to distinguish from mycelial matter. As growth progressed, multiple buds became evident on blastospores as can be seen in Figure 3.2.1B(D) taken at 48 hours.

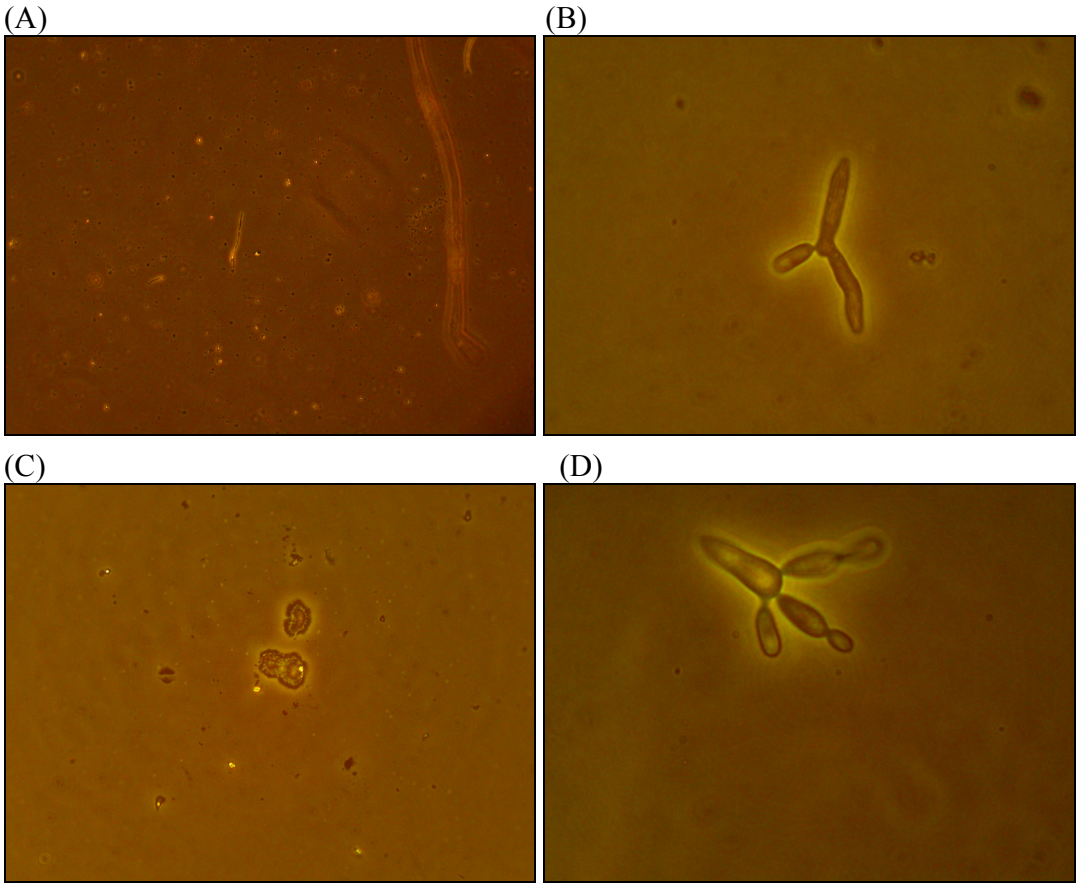
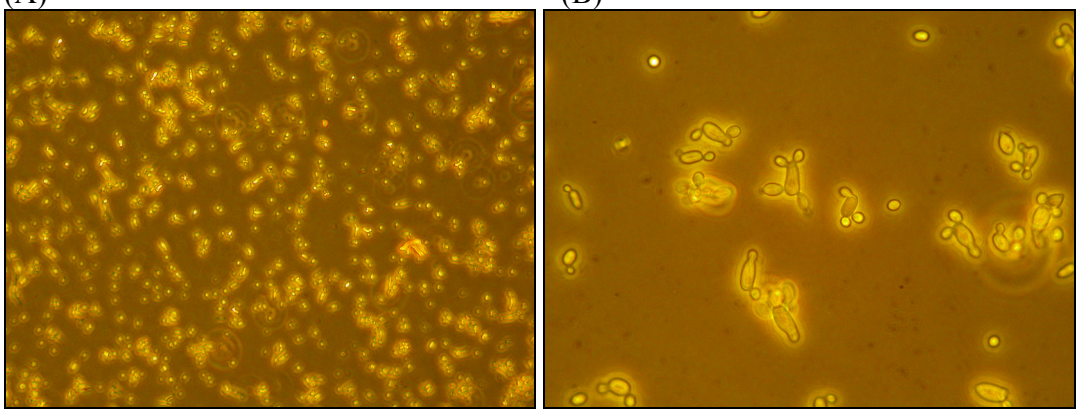


Figure 3.2.1B *O. floccosum* early phase growth. (A) 40X magnification, 2 X digital magnification. Early phase culture taken at 24 hours containing few small mycelial fragments and detritus material. No blastospores were present in this image.
(B) 40X magnification, 4 X digital magnification. Elongated blastospore present at 24 hours.
(C) 100X magnification, 2 X digital magnification. Detritus material observed within the media at 24 hours.
(D) 100X magnification, 4X digital magnification. Elongated blastospore containing multiple buds taken at 48 hours of growth.

Mid Phase of Growth (48-96 hours):

Mycelial material was very apparent in cultures at the mid phase of growth. It was typically clumped into small balls, presumably due to the circular motion of the shaking liquid (this tight circular clumping was not observed under static conditions). There was a greater proportion of blastospores to mycelia in the media (as can be seen in Figure 3.2.1C(A)) compared to early phase growth. Blastospores typically produced multiple buds, displaying characteristic 'foot shaped' cells as shown in Figures 3.2.1C (B) and (C). The culture became increasingly more turbid in appearance as growth progressed.

(A) (B)



(C)

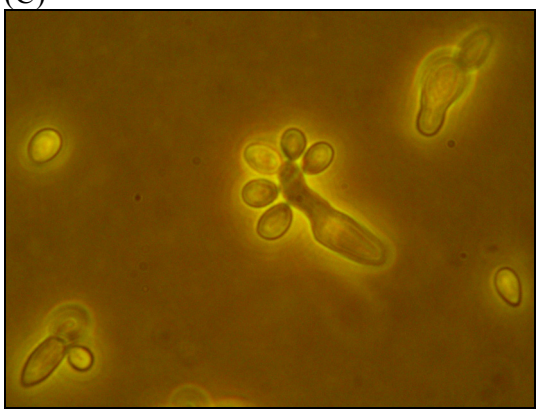


Figure 3.2.1C *O. floccosum* mid phase growth**. (A)** 40X magnification, 3 x digital magnification. Mid phase culture photographed at 72 hours. Large numbers of blastospores evident.
(B)100X magnification, 2X digital magnification. Blastospores displaying multiple budding taken at 72 hours of growth.
(C) 100X magnification, 4X digital magnification. Blastospores at 72 hours of growth.

Late phase of growth – stationary phase (96 + hours):

The cultures were very turbid in appearance. Upon microscopic examination at 40 X magnification, blastospores were smaller with fewer if any buds as shown in Figure 3.2.1D. In many cultures, fewer mycelial balls were present with mycelia appearing more evenly distributed throughout the culture as shown in Figures 3.2.1D(A) and (C).

Under microscopic investigation, a number of hyphal fragments were present (Figures 3.2.1D(B) and (E)). These often produced short buds rather than the typical elongated apical ends usually observed in actively growing mycelia.

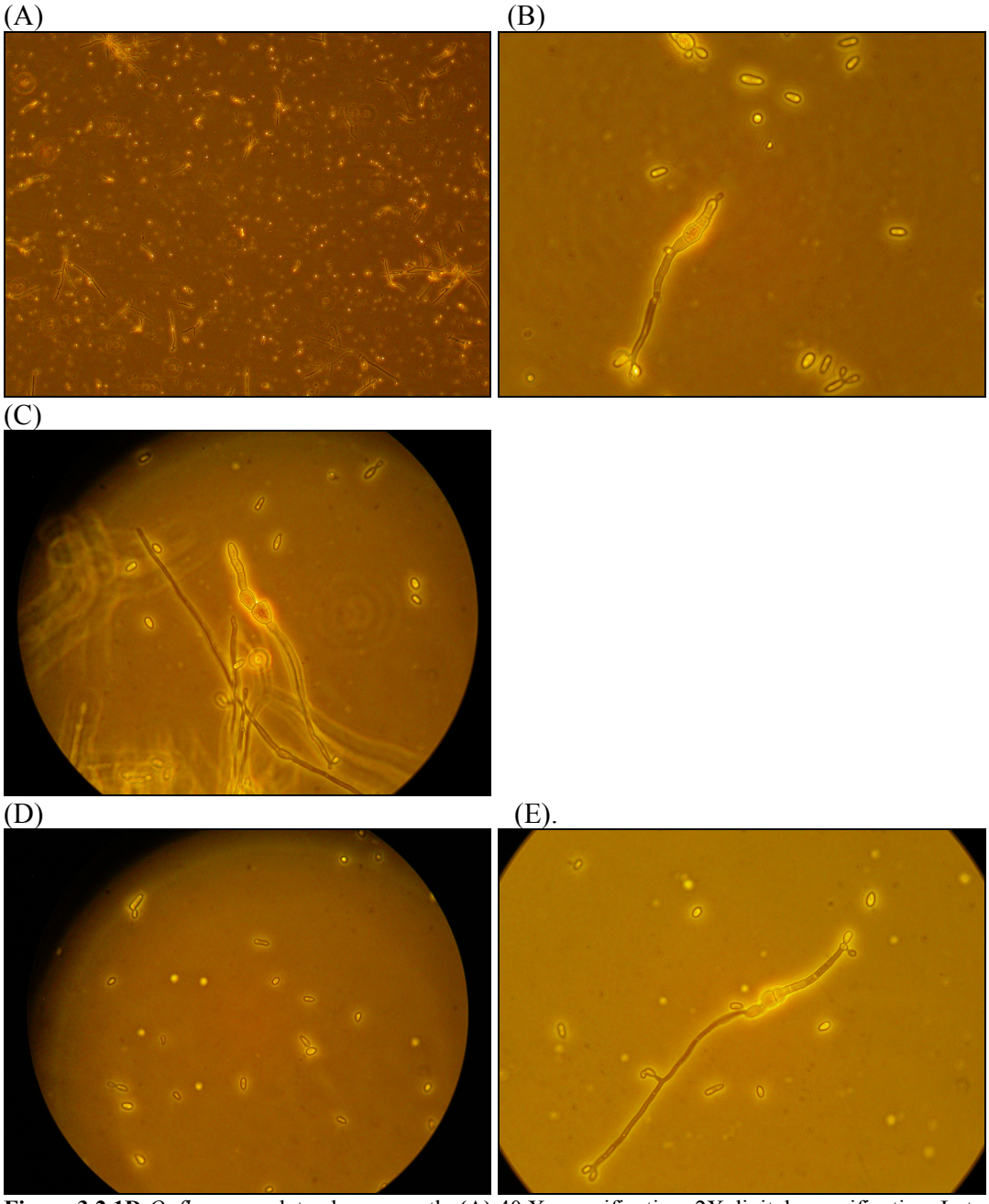


Figure 3.2.1D *O. floccosum* late phase growth. **(A)** 40 X magnification, 2X digital magnification. Late phase culture at 168 hours of growth.
(B) 100X magnification, 2X digital magnification. Blastospores with small mycelial fragment possessing budding apical ends at 168 hours of growth
(C) 100X magnification. Mycelium and blastospores at 168 hours of growth.
(D) 100X magnification. Blastospores at 192 hours of growth
(E) 100X magnification. Blastospores and a mycelial fragment possessing budding apical ends at 192 hours of growth

3.2.2 Comparison Of *O. floccosum* Growth Under Conditions 1, 2 and 3.

During growth in the stirred tank bioreactor (fermenter), pH as well as temperature, aeration and agitation were constantly monitored. As pH was not actively monitored under conditions 1 and 3, growth data produced under fermenter conditions is listed separately in this section.

Sterile YM media was prepared using standard methods as described in Section 2.1.5. Duplicate fermentation runs produced consistent results in terms of rate of growth, biomass, viability and pH and were used to determine average growth rate values, as shown in Figures 3.2.2A and 3.2.4A.

During growth in the stirred tank fermenter, *O. floccosum* J2122 strain produced between 10^8 and 10^9 blastospores /ml. Maximal viability was observed at 96 hours with 2.36 x 10^9 CFU per ml. The highest level of biomass was achieved at 84 hours with an average of 7.36 g/L of culture. pH steadily decreased ranging from an initial pH 5.8 to 3.9 at its lowest recorded level. The fermenter cultures were stopped after a minimum of 96 hours at which stage the culture began to show decreases in all three of the growth factors measured.

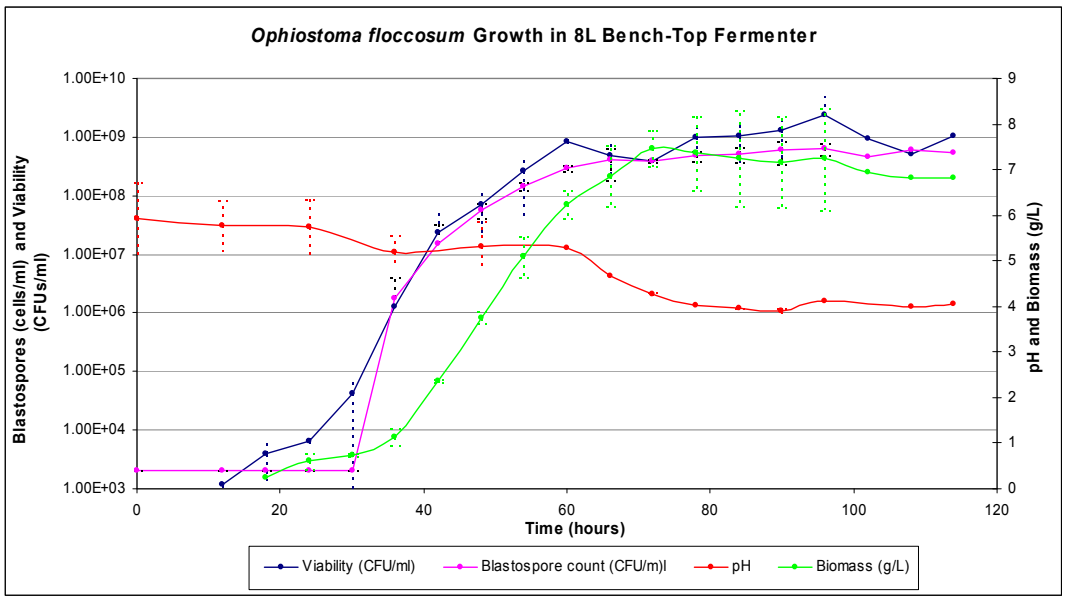


Figure 3.2.2A *O. floccosum* strain J2122 growth in a 8L working volume bench top stirred tank fermenter.

Three separate graphs summarising growth in terms of the three factors measured were plotted to compare *O. floccosum* J2122 growth under conditions 1, 2 and 3 (Figures 3.2.2B, 3.2.2C and 3.2.2D).

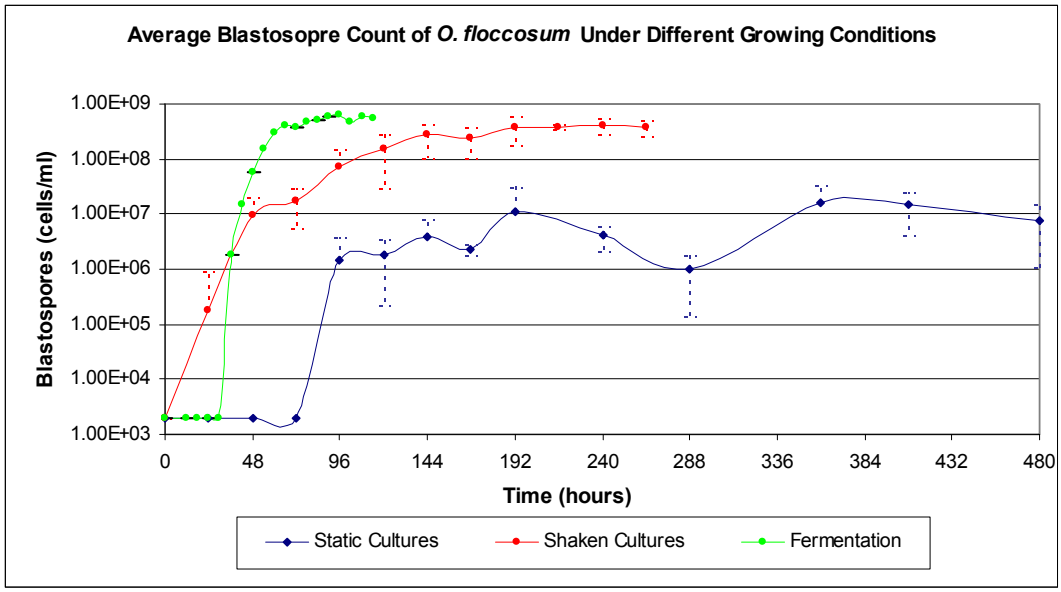


Figure 3.2.2B Comparison of the average blastospore count measured in cells/ml of *O. floccosum* strain J2122 over time under conditions 1, 2 and 3 for growth.

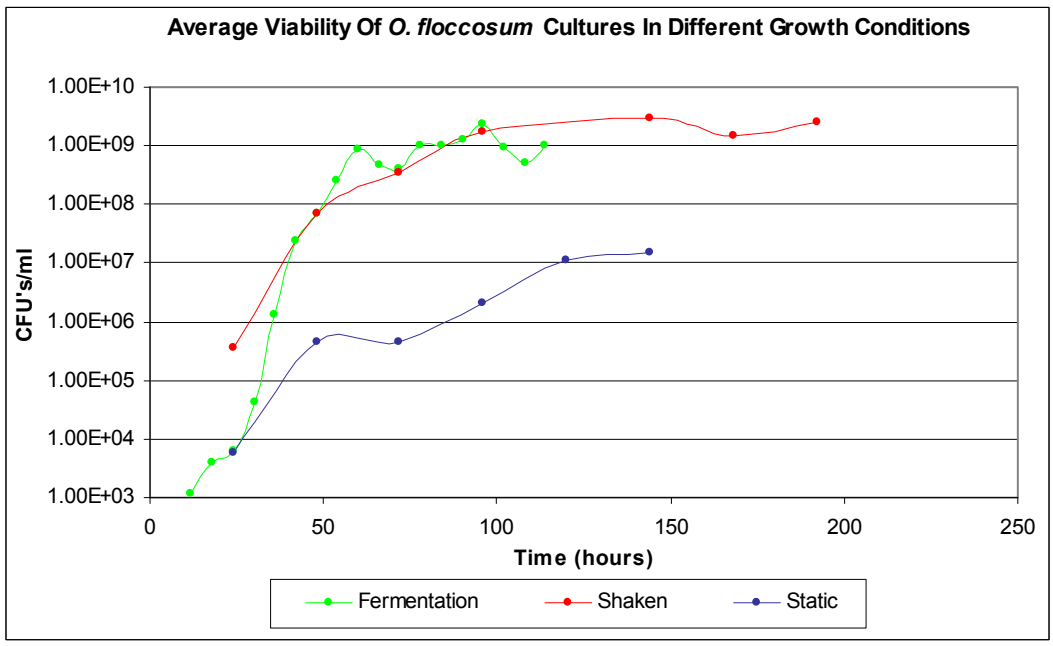


Figure 3.2.2C Comparison of the average viability measured in CFUs of *O. floccosum* strain J2122 over time under conditions 1, 2 and 3 for growth.

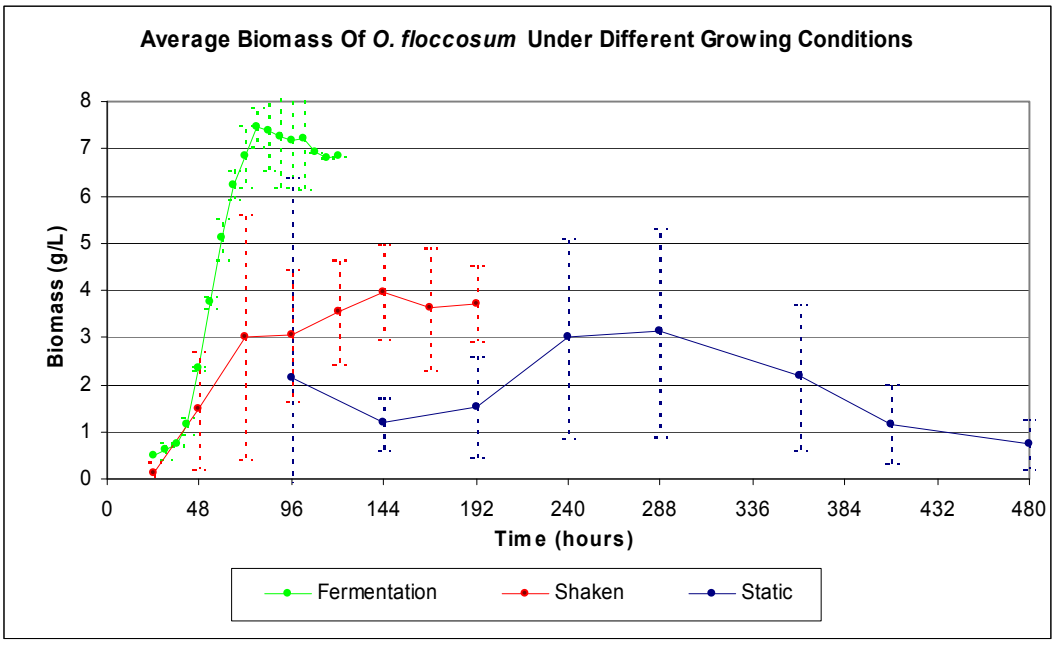


Figure 3.2.2D Comparison of the average biomass measured in g/L of *O. floccosum* strain J2122 over time under conditions 1, 2 and 3 for growth.

From the growth data we can determine that both cell density and viability were less for all growth factors measured in condition 3, stationary growth, compared to the other two conditions. For *O. floccosum* strain J2122, stirred tank fermentation conditions produced the highest blastospore count (6.43 x 10^8 cells/ml) and the greatest biomass (7.46 g/L) in the shortest period of time compared to both shaking and static growth conditions. However, viability was slightly higher (2.95 x 10^9 compared to 2.36 x 10^9) under shaken growth (condition 2) compared to stirred tank fermentation.

The blastospore doubling rate during exponential growth was calculated for the three conditions using the following equation:

$$\text{Doubling time} = \ln 2 / ((\ln (A/Ao)) / t)$$

where A = cell density (e.g., OD) at time t; Ao = initial cell density

The blastospore doubling rate during exponential growth was calculated for the three conditions.

Condition 1 = 3.92 hours

Condition 2 = 1.48 hours

Condition 3 = 2.51 hours

Growth under stirred tank fermentation conditions produced the highest rate of blastospore production with a blastospore count doubling time of 1.48 hours during exponential growth.

3.2.3 *O. piliferum* Growth Under Shaking Conditions in Liquid Culture (Condition 1)

The standard growth of *O. piliferum* strain 97 was measured and evaluated by the same methodology as previously described for *O. floccosum* strain J2122.

Under condition 1, *O. piliferum* strain 97 late phase growth typically produced between 10^8 and 10^9 blastospores per ml with an approximate 100 fold greater number (between 10^{10} and 10^{11}) of CFU per ml measured in the same phase. The highest level of biomass was achieved at 144 hours, well within the late phase, with an average of 6.1 g/L of culture.

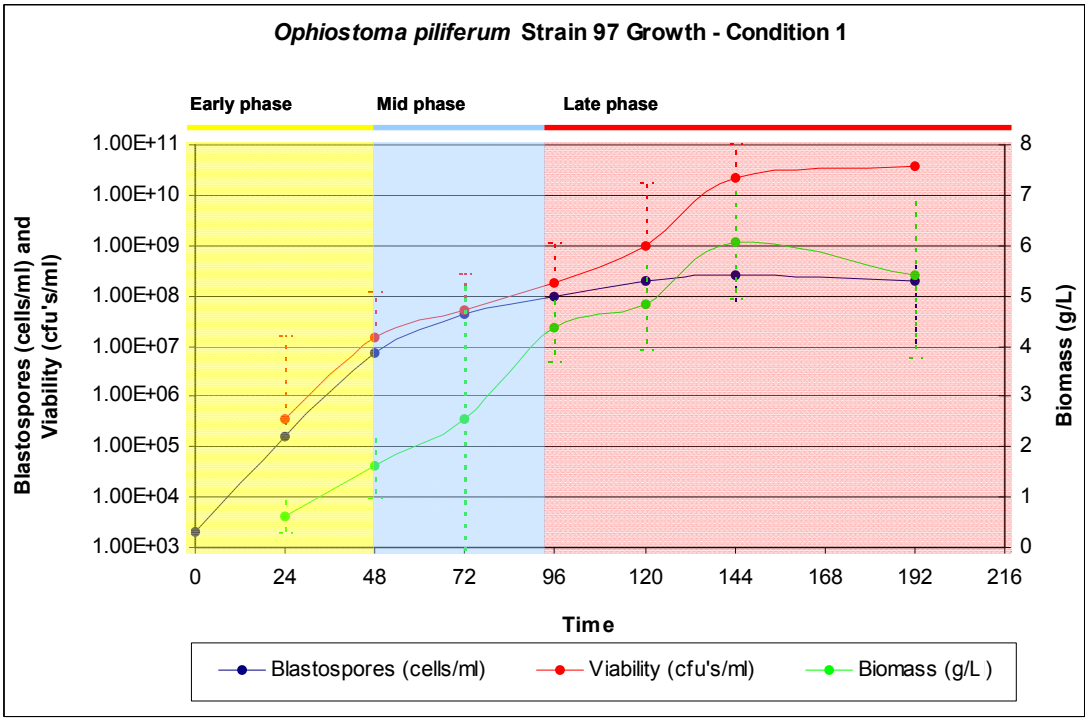


Figure 3.2.3A. Standard growth curve of *O. piliferum* 97 strain.

Cultures were examined microscopically throughout growth. Morphologically many distinguishing features were identified that were characteristic of each growth phase.

Early phase of growth (0-48 hours):

The cultures were typically comprised of small fragments with lots of 'detritus' material as shown in Figures 3.2.3B (A) and (B). The media was clear in appearance with turbidity increasing as growth progressed. Some mycelial growth was evident during microscopic examination at 40X and 100X magnification. These mycelia fragments were typically not clumped. Some blastospores were evident, these were usually elongated with only a single bud (Figure 3.2.3B (D)).

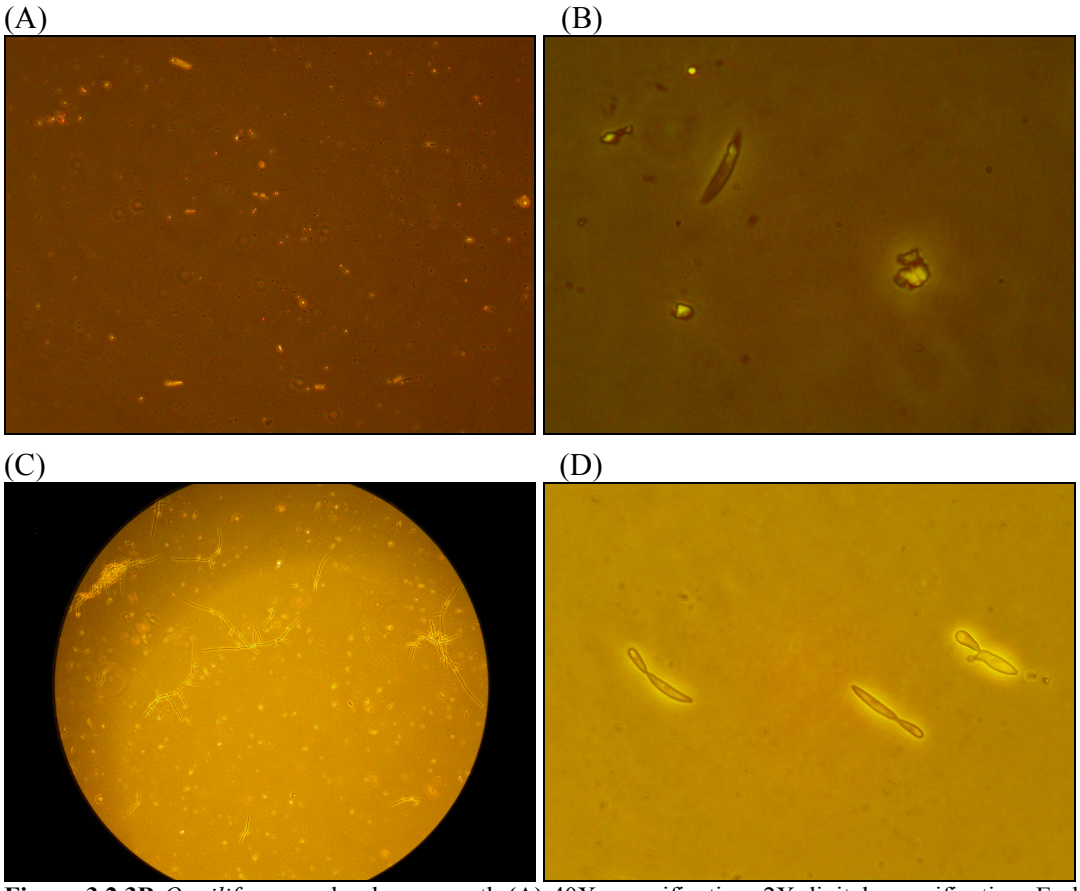


Figure 3.2.3B *O. piliferum* early phase growth **(A)** 40X magnification, 2X digital magnification. Early phase culture at 24 hours.
(B) 100X magnification, 3X digital magnification. A photograph taken at 24hours. A single blastospores and some detritus material were observed.
(C) 40X magnification. Photograph of a typical culture at 48 hours growth.
(D) 100X magnification, 3X digital magnification. Elongated blastospores at 48 hours growth.

Mid Phase of Growth (48 - 96 hours):

Some mycelial clumping into small balls was observed. There were much greater numbers of blastospores as shown in Figure 3.2.3C (A) and (B) compared to early phase cultures. Blastospores produced multiple buds, forming 'foot-shaped' cells.

These blastospores were less elongated than in earlier stages of growth. However, they were proportionally larger than blastospores observed at the same stage in *O. floccosum*. The cultures became increasingly more turbid as growth progressed.

(A) (B)

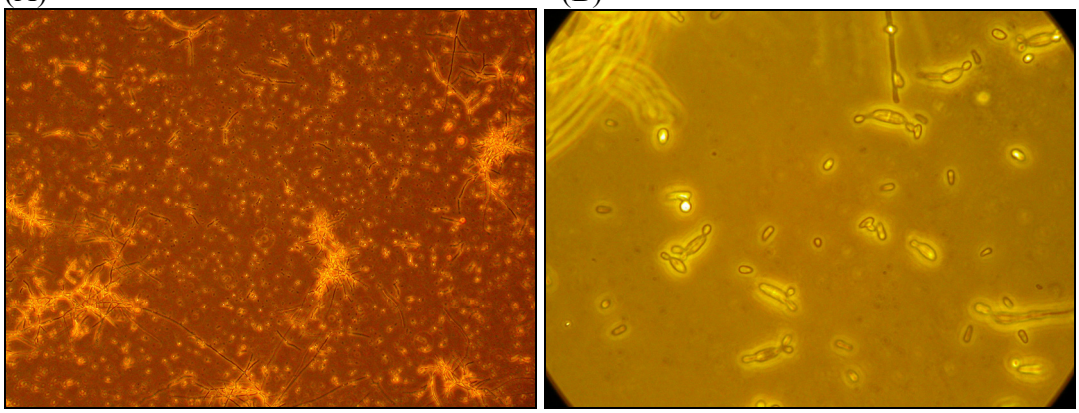


Figure 3.2.3C *O. piliferum* mid phase growth. **(A)** 40X magnification, 2x digital magnification. Typical mid phase culture photographed at 72 hours.
(B) 100X magnification. Mycelia and foot shaped blastospores at 72 hours of growth.

Late phase of growth (96 + hours):

The cultures were very turbid in appearance with fewer mycelial balls observed compared to early and mid phase growth. Blastospores were typically smaller with few if any buds. Compared to *O. floccosum* J2122 cultures at the same stage, there was proportionally a much greater volume of mycelia present (Figure 3.2.3D (B) and (C)). This was likely reflected in the greater levels of biomass but relatively equal numbers of blastospores in the *O. piliferum* 97 cultures as compared to *O. floccosum* J2122 cultures for the same stage of growth (Figures 3.2.1A and 3.2.3A).

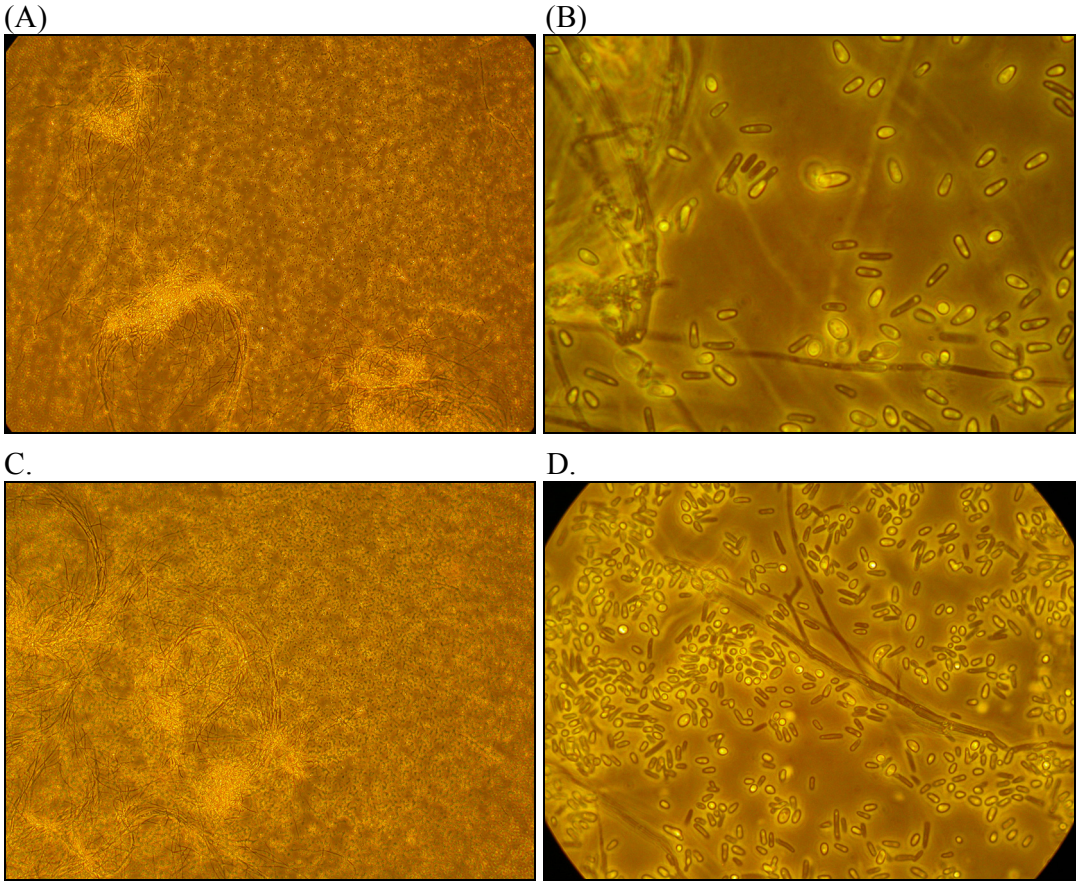


Figure 3.2.3D *O. piliferum* late phase growth. **(A)** 40X magnification, 2X digital magnification. Typical *O. piliferum* 97 culture at 144 hours.
(B) 100X magnification, 2X digital magnification, Blastospores and mycelia at 144 hours.
(C) 40X concentration, 2X digital magnification. Typical *O. piliferum* 97 culture at 168 hours.
(D) 100X magnification. Mycelia and blastospores at 192 hours.

3.2.4 Comparison of *O. piliferum* Growth Among Conditions 1, 2 and 3.

Stirred tank fermentation was performed as described in Sections 2.1.8. and 3.2.2. During fermentation growth (figure 3.2.4A) *O. piliferum* strain 97 produced between 10^8 and 10^9 blastospores /ml. Viability peaked at 98 hours with 5.7 x 10^{11} CFU more than 100 fold greater than the highest number of blastospores recorded during growth. The highest level of biomass was achieved at 54 hours with an average of 2.43 g/L of culture. The culture pH ranged from 4.6 to 7.93.

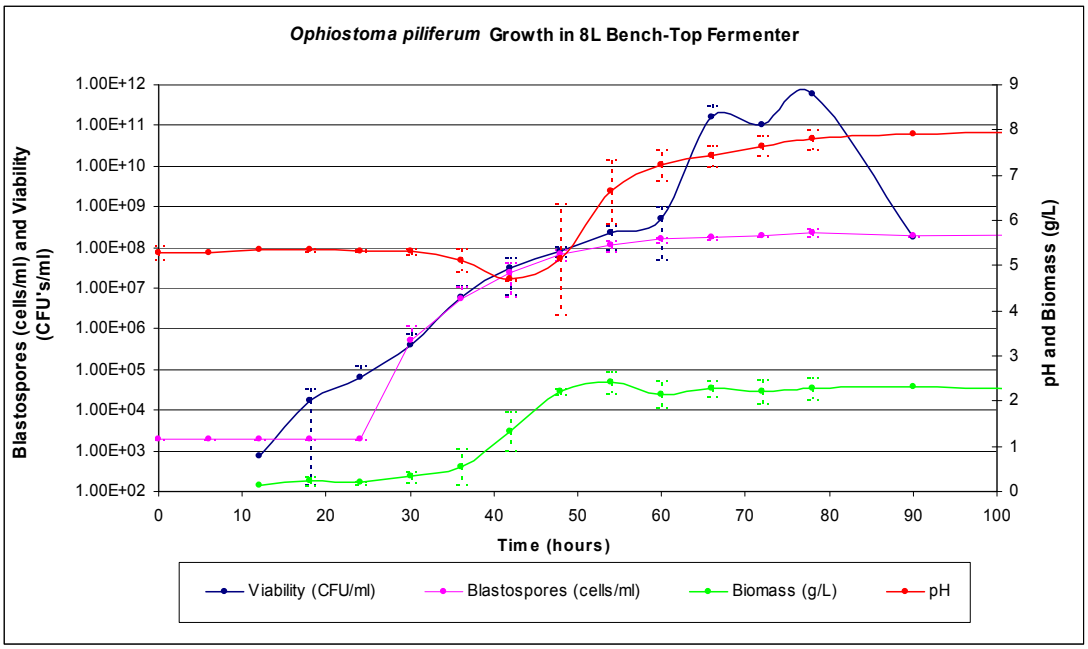


Figure 3.2.4A *O. piliferum* 97 growth in a 8L working volume bench top bioreactor.

Growth in terms of blastospore production, viability and biomass was compared for the three conditions (Figures 3.2.4B, 3.2.4C and 3.2.4D, respectively).

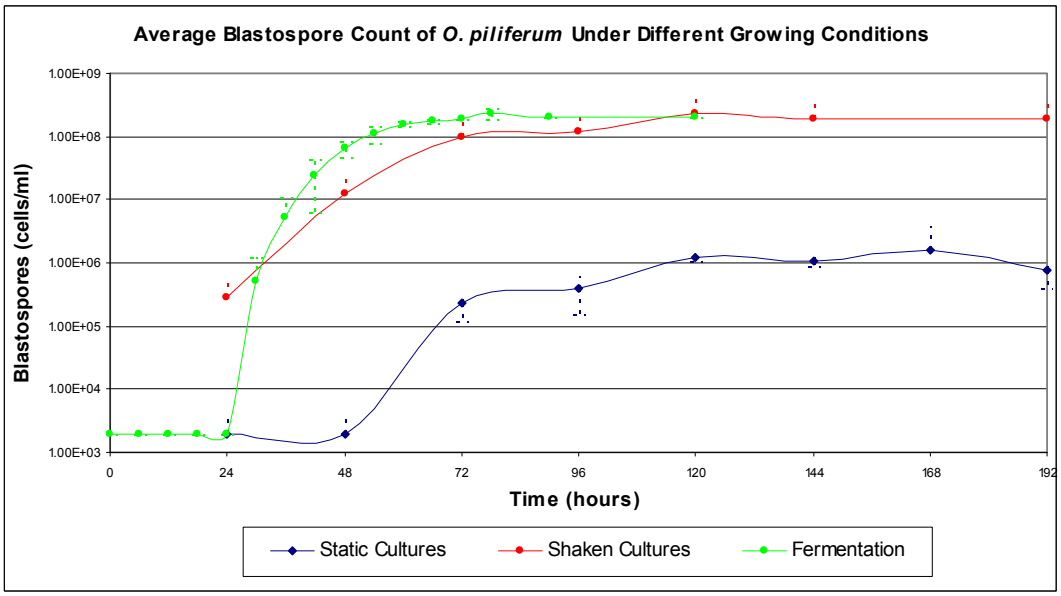


Figure 3.2.4B Comparisons of the average blastospore count measured in cells/ml of *O. piliferum* strain 97 over time under conditions 1, 2 and 3 for growth.

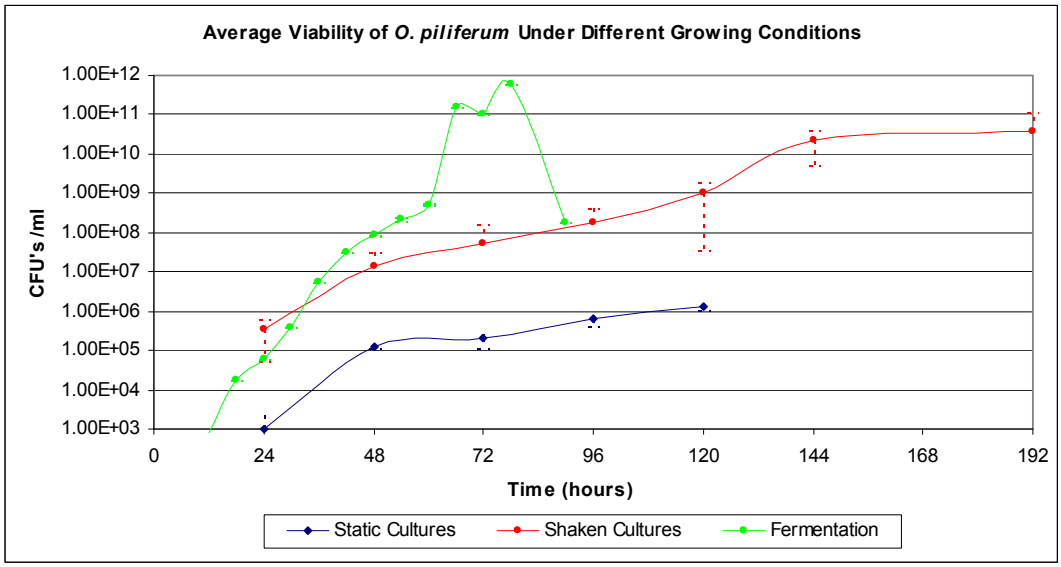


Figure 3.2.4C Comparisons of average viability measured in CFUs of *O. piliferum* strain 97 over time under conditions 1, 2 and 3 for growth.

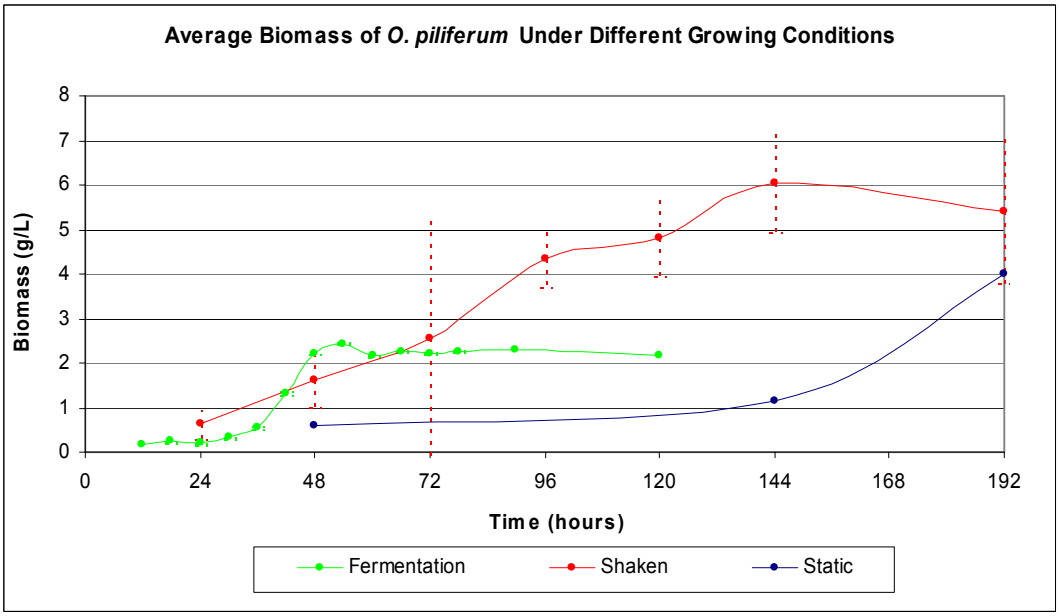


Figure 3.2.4D Comparison of the average biomass in mg/ml of *O. piliferum* strain 97 over time under conditions 1, 2 and 3 for growth.

From the growth data it was observed that cultures grown under stationary growth (condition 3) had lower levels of viability and less cell density compared to conditions 1 and 2. The highest level of biomass (6.1 g/L) was measured at 144 hours in shaken cultures (condition 1). The total levels of blastospores produced were comparable between stirred tank fermentation and shaken growth with between 1 x 10^8 and 5 x 10^8 cells per ml.

The blastospore doubling rate during exponential growth was calculated for the three conditions.

Condition 1 = 4.06 hours

Condition 2 = 2.11 hours

Condition 3 = 3.52 hours

Growth under stirred tank fermentation conditions produced the highest rate of blastospore production with a blastospore count doubling time of 2.11 hours during exponential growth.

3.3 Comparisons Between *O. floccosum* Strain J2122 and *O. piliferum* Strain 97 Growth

Overall, growth of *O. floccosum* strain J2122 and *O. piliferum* strain 97 was comparable in terms of the three factors measured. Microscopic analysis of growth under condition 1 for the two strains showed that both species were morphologically similar and shared many time-dependent characteristics of growth. Each stage of growth appeared particularly evident in terms of blastospore size and degree of budding. In early growth, blastospores were typically elongated and mycelial fragments had elongated apical ends, likely due to their active growth. By contrast mid phase blastospores were typically much rounder and more yeast-like in appearance, they were actively budding, often producing multiple buds. By late phase blastospore budding was much reduced and blastospores appeared much smaller. During late phase growth mycelial fragments were generally much wider and typically contained buds at the apical ends.

In conclusion it would appear that stirred tank fermentation was a much faster and more efficient method of accumulating cell density as one would suspect. What was surprising was the extent to which levels of total biomass differed between stirred tank fermentation and shaking methodologies for the two species investigated. Biomass accumulation is influenced by many factors including mass transfer, oxygenation, pH and temperature. The differences in biomass observed between stirred tank fermentation and benchtop shaken cultures in both species was neither reflective of the total blastospore levels achieved, as these were relatively proportional between the

stirred tank fermentation and shaken cultures, or the viability levels between the conditions as the level of viable colony forming units was greater in *O. piliferum* stirred tank fermentation cultures (1.7×10^{11} CFUs compared to 2.36×10^{9} CFUs in *O. floccosum*). Therefore, the differences in biomass were not due to increased filamentous or yeast like growth. Exopolysaccharide production has previously been reported in *O. novo-ulmi* and *O. ulmi* (Jeng *et al.*, 2007). Exopolysaccharides are typically high-molecular-weight polymers composed of sugar residues that are secreted by a micro-organism, including both yeasts and filamentous fungi, into the surrounding environment. Polysaccharide produced by *O. novo-ulmi* and *O. ulmi* was hypothesised to have a role in pathogenesis, likely because the presence of polysaccharide sheaths on hyphae and spores has been demonstrated in a large number of plant pathogenic fungi (Viret *et al.*, 1994). The production of exopolysaccharides by *O. floccosum* would likely explain the significant level of biomass observed as levels of approximately 6.9 g/L have been reported for some fungi (Reeslev *et al.*, 1997). Further evidence to support this explanation includes:

1) The timing of peak biomass production in *O. floccosum* and *O. piliferum*. The production of exopolysaccharides was shown to peak in early to mid phase growth in *O. novo-ulmi* and *O. ulmi* and many other fungi examined (Reeslev *et al.*, 1997; Jeng *et al.*, 2007). Exopolyssacharides have also been demonstrated to be reduced due to autolysis as a culture ages.

2) The differences observed in pH level between *O. piliferum* and *O. floccosum* stirred tank fermenter cultures. Different exopolysaccharides produced by fungi have different properties. For instance in some fungi the production of exopolysaccharides may reduce the pH level of the medium (Reeslev *et al.*, 1997)

3) *O. floccosum* cultures were typically more viscous compared to *O. piliferum* despite the appearance of fewer mycelia.

Within *O. floccosum* and *O. piliferum* cultures, growth was not exclusive to one somatic form at a time. In liquid culture both blastospores and mycelia were always present, but the proportional amounts of these forms varied throughout growth and between the conditions.

The blastospore doubling rate calculated during exponential growth indicated that a much higher rate of blastospore production occurred during growth in a stirred tank bioreactor for both species compared to shaken and static cultures. Blastospore doubling times were relatively comparable between the two species. The doubling times calculated for static growth were higher in both species than under shaking conditions despite cell density and viability being lower compared to shaken and fermenter growth. This is likely because the period of exponential growth for static cultures was over a much shorter period of time compared to shaken cultures. Cell doubling times during exponential growth for *S. cerevisiae* were typically reported at being between 90 and 140 min in nutrient rich media (Wheals *et al.*, 1995) indicating that both *O. piliferum* and *O. floccosum* are capable of relatively rapid blastospore production during stirred tank fermentative growth.

It is important to note that the growth of *O. floccosum* strain J2122 and *O. piliferum* strain 97 did not represent optimised growth as conditions of pH as well as media type and concentration were not optimised. However, previous research regarding temperature did identify 25 ºC as the optimal temperature for many strains of *Ophiostoma* species (Shirp, 2001), but not for all species, e.g. 28 °C for *O. ulmi* (Brasier *et al.*, 1981) 22 ºC for *O. piceae* (Brasier and Stephens, 1993) and *O. arduennense* (Calier *et al.*, 2006). Therefore, no comparisons have been made between *O. piliferum* and *O. floccosum* growth both within the Ophiostomataceae family and with other Ascomycete fungi.

4 Transcriptional Profiling

4.1 Introduction

At the commencement of this PhD thesis research, August 2004, little was known about *Ophiostoma floccosum*. Although *Ophiostoma piliferum* was first described in 1822 by Fries, little investigation had been made into its genetic composition, and there was little comparative molecular research done among species members of the Ophiostomataceae family. In August 2004, 34 entries representing *Ophiostoma sp*. ESTs were deposited in NCBI dbEST.

Understanding the mechanisms of transcriptional regulation is critical to understanding molecular functioning in an organism and protein expression. Advances in bioinformatics and computational tools have enabled the identification of many genes using high throughput sequencing EST analysis. Given the global impact of some members of the genus with regard to forest health and timber production and the potential for some members of the genus to be of biotechnological significance (as biocontrol agents and expression hosts), there has been a significant drive in the past 4 years to further understand the molecular biology of this genus. As of May 2008, there were 17,500 entries for *Ophiostoma* species in the NCBI dbEST including 34 entries for *O. piceae*, 9589 for *O. piliferum,* 5974 for *O. calvigerum* (*Grosmannia calvigera)* and 1277 associated with *O. novo ulmi* and *Americanas* callus production as a result of *O. novo ulmi* infection. No EST sequences for *Ophiostoma floccosum* species were deposited in the NCBI dbEST at this time.

Analysis of growth in early phase blastospores versus other phases or morphological forms was two fold in reasoning. From a biotechnological point of view, it was advantageous to understand gene expression and to identify up-regulated transcripts that could be indicative of a strong promoter at an early stage in a fermentable system. Also, EST datasets produced from other *Ophiostoma* species were generally produced at later stages of growth, often on solid media and typically isolated from or including

the mycelial form. Based on the information available regarding the current *Ophiostoma* EST libraries, no one had investigated early stage growth.

4.1.1 Specific Aims and Objectives for Transcriptional Profiling

There were two specific objectives for the Transcriptional Profiling chapter of this PhD thesis research, as follows:

1) To profile mRNA transcripts using EST methods in *O. floccosum* early log phase blastospore development. Analysis of the transcripts included:

- The determination of unique putative transcript (UPT) frequency.
- The identification of open reading frames (ORFs).
- Prediction and annotation of putative gene function.
- Analysis of predictive codon frequency for protein expression.

2) Compare and contrast *O. floccosum* transcript profiles with EST data from other *Ophiostoma* species.

4.2 Profile of *Ophiostoma floccosum* Transcription Using EST Sequence Data

4.2.1 cDNA Library Construction and EST Generation

A cDNA library from *O. floccosum* blastospores and small mycelial fragments cultivated for 48 hours post inoculation was used to generate ESTs.

RNA was isolated from blastospores which had been incubating for 48 hours in YM medium shaken at 180rpm as described in Sections 2.1.7 and 2.2.2. This represented early stage fermentation-type growth. mRNA was then further purified from the RNA according to the method described in Section 2.2.3. cDNA was constructed from the mRNA using OrientExpress™ cDNA system (Novagen). cDNA synthesis was primed using oligo (dT) primers. The cDNA was then cloned using OrientExpress ™ (Novagen) phage vector system. Clones containing the cDNA inserts were sequenced (Macrogen, Korea). A total of 1165 48Sp clones were sequenced in both the forward

and reverse directions. From this library, 1527 ESTs were generated with an average read length of 671 bp.

4.2.2 Cluster Formation and EST Sequence Analysis

The 1,527 EST raw sequences were edited to remove vector and poor quality sequence as described in Section 2.5.6 resulting in a subset of 1,207 EST fragments, with an average read length of 713 bp.

The ESTs from OR48 were clustered into consensus sequences based on sequence similarity in order to determine how many unique putative transcripts (UPTs) were represented by the EST dataset using Vector NTI sequence assembly software. A consensus sequence (cluster) containing 2 or more EST fragments was termed a contig. A cluster containing only a single EST fragment was termed a singlet. A total of 225 contigs and 372 singlet sequences were determined. Open reading frames (ORFs) were predicted using the following two methods, as described in Section 2.5.6.3:

1) OrfPredictor (Fungal Genomics Project, Concordia University),
2) ORFinder (NCBI).

Using ORFinder methods of predicting ORFs, a total of 1,243 ORFs greater than 100bp in read length were predicted from OF48 UPTs. A summary of the OF48 EST dataset is given in Table 4.3.1.

UPTs were annotated based on homology to known protein listed in the Genbank Non-redundant database using TargetIdentifier (Fungal Genomics Project, Concordia University) and manually from NCBI BLASTx results. TargetIdentifier used the frames predicted in the pre-run BLASTX (NCBI BLASTall package) results to identify and annotate full-length EST-derived sequences according to the predicted ORF length and completeness. All 599 UPTs were analysed using TargetIdentifier software as described in Section 2.5.6.4. Of these, 386, representing 65% of the UPTs, did not have a highest score pair with an *E*-value greater than 1e-5 and were classed as having no hit. Independent of the TargetIdentifier results, BLASTx analysis identified 72% of these UPTs as being significantly homologous (an *E*-value greater than 1e-15)

to known or predicted protein sequences in the public databases. This level of homology to known proteins in the public databases is relatively standard if not high compared to the alignment of other Pezizomycotina ESTs annotated. Within the *O. calvigerum* EST dataset, approximately 67% of the UPTs identified were matched by BLASTx analysis to known or predicted sequences (DiGuistini *et al.*, 2007). In *Trichoderma* species, BLASTx analysis of 3,478 UPTs identified in *T. harzianum* ESTs resulted in the annotation of 65% of UPTs whereas in *T. reesei* putative function could only be assigned to 36% of 1151 UPTs (Chambergo, *et al.*, 2002; Suárez *et al.*, 2007).

A summary of *E*-values for all significant matches using TargetIdentifier methods of annotation is presented in Figure 4.2.2. Of the 213 UPTs with significant hits, 97 were considered full length, 15 were short full length, 4 were categorized as ambiguous and 97 were determined to be partial. A complete list of all significant *O. floccosum* UPTs annotated using TargetIdentifier is presented in Appendix 1.

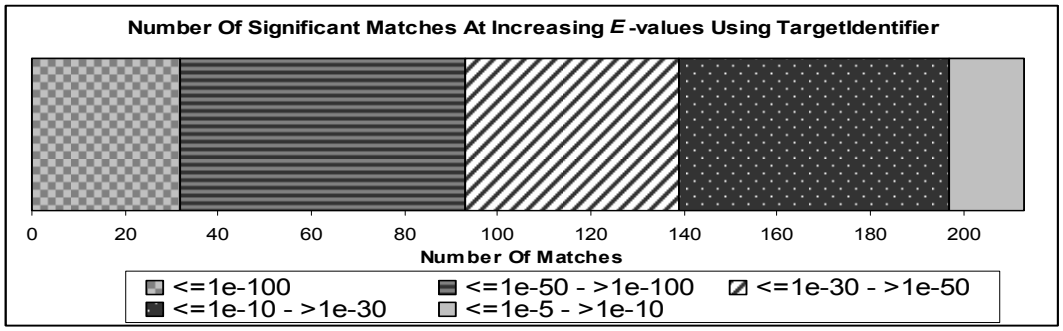


Figure 4.2.2 The number of matches at increasing *E*-values for BLASTx analysis of OF48 UPTs against Genbank non redundant database using TargetIdentifier methods of classification and annotation.

4.2.3 Transcript Abundance and Highly Represented Genes

The frequency of transcript was determined by calculating the number of EST fragments used to compose a contig. This data is given in Figure 4.2.3A.

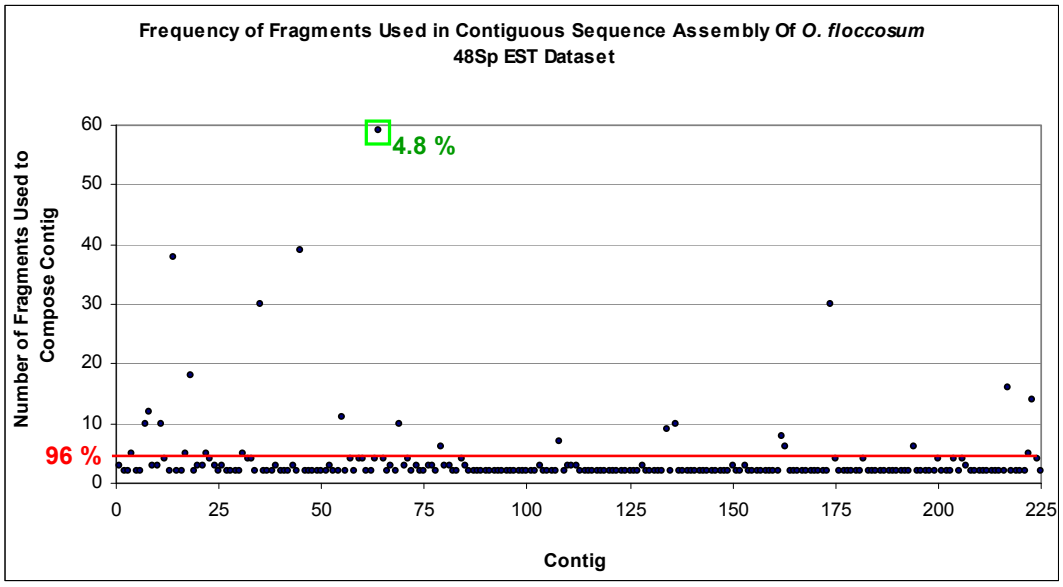


Figure 4.2.3A Graph showing the frequency of OF48 EST fragments used in alignments to form contigs.

A fragment was considered up-regulated if the total number of fragments used in the alignment was greater than 1% of the total EST dataset. Only 2% of UPTs identified coded for up-regulated transcripts. The contig containing the highest frequency (OF0064), with 59 fragments, (highlighted in a green box in Figure 4.3.3A), represented 4.8% of the total OF48 EST dataset. 96% of UPTs contained >5 EST fragments.

Assessing the completeness of the OF48 dataset could only be achieved using theoretical methods. Previously, the completeness of the *O. calvigerum* EST dataset was assessed based on the number of genes present from the 1,8-dihydroxynaphthalene-melanin biosynthetic pathway (DiGuistini *et al.*, 2007). However, the yeast form of *Ophiostoma* does not produce melanin (Dogra and Breuil, 2004) and therefore this method, using expression of melanin biosynthetic pathway genes, was not suitable to assess the completeness of the OP48 EST dataset. Therefore, assuming 9,000 – 14,000 total ORFs in Pezizomycotina genomes, as presented in the literature (Bruce and Birren, 2003; Mannhaupt *et al.*, 2003; Martinez *et al.*, 2004; Arvas, 2007; Cuomo *et al.*, 2007; Pel *et al.*, 2007), the OF48 clusters identified were likely to represent approximately 4.3 – 6.6% of all *O. floccosum* genes.

Those UPTs identified that contained 5 or more aligned fragments were analysed using BLASTx independent of TargetIdentifier findings, and summarized in Table 4.2.3 with the UPTs listed in order of the greatest frequency. BLASTx results with an *E*-value of <1e-15 were not considered to be significantly homologous and were described as having no significant similarity.

			Non-redundant Genbank			
Cluster ID	**Length**	**Frequency**	**Best identity descriptor**	**Accession**	**Score**	***E*-value**
OF480064	692	59	NADH dehydrogenase subunit 4 [Verticillium dahliae].	YP_667826	270	2.00E-71
OF480045	851	39	cytochrome oxidase subunit I [Neurospora crassa]	CAA32799	360	3.00E-98
OF480014	1293	38	NADH dehydrogenase subunit 5 [Phaeosphaeria nodorum SN15].	YP_001427401	657	0
OF480035	950	30	NADH dehydrogenase subunit 5 [Neurospora crassa]	AAT74903	125	2.00E-27
OF480174	1003	30	putative DNA recombination-associated ATPase RarA [Escherichia albertii TW07627]	ZP_02903747	381	1.00E-104
OF480018	994	18	ATP synthase F0 subunit 9 [Verticillium dahliae].	YP_667831	90.5	4.00E-17
OF480217	525	16	hypothetical protein AN5245.2 [Aspergillus nidulans FGSC A4].	XP_662849	84	1.00E-18
OF480223	662	14	DNA-directed RNA polymerase, beta' subunit [Escherichia albertii TW07627]	ZP_02904574	125	9.00E-28
OF480008	502	12	No significant similarity	NA		
OF480055	518	11	No significant similarity	NA		
OF480007	523	10	unnamed protein product [Podospora anserina], possible glucose-repressible protein	XP_001912006	96.7	6.00E-19
OF480011	934	10	No significant similarity	NA		
OF480069	407	10	No significant similarity	NA		
OF480136	649	10	orf301 - mitochondrion protein [Podospora anserina], possible endonuclease	NP_074923	107	3.00E-22
OF480134	2804	9	COG3696: Putative silver efflux pump [Escherichia coli E110019].	ZP_00720644	501	1.00E-149
OF480162	582	8	formate transporter [Shigella flexneri 5 str. 8401]	YP_688432	239	1.00E-61
OF480108	2069	7	LAGLIDADG endonuclease [Gibberella zeae]. - mitochondrion protein	YP_001249331	506	6.00E-142
OF480079	872	6	orf301 [Podospora anserina] possible GIY-YIG endonuclease	NP_074923	242	5.00E-85
OF480163	1672	6	predicted peptide transporter subunit: periplasmic-binding component of ABC superfamily [Escherichia coli APEC O1]	YP_851922	496	6.00E-139
OF480194	409	6	no significant match			
OF480004	835	5	40s ribosomal protein S5 [Chaetomium globosum CBS 148.51].	XP_001228475	352	2.00E-95
OF480017	1826	5	translation elongation factor 1 alpha [Metarhizium anisopliae].	AAR16425	813	0
OF480022	199	5	No significant similarity	AAK67169	38.9	0.43
OF480031	638	5	hypothetical protein AN5245.2 [Aspergillus nidulans FGSC A4].	XP_662849	84	9.00E-24
OF480222	313	5	COG1501: Alpha-glucosidases, family 31 of glycosyl hydrolases [Escherichia coli B171]	ZP_00710871	148	5.00E-34

Table 4.2.3 'Best Match' of OF48 contigs containing =< 5 EST fragments determined by BLASTx analysis against Genbanks Non-redundant database. Hits with an *E*-value of < 1e-15 were determined to have no significant match.

A third method, independent of BLASTx and TargetIdentifier homology searching was also used to putatively predict the cellular role of ORFs identified in OF48. The most likely ORF for each UPT was determined using OrFinder (Fungal Genomics Project, Concordia University, Montreal, Canada) and submitted for on-line analysis using the ProtFun 2.2 server (CBS). This program produced *ab initio* predictions of protein function from sequence by querying a large number of other feature prediction servers to obtain information on various post-translational and localisational aspects of the protein. These predictions were then integrated into final predictions of the cellular role. The ProtFun results (Figure 4.2.3B) were divided into three categories, metabolism, protein fate and unclassified, based on the following criteria:

Metabolism – Transcripts with a predicted gene function involved with one of the following roles: cell cycle and DNA processing, protein synthesis, cell type differentiation, energy and transcription.

Protein fate - Transcripts with a predicted gene function involved with one of the following roles: cellular communication, signal transduction, regulation, interaction with cellular environment, systematic regulation, interaction with the environment, control of cellular organization, subcellular organization and proteins with binding functions.

Unclassified – anything that did not fit into one of the above categories or was unable to be classified using ProtFun 2.2.

Based on these criteria it was apparent that the majority of proteins produced in early phase yeast cells were involved in metabolic processes within the cell, which are underlined in red in Figure 4.2.3B.

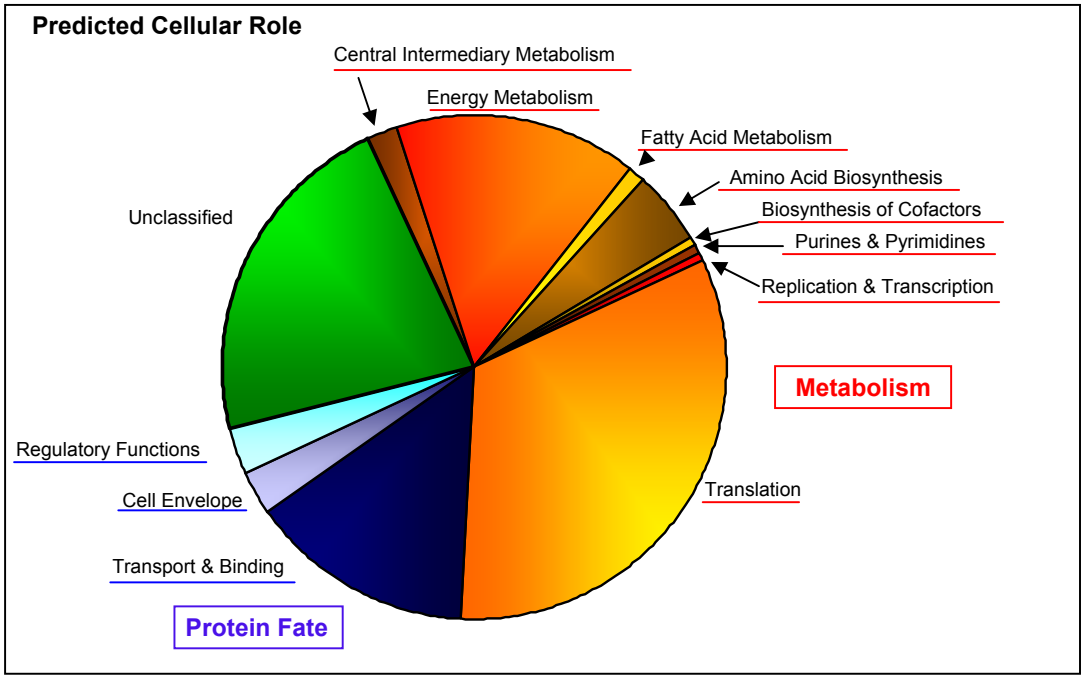


Figure 4.2.3B Predicted cellular role of OrFinder predicted genes within OP48 dataset. Predictions were made independently of BLASTx analysis using ProtFun *ab initio* methods.

Sequence homology does not automatically imply functional similarity. However, it is essentially the only practical existing method that allows large scale annotation of novel genes. Annotation based on sequence homology of the OF48 EST dataset indicated that the majority of transcripts putatively identified were involved with rapid cell proliferation and carbohydrate metabolism .

Within the OF48 dataset genes were identified that were involved in glycolysis, the tricarboxylic acid (TCA) cycle and pyruvate oxidation. The presence of transcripts putatively identified as being involved in the TCA cycle indicated that respiration is likely to be favored in *O. floccosum* in the blastospore form of growth. In the yeast *S. cerevisiae* which prefers anaerobic fermentation, high concentrations of glucose, as would have been present in the YM media used to cultivate *O. floccosum*, would repress genes encoding enzymes of the TCA cycle (Chambergo *et al.*, 2002). Up-regulation was observed in a number of genes encoding mitochondrial proteins in OF48. Aerobic metabolism requires the expression of protein involved in mitochondrial activity and the flow of electrons and protons through the complex of respiratory chain proteins that are encoded by mitochondrial and nuclear genes (Chambergo *et al.*, 2002). The high numbers of transcripts homologous to cytochrome *c* oxidase and NADH-ubiquinone oxidoreductase subunits further supported the evidence that the glucose-rich medium did not repress mitochondrial gene expression in *O. floccosum* early stage blastospores.

Few genes were identified by this PhD thesis transcriptomic analysis as involved in biomass degradation despite the identification of hydrolases previously characterised in the extracellular supernatant in *O. floccosum* (Schirp *et al*., 2003b; Wu *et al.*, 2006). This may be because there are ample nutrients available in the media employed in this thesis research or that growth was predominantly in the yeast form. Genome analysis of Saccharomycotina has shown that they produce fewer genes involved in biomass degradation such as glycoside hydrolases compared to Pezizomycotina (this is discussed in Section 1.2.3.4).

One unexpected result was the high number of UPTs that aligned most significantly to reported bacterial sequences. As has been previously mentioned, this may be due to a

lack of fungal sequences in the databases or it may be due to bacterial contamination. Addressing the latter, though, during construction of the *O. floccosum* OF48 cDNA library, it was probed using PCR amplification of universal 16S primers to identify any potential bacterial contamination. This did not result in any amplification although a positive control containing bacteria did show amplification. These experimental results indicate that it was unlikely there was bacterial contamination in the cells from which the OF48 cDNA library was made.

4.3 Comparison Among EST Data From Different *Ophiostoma* Species

OF48 EST data was compared with EST datasets from four other *Ophiostoma* species including *O. piliferum, O. novo-ulmi, O. clavigerum* and *O. piceae*. A description of these datasets is given in Table 4.3.1

4.3.1 cDNA Libraries and EST Generation

Dataset OF48 was made as part of this PhD thesis research. The HN dataset, prepared from *O. novo-ulmi* strain H327, was kindly obtained from Professor W. E. Hintz of University of Victoria, Canada. All other datasets were obtained from publicly available databases of NCBI Genbank. The data sets are summarised in Table 4.3.1.

Species	Library name	Strain	Tissue type	Description	Submitter	ESTs	Clusters	Contigs	Singlets
O. floccosum	OF48	J2122	Yeast	Grown for 48 hours under shaking conditions, filtered to remove large fragments	Farrell, R. (not published)	1207	598	225	373
O. piliferum	OP	unknown		Cultured under different carbon conditions	Tsang, A.	9589	3500	1482	2077
O. clavigera (G. clavigera)	CV	SLKw1407	Mycelium	mycelium from 4 different types of solid media	Breuil, C.	5974	2444	1164	1280
O. novo-ulmi	UA	H327	Mycelium	Onu per: supressive subtractive cDNA library of perithecia (sexual fruiting body) against driver (mycleium grown at 22-24 ºC)	Bernier, L.	128			
				Onu-syn: supressive subtractive cDNA library of synnemata (asexual fruiting body) aginst driver population (mycelium grown at 22-24 ºC)		181			
				Onu-t15: supressive substractive cDNA library with driver population (mycleium grown at 22-24 ºC)		156			
				Onu-t31: supressive substractive cDNA library with driver population (mycleium grown at 22-24 ºC)		96			
						561	430	72	357
O. novo-ulmi	HN	unknown	Yeast		Hintz, W. (not published)	4554			
			Mycelium			71			
						4625	2540	617	1923
O. piceae	OPC	unknown	Yeast	Grown under shaking conditions in liquid complete medium for 48 hours at 23 ºC (250 rpm) and filtered to remove large fragments	Breuil, C.	22			
			Mycelium	Grown for 48 hours on complete medium		12			
						34	19	5	14

Table 4.3.1 Summary of *Ophiostoma* EST datasets used in this PhD thesis research.

Each of the datasets was produced from cDNA libraries constructed from cultures grown on a variety of carbon sources in both liquid and solid medium. Within each dataset, a variety of cDNA libraries were used. It was not always clear as to what the conditions for growth were and the methodologies used for EST production particularly EST data obtained from public databases. Most of the cDNA libraries were likely constructed from *Ophiostoma* mycelia or a mixture of mycelia and blastospores in mid to late phase growth. Only two of the subsets, HN-yeast and OPC-yeast, were constructed from isolated blastospores.

4.3.2 Cluster Formation

ESTs contained within individual datasets were clustered and assembled into consensus sequences based on sequence similarity using Vector NTI software (Invitrogen). A summary of the number of contigs and singlets identified within each of these datasets is presented in Table 4.3.1. In addition to the results summarized in Table 4.3.1, UA and HN datasets, both containing EST sequence data from *O. novo ulmi*, were aligned to determine UPTs. From this combined *O. novo ulmi* dataset, (named ANU), a total of 5,186 EST fragments, 2,196 singlets and 696 contigs containing 2 or more fragments were identified.

4.3.3 Transcript Abundance And Highly Represented Genes

UPTs within individual datasets containing theorized up-regulated transcripts were determined by calculating the number of EST fragments used to compose a contig. For analysis of these datasets, this method was not assumed to calculate frequency as the methods for cDNA construction, sequence quality and editing were not known and many of the sequences represented normalized libraries constructed from strains grown under different conditions in terms of media, time, temperature etc. Therefore, any frequency calculations would be speculative as the accuracy of such predictions could not be determined.

Contigs within each dataset containing a number of EST fragments representing more than approximately 0.4% of the total dataset were identified and are summarized in Appendix 2.

4.3.4 Comparison Of Transcripts Expressed in *Ophiostoma* EST Datasets

All UPTs identified within individual EST datasets were aligned based on sequence homology to identify UPTs coding for genes common between the *Ophiostoma* species. Of the 9,531 individual UPTs aligned, 484 contigs containing 2 or more UPTs were identified. Of these 484 contigs 137 were alignments of UPTs from within the same dataset and therefore these 137 were removed from the overall analysis. A Venn diagram representing the number of contigs identified with UPTs in common between the datasets (not including the 137 alignments from the same dataset) is presented in Figure 4.4.4A.

For analysis, the dataset ANU representing all *O. novo ulmi* EST sequences from HN and NU datasets was used. No contigs were identified containing UPTs from all 5 of the datasets. 6 contigs were identified containing UPTs from the four main datasets OF48, OP, ANU and CV. These contigs were described as MX11, MX72, MX84, MX215, MX422 and MX433 and are listed in Table 4.3.4A. Three of these contigs, MX11, MX72 and MX215, contained putative genes functioning in core metabolism with MX11 and MX72 involved in glycolysis and MX215 identified as a transcript coding for a 60S Ribosomal protein. The remaining three, MX84, MX422 and MX433, were predicted based on sequence homology to have a role in protein fate with the transcripts identified as a putative superoxide dismutase, heat shock protein and a structural alpha-B chain tubulin gene, respectively. MX422, the putative heat shock protein, was up-regulated in the OP dataset, containing 72 EST fragments with homology to only 1 fragment in each of the OF48, CV and ANU datasets. MX72 identified as enolase, an enzyme functioning in the conversion of glucose to pyruvate in the glycolytic pathway, was found to be reasonably up-regulated in OF48 (as 7 fragments were identified), OP (38 fragments) and ANU (8 fragments) with only 2 fragments having homology from the CV dataset.

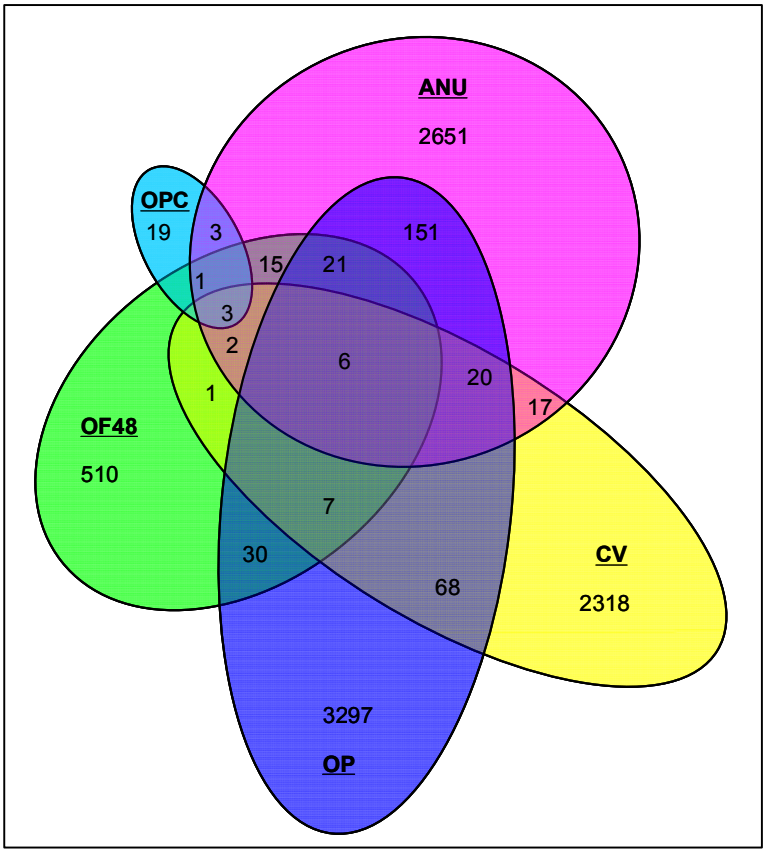


Figure 4.3.4A Venn diagram representing the number of contigs identified with UPTs in common between the datasets.

A Venn diagram representing the number of EST fragments homologous between the datasets as an indication of frequency is given in Figure 4.3.4B. Approximately 13% of all *Ophiostoma* EST sequences analysed in this research had homology to at least one UPT identified within EST data from another *Ophiostoma* species. Interestingly, up-regulated transcript identified in the datasets did not align more readily with UPTs in other datasets.

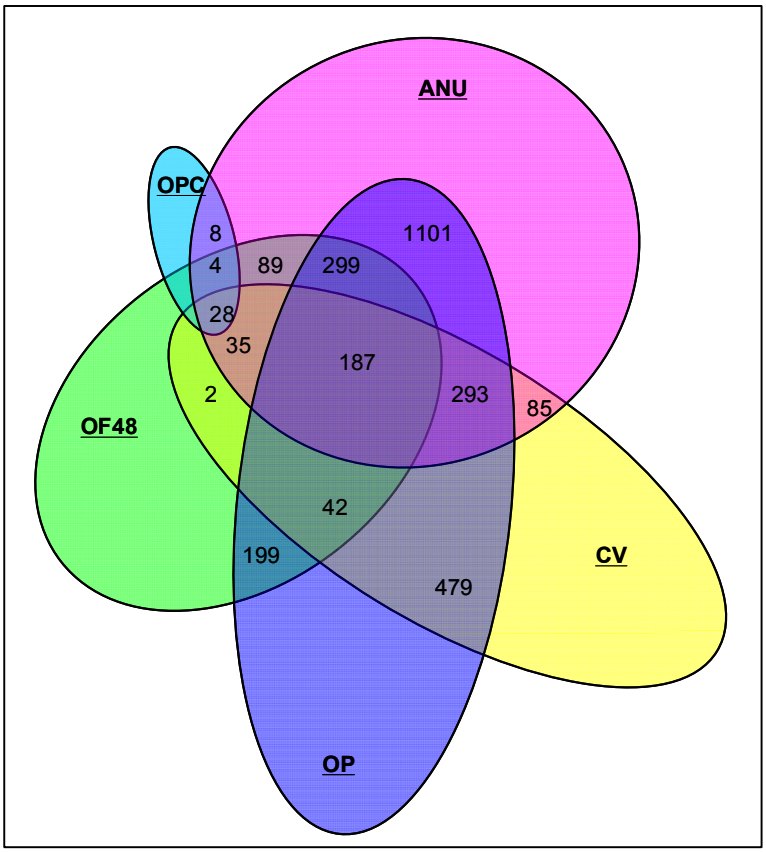
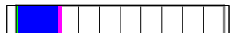
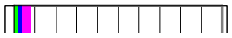
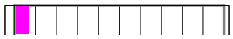
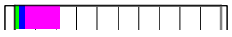
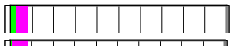
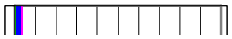
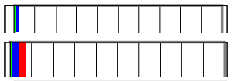
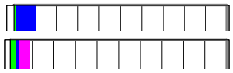
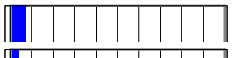
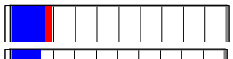


Figure 4.3.4B Venn diagram representing the number of EST fragments included in homologous UPTs between datasets.

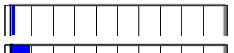
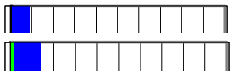
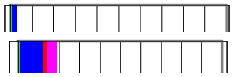
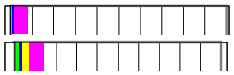
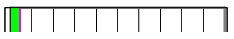
All contigs containing OF48 UPTs were identified and summarised in Table 4.3.4. The number of individual EST fragments within each UPT from different datasets was calculated and graphed beside each entry in the table to indicate the degree of commonality between the datasets. In the graphs contained within Table 4.3.4, each bar represented 10 EST fragments.

METABOLISM

	ID.	BP	Putative function	Accession	Score	*E*-value
	MX10	726	40S ribosomal protein	XP_369965	255	2.00E-66
	MX106	1764	40S ribosomal protein S1-like protein	AAX07667	415	5.00E-114
	MX141	937	40S ribosomal protein S9	XP_001275347	198	9.00E-49
	MX159	906	40S ribosomal protein S22	XP_369964	246	2.00E-63
	MX192	893	40S ribosomal protein S4-A-like protein	AAW69345	205	6.00E-51
	MX435	709	40S ribosomal protein S25	XP_001228474	142	3.00E-32
	MX407	1345	40s ribosomal protein S5	XP_001228475	392	2.00E-107
	MX27	1302	60S ribosomal protein L15	P_957199	359	1.00E-97
	MX83	1112	60S ribosomal protein L5	XP_960578	387	1.00E-105
	MX114	1065	60S ribosomal protein L32	Q7RXY1	194	1.00E-47
	MX147	1099	60S ribosomal protein L11	XP_001267248	233	8.00E-68
	MX215	1142	60S ribosomal protein L8	XP_001273971	438	4.00E-121
	MX271	752	60S ribosomal protein L44	XP_663785	205	6.00E-51
	MX421	1410	60S acidic ribosomal protein P0	XP_001220197	434	6.00E-120
	MX440	707	60S ribosomal protein L34	XP_359481	195	4.00E-48
	MX117	1803	Ribosomal L18ae protein family	XP_001269707	273	2.00E-71
	MX138	1037	ribosomal protein L19	XP_001404186	212	1.00E-53
	MX450	839	ribosomal protein 3	AAY59060	116	2.00E-24
	MX438	859	RNP domain protein	EDP49798	196	2.00E-48
	MX62	2178	eukaryotic initiation factor	XP_001220184	535	0
	MX71	890	eukaryotic translation initiation factor	XP_001245555	275	3.00E-72
	MX89	1108	eukaryotic translation elongation factor 1-β	XP_959625	192	8.00E-61
	MX423	1167	protein translation factor sui1	XP_964345	176	2.00E-42
	MX199	1910	Elongation factor 1-alpha	P34825	798	0
	MX280	701	Elongation factor 2	Q96X45	143	2.00E-43
	MX15	3017	elongation factor 2	XP_962286	1457	0
	MX2	837	CHCH domain protein	XP_001263816	108	6.00E-22
	MX9	1745	cytochrome c oxidase assembly cox15	XP_750658	484	3.00E-135
	MX434	648	ubiquinol-cytochrome-C oxidoreductase	XP_960814	141	6.00E-32
	MX22	923	NADH-quinone oxidoreductase	XP_001269854	268	3.00E-70
	MX24	1177	possible NADH-cytochrome b5 reductase	XP_001549033	275	3.00E-78
	MX415	1299	NADH dehydrogenase subunit 2	AAX21824	254	5.00E-66
	MX25	1546	Phosphoglycerate kinase	P14228	587	4.00E-166
	MX70	1587	fructose-1,6-bisphosphatase	XP_960423	575	1.00E-162
	MX42	1976	6-phosphogluconate dehydrogenase	XP_964959	864	0
	MX60	2819	glycosyl transferase	XP_001259116	952	0
	MX72	2068	enolase	XP_957959	757	0
	MX11	1484	fructose-bisphosphate aldolase	XP_958393	484	2.00E-135
	MX14	1532	alcohol dehydrogenase I	XP001228112	362	2.00E-140
	MX441	1574	possible aldehyde dehydrogenase family	XP_001275849	484	4.00E-135
	MX38	1590	pyruvate dehydrogenase E1 component α	XP_957122 630	630	3.00E-179

	ID.	BP	Putative function	Accession	Score	*E*-value
	MX49	2581	ATP-citrate synthase subunit 1	XP_001228353	811	0
	MX64	3272	fatty acid synthase beta subunit dehydratase	XP_962465	3219	0
	MX67	1281	glutamyl-tRNA synthetase	XP_958832	608	2.00E-172
	MX73	1396	3-ketoacyl-CoA ketothiolase	XP_752635	276	3.00E-141
	MX74	1703	Glutamine synthetase	AAB00322	651	0
	MX102	945	vacuolar ATP synthase subunit c	XP_001269180	315	5.00E-84
	MX419	1256	3-hydroxybutyryl-CoA dehydrogenase	XP_001274777	404	5.00E-11
	MX436	1202	possible Vacuolar ATP synthase subunit H	XP_752433	198	4.00E-49
	MX409	2298	inosine-5'-monophosphate dehydrogenase	XP_964976	896	0

PROTEIN FATE

	ID.	BP	Putative function	Accession	Score	*E*-value
	MX389	1004	histone H4	XP_368084	161	6.00E-38
	MX413	1052	histone H2A	XP_959442	191	5.00E-47
	MX427	828	histone H2B	XP_959440	213	1.00E-53
	MX91	1263	nucleotide binding protein	XP_001268037	221	5.00E-69
	MX3	448	CipC-like antibiotic response protein	XP_753706	95.9	3.00E-18
	MX36	1102	rapamycin binding protein FKBP12	CAK54362	137	3.00E-34
	MX43	912	ribonucleoprotein complex protein Nhp2,	XP_001268897	104	8.00E-21
	MX84	1378	Superoxide dismutase	AAN75577	285	8.00E-75
	MX164	997	Translationally controlled tumor protein	XP_001385879	191	1.00E-46
	MX178	1471	polyubiquitin	AAK19308	323	1.00E-127
	MX183	939	cyclophilin	CAA35682	258	6.00E-67
	MX186	1107	Coatomer subunit alpha	XP_001270903	290	1.00E-76
	MX234	894	Patatin-like serine hydrolase	XP_746486	206	2.00E-73
	MX285	1147	outer mitochondrial membrane protein porin	XP_960950	404	1.00E-110
	MX406	1019	mitochondrial dicarboxylate carrier protein	XP_750662	258	6.00E-67
	MX410	878	metacaspase CasA	XP_001269487	147	9.00E-34
	MX418	1433	mitochondrial carrier- GTP/GDP transporter	XP_001225893	491	3.00E-137
	MX422	2469	heat shock protein 90	XP_961298	1008	0
	MX425	702	Nascent polypeptide complex subunit β	Q7SDU4	224	6.00E-57
	MX426	863	Guanine nucleotide-binding protein	BAD44728	220	3.00E-61
	MX428	874	proteasome regulatory particle subunit	XP_001270680	331	3.00E-89
	MX430	663	nuclear transport factor 2	XP_001558550	176	1.00E-42
	MX432	1027	proteasome component PRE6	XP_001543301	439	1.00E-121
	MX433	1781	tubulin alpha-B chain	XP_963223	855	0
	MX437	1343	serine/threonine protein phosphatase PP1	XP_001245465	524	1.00E-159
	MX455	1302	putative senescence-associated protein	BAB33421	249	2.00E-64

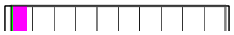
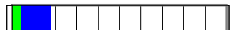
UNKNOWN						
	ID.	BP	Putative function	Accession	Score	*E*-value
	MX50	1292	unknown	NA		
	MX133	775	unknown	NA		
	MX330	1984	Unknown	NA		
	MX412	764	unknown	NA		
	MX420	716	no significant match	NA		
	MX431	666	unknown	NA		
	MX439	1130	unknown	NA		
	MX443	923	unknown	NA		
	MX473	1345	unknown	NA		
	MX12	1697	unknown	NA		

Table 4.3.4 All contigs identified from alignments between *Ophiostoma* datasets containing OF48 UPTs. Each bar within the graph represents 10 EST fragments.

Of the 86 *O. floccosum* UPTs identified as being homologous to one or more of the other *Ophiostoma* datasets, only 10, representing 11.6%, were of unknown function or did not align with any significant homology to sequences in the public databases. This indicated that many of the transcripts identified were conserved in other fungal species. Many of these putative genes were involved in protein production and core metabolism. Given the dominance of transcripts identified within OF48 that were predicted to have a cellular role in these two categories, it is not surprising that the findings were biased toward these two functional groups. Interestingly, greater homology between OF48 and the yeast specific EST subsets in *O. novo-ulmi* and *O. piceae* was not observed. As previously mentioned, a lack of homology with the *O. piceae* dataset could be due to the datasets small size. According to the results shown in Figure 4.3.4B, the numbers of ESTs common to the OF48 and ANU libraries represent 53 and 12% of the total ESTs in the libraries respectively. Figures for OF48-CV are only 41 and 8% respectively. Given that the total numbers of ANU and CV libraries are similar (5186 and 5974), this suggests a closer similarity between ANU and OF48 than to OF48 and CV. This is logical as both *O. floccosum* and *O. novo-ulmi* are both members of the *O. piceae* complex (Harrington *et al.*, 2001).

4.3.4.1 Identification of Putative Peptidases

Of particular interest in the analysis of the *Ophiostoma* datasets was the identification of genes putatively encoding peptidases (also termed proteases, proteinases and proteolytic enzymes). Peptidases are degradative enzymes that cleave proteins into smaller peptides and amino acids (North, 1982). They have a range of biological

functions including cell autolysis, nutrient cycling, activation of zymogens and post-translational processing (Hoffman and Breuil, 2002). Peptidases are widely produced in fungi (Hoffman and Breuil, 2002; James, 2006; Suárez, 2007; Wu *et al.*, 2007). As most of the nitrogen in wood is in an organic form, proteinases are essential for *Ophiostoma* species to access the nitrogen necessary for growth (Abraham *et al.*, 1996; Hoffman and Breuil, 2004; Wu *et al.*, 2006). While some may serve a useful application in industry, peptidase production can severely inhibit the production of recombinant protein in fungal hosts.

To identify putative peptidases within the datasets, UPTs were aligned against the *MEROPS* database as described in Section 2.5.6.5 using BLASTx alignment algorithms. The *MEROPS* database uses a hierarchical, structure-based classification of the peptidases where each peptidase is assigned to a Family on the basis of statistically significant similarities in amino acid sequence. Proteolytic enzymes are grouped into six families, aspartic peptidases, cysteine peptidases, glutamic peptidases, metallo peptidases, serine peptidases and threonine peptidases. Families that are thought to be homologous are then grouped together in a Clan. A total of 32 Eukaryotes, including 19 fungal species have sequences listed in the database.

The UPTs identified within each of the *Ophiostoma* EST datasets that are homologous to peptidase families are summarised in Table 4.3.4.1A. The number of EST fragments used to construct the UPTs is also included in the Table to give an indication of the frequency of the predicted peptidase transcripts. Relative frequency (Reltv. Freq) was calculated by determining the percentage of total fragments homologous to sequences in the MEROPS database. Proportionally, the *O. calvigerum* dataset contained the greatest number of predicted peptidases with 174 fragments, representing 2.9% of the total CV EST dataset. No putative peptidases were predicted in the OPC dataset.

Dataset	No. of UPTs and Fragments Homologous To The Different Peptidase Families															
	Aspartic		Cystine		Glutamic		Mettalo		Serine		Threonine		Tot. UPTs	% UPTs	Tot. Frag.	Reltv. Freq
	UPTs	Frag.	UPTs	Frag.	UPTs	Frag.	UPTs	Frag.	UPTs	Frag.	UPTs	Frag.				
OF48	-		3	6	-	-	4	7	4	6	1	1	**12**	**2**	**20**	**1.7**
OP	-	-	7	13	-	-	37	58	34	50	6	6	**84**	**2.4**	**127**	**2.4**
ANU	8	57	6	6	-	-	13	16	27	39	6	8	**60**	**2.1**	**126**	**2.1**
CV	5	13	14	28	1	1	20	48	30	64	10	33	**80**	**3.3**	**187**	**3.1**
OPC	-		-		-	-	-		-		-		**0**	**0**	**0**	**0**

Table 4.3.4.1A The number of UPTs homologous to known peptidases identified within each of the *Ophiostoma* EST datasets and the subsequent number of EST fragments used to construct the UPTs.

A list of peptidase homologues identified within OF48 and similar homologues identified within the other *Ophiostoma* datasets is given in Table 4.3.4.1B. A complete list of all peptidases identified in all of the *Ophiostoma* datasets are listed in Appendix 3. In Table 4.3.4.1B the predicted active site and metal ligands were listed as given by the MEROPS database. An active site residue or metal ligand was shown in single letter code followed by the residue number. Where a residue did not match any of those permitted for the active site residue or metal ligand at that position, the permitted amino acids were shown following the residue number. A hyphen preceding a residue number meant that the active site residue or metal ligand in the submitted sequence was missing. An angled bracket preceding an active site residue or metal ligand indicated that the submitted sequence contained only a fragment of the peptidase unit.

	Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
Cysteine Peptidases	07_F2-H-SP6	C14	MER039477	5.00E-23	metacaspase-1	*Gibberella zeae*	<H, <C	
	OF480033	C14	MER039477	1.70E-37	metacaspase-1	*Gibberella zeae*	<H, <C	
	Clv_114221981_gb_EE728655	C14B	MER039472	4.30E-63	metacaspase-1	*Podospora anserina*	H468, >C	
	LMW_58_F_A08_57_053	C14B	MER093173	4.40E-15	metacaspase-1	*Aspergillus niger*	<H, <C	
	LMW_57_F_E01_5_003	C14B	MER039477	1.30E-12	metacaspase-1	*Gibberella zeae*	<H, <C	
	OF480177	C44	MER029814	8.50E-12	family C44 non-peptidase homologues	*Escherichia coli*	C844	
	PIL_90615213_gb_EB048656	C44	MER033304	6.50E-112	family C44 non-peptidase homologues	*Neurospora crassa*	C193	
	PIL_90614304_gb_EB047747	C44	MER033362	6.30E-29	family C44 non-peptidase homologues	*Neurospora crassa*	<C	
Metallo Peptidases	07_C8-AX	M20D	MER092183	3.10E-32	subfamily M20D non-peptidase homologues	*Chaetomium globosum*	<D, <E <D/E, <D, <E, H117	
	CV0793	M20D	MER092128	4.30E-65	subfamily M20D non-peptidase homologues	*Phaeosphaeria nodorum*	<D, <E <D/E, <D, <E, H588	
	PiL_90611981_gb_EB045424	M20D	MER092183	4.10E-70	subfamily M20D non-peptidase homologues	*Chaetomium globosum*	D104, X152E	E102, N116D, X153E, >H
	OP0664	M38	MER033125	5.20E-112	family M38 non-peptidase homologues	*Magnaporthe grisea*	A584H, F586H	
	OP1161	M38	MER033407	4.50E-105	family M38 non-peptidase homologues	*Neurospora crassa*	<H, <H, K5, D49H, Y98H	
	PiL_90612045_gb_EB045488	M38	MER058115	7.00E-05	family M38 non-peptidase homologues	*Sphingomonas wittichii*	>D	<H, <H, L405K, R437H, H451
	YLMW25_G05_39_036	M38	MER060262	6.20E-31	family M38 non-peptidase homologues	*Gibberella zeae*	>D	H201, H203, >K, >H, >H
	CV0834	M38	MER065489	3.80E-25	Pro-Hyp dipeptidase	*Acidobacteria bacterium*	D279	<H, <H, <K, Q194H, H203
	CV1069	M38	MER075387	6.10E-20	family M38 non-peptidase homologues	*Azoarcus sp. BH72*	D71	<H, <H, <K, <H, H1
	07_I7-F-SP6	M38	MER076986	1.60E-43	family M38 non-peptidase homologues	*Acidovorax sp. JS42*	S270D	<H, <H, A125K, L166H, K200H

	Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
Serine Peptidases	PiL_90613650_gb_EB047093	M38	MER090098	1.10E-07	family M38 non-peptidase homologues	*Emericella nidulans*	<D	<H, <H, <K, <H, <H
	OP1326	M38	MER090098	1.70E-53	family M38 non-peptidase homologues	*Emericella nidulans*	D158	<H, <H, <K, H42, H83
	YeastLMW7_C11_83_082	M38	MER090310	7.80E-66	urease	*Emericella nidulans*	<D	<H, <H, <K, <H, <H
	PiL_90610577_gb_EB044020	M38	MER090310	2.30E-118	urease	*Emericella nidulans*	D145	<H, <H, L3K, H57, H105
	PIL_90614535_gb_EB047978	M38	MER090322	1.10E-74	family M38 non-peptidase homologues	*Emericella nidulans*	>D	H201, H203, K292, K340H, >H
	YLMW23_G08_63_056	M38	MER096733	1.40E-06	family M38 non-peptidase homologues	*Maricaulis maris*	>D	<H, <H, S92K, K125H, H156
	YeastLMW7_C11_83_082	M38	MER090310	3.60E-05	urease	*Emericella nidulans*	<D	<H, <H, <K, <H, <H
	OF480024	M41	MER014133	7.40E-50	i-AAA peptidase	*Neurospora crassa*	E515	H514, H518, >D
	OF480125	M74	MER01298	3.30E-119	murein endopeptidase	*Escherichia coli*	H605	H509, D516, H607
	PIL_90614095_gb_EB047538	S09X	MER033025	3.90E-19	family S9 unassigned peptidases	*Magnaporthe grisea*	<S, <D, G265H	
	Clv_114219978_gb_EE726652	S09X	MER033307	1.70E-15	family S9 unassigned peptidases	*Neurospora crassa*	<S, D211, H241	
	Clv_114219978_gb_EE726652	S09X	MER033307	3.00E-09	family S9 unassigned peptidases	*Neurospora crassa*	X682S, >D, >H	
	CV0479	S09X	MER033307	2.20E-29	family S9 unassigned peptidases	*Neurospora crassa*	S402, >D, >H	
	CV0479	S09X	MER033307	3.60E-13	family S9 unassigned peptidases	*Neurospora crassa*	<S, D979, H1009	
	CV0407	S09X	MER034143	1.70E-37	family S9 unassigned peptidases	*Mycobacterium bovis*	S240, D357, H387	
	CV0683	S09X	MER035051	3.60E-33	family S9 non-peptidase homologues	*Gibberella zeae*	<S, <D, H876	
	07_C5-GX	S09X	MER047694	7.10E-09	family S9 unassigned peptidases	*Trichodesmium erythraeum*	V555S, D596, >H	

Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
ANU_631	S09X	MER065419	2.00E-52	family S9 unassigned peptidases	*Gibberella zeae*	<S, D186, H216	
PiL_90608471_gb_EB041916	S09X	MER066433	1.70E-05	family S9 unassigned peptidases	*Ignicoccus hospitalis*	<S, D685, H713	
OP0482	S09X	MER067920	5.50E-05	family S9 non-peptidase homologues	*Myxococcus xanthus*	>S, >D, >H	
LMW_47_G03_23_021	S09X	MER072467	5.40E-05	family S9 unassigned peptidases	*Rickettsia akari*	S90, >D, >H	
CV0283	S09X	MER075602	1.30E-08	family S9 non-peptidase homologues	*Nocardioides sp. JS614*	S649, >D, >H	
Ylmw3_B07_50_057	S09X	MER078361	3.80E-09	family S9 unassigned peptidases	*Methylibium petroleiphilum*	<S, D305, H337	
PiL_90610459_gb_EB043902	S09X	MER093135	3.70E-11	family S9 non-peptidase homologues	*Aspergillus niger*	C319S, E367D, >H	
PiL_90613484_gb_EB046927	S09X	MER101613	7.80E-13	family S9 non-peptidase homologues	*Acaryochloris marina*	T268S, G375D, >H	
PIL_90615672_gb_EB049115	S09X	MER101613	8.30E-05	family S9 non-peptidase homologues	*Acaryochloris marina*	<S, >D, >H	
PiL_90617248_gb_EB050691	S09X	MER101613	1.90E-15	family S9 non-peptidase homologues	*Acaryochloris marina*	E261S, >D, >H	
PiL_90610896_gb_EB044339	S09X	MER101638	1.90E-06	family S9 non-peptidase homologues	*Acaryochloris marina*	E67S, >D, >H	
CV0482	S09X	MER101643	1.20E-15	family S9 non-peptidase homologues	*Acaryochloris marina*	A404S, S480D, D526H	
PiL_90613512_gb_EB046955	S09X	MER101644	2.80E-10	family S9 non-peptidase homologues	*Acaryochloris marina*	<S, A476D, G513H	
07_D9-HX	S09X	MER101652	7.30E-14	family S9 unassigned peptidases	*Acaryochloris marina*	G88S, >D, >H	
LMW_41_B05_34_042	S09X	MER101668	3.50E-07	family S9 unassigned peptidases	*Acaryochloris marina*	<S, G66D, G105H	
CV0277	S09X	MER101719	1.60E-06	family S9 unassigned peptidases	*Acaryochloris marina*	<S, V730D, >H	
OP1158	S09X	MER101719	4.00E-05	family S9 unassigned peptidases	*Acaryochloris marina*	G359S, >D, >H	
PiL_90613888_gb_EB047331	S09X	MER101720	8.20E-05	family S9 unassigned peptidases	*Acaryochloris marina*	<S, A629D, >H	
OP0584	S09X	MER101724	8.20E-07	family S9 unassigned peptidases	*Acaryochloris marina*	-492S, >D, >H	
CV0303	S09X	MER101750	1.50E-13	family S9 unassigned peptidases	*Acaryochloris marina*	<S, A550D, >H	

Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
PIL_90615478_gb_EB048921	S09X	MER101750	1.30E-12	family S9 unassigned peptidases	*Acaryochloris marina*	A23S, >D, >H	
ANU_366	S09X	MER101750	5.40E-08	family S9 unassigned peptidases	*Acaryochloris marina*	432-734 <S, V489D, >H	
CV0971	S09X	MER101958	2.90E-05	family S9 unassigned peptidases	*Cyanothece sp. ATCC 51142*	<S, >D, >H	
Clv_114218419_gb_EE725093	S09X	MER114060	1.10E-25	family S9 unassigned peptidases	*Coxiella burnetii*	S270, D327, >H	
PiL_90611895_gb_EB045338	S09X	MER114216	1.50E-05	family S9 non-peptidase homologues	*Sorangium cellulosum*	<S, I61D, >H	
Clv_114221825_gb_EE728499	S09X	MER114216	2.80E-16	family S9 non-peptidase homologues	*Sorangium cellulosum*	P185S, V293D, >H	
PiL_90616198_gb_EB049641	S09X	MER114216	4.00E-06	family S9 non-peptidase homologues	*Sorangium cellulosum*	>S, >D, >H	
PiL_90616748_gb_EB050191	S09X	MER114450	8.20E-05	family S9 non-peptidase homologues	*Sorangium cellulosum*	<S, >D, >H	
PiL_90609108_gb_EB042553	S09X	MER114493	2.60E-17	family S9 non-peptidase homologues	*Emericella nidulans*	E137S, G221D, >H	
PiL_90609805_gb_EB043248	S09X	MER114556	3.30E-07	family S9 non-peptidase homologues	*Emericella nidulans*	X477S, >D, >H	
OP0667	S09X	MER115415	4.70E-18	family S9 non-peptidase homologues	*Microcystis aeruginosa*	G143S, I255D, >H	
LMW_40_E06_45_040	S09X	MER064110	5.50E-61	esterase D (*Homo sapiens*)	*Gibberella zeae*	S120, D207, H242	
PIL_90614450_gb_EB047893	S09X	MER068641	3.00E-05	family S9 non-peptidase homologues	*Sinorhizobium meliloti*	C217S, >D, >H	
LMW_49_D09_68_074	S09X	MER101719	6.40E-12	family S9 unassigned peptidases	*Acaryochloris marina*	G192S, >D, >H	
OP1057	S09X	MER114450	2.30E-21	family S9 non-peptidase homologues	*Sorangium cellulosum*	A493S, L602D, >H	
ANU_666	S09X	MER033309	3.20E-30	family S9 unassigned peptidases	*Neurospora crassa*	<S, D393, Y416H	
LMW_60_F_E05_37_035	S09X	MER033309	1.70E-56	family S9 unassigned peptidases	*Neurospora crassa*	S116, D228, Y251H	
OP0004	S09X	MER114250	2.90E-06	family S9 non-peptidase homologues	*Sorangium cellulosum*	A847S, >D, >H	
LMW_135_C07_51_050	S09X	MER116474	2.90E-16	family S9 non-peptidase homologues	*Caulobacter sp. K31*	S19, >D, >H	

	Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
Threonine Peptidases	OF480180	S11	MER088022	5.40E-28	family S11 unassigned peptidases	*Escherichia coli*	<S, <K, L83S	
	OF480193	S11	MER088022	1.50E-66	family S11 unassigned peptidases	*Escherichia coli*	V468S, C471K, L532S	
	CV1065	T01A	MER049775	1.30E-89	proteasome subunit beta 2	*Aspergillus fumigatus*		
	LMW_58_F_D07_52_058	T01A	MER064094	1.50E-69	proteasome catalytic subunit 3	*Gibberella zeae*	<T	
	CV0782	T01A	MER064166	8.10E-99	proteasome catalytic subunit 1	*Gibberella zeae*	T226	
	CV0252	T01A	MER083526	9.70E-84	proteasome catalytic subunit 2	*Chaetomium globosum*	<T	
	YeastLMW9_B07_50_057	T01A	MER083526	9.40E-21	proteasome catalytic subunit 2	*Chaetomium globosum*	<T	
	Clv_114222659_gb_EE729333	T01A	MER086130	3.50E-103	proteasome subunit alpha 7	*Aspergillus terreus*	T214	
	07_F11-A-SP6	T01A	MER086130	5.30E-102	proteasome subunit alpha 7	*Aspergillus terreus*	T202	
	PiL_90609540_gb_EB042985	T01A	MER087520	3.80E-60	proteasome subunit beta 3	*Aspergillus clavatus*	G338T	
	CV0079	T01A	MER087527	2.80E-83	proteasome subunit alpha 3	*Aspergillus clavatus*	G426T	
	Clv_114218930_gb_EE725604	T01A	MER087625	1.00E-85	proteasome subunit alpha 6	*Neosartorya fischeri*	M30T	
	CV0932	T01A	MER089994	3.80E-77	proteasome subunit alpha 1	*Emericella nidulans*	<T	
	Ylmw21_F04_30_041	T01A	MER091363	6.20E-58	proteasome subunit alpha 4	*Neosartorya fischeri*	<T	
	Clv_114222302_gb_EE728976	T01A	MER093154	1.80E-44	proteasome subunit alpha 2	*Aspergillus niger*	<T	
	Ylmw21_F04_30_041	T01A	MER091363	4.60E-18	proteasome subunit alpha 4	*Neosartorya fischeri*	<T	
	LMW_30_A07_49_049	T01A	MER093154	4.70E-45	proteasome subunit alpha 2	*Aspergillus niger*	<T	
	ANU_223	T01A	MER087575	5.20E-97	proteasome subunit beta 1	*Aspergillus clavatus*	S191T	

Table 4.3.4.1B List of peptidase homologues identified within OF48 and similar homologues identified within the other *Ophiostoma* datasets.

At least one homologue representing all six of the peptidases families were identified within the *Ophiostoma* EST datasets. 12 OF48 UPTs in total aligned with significant homology to sequences in the MEROPS database. Of these, three were predicted to be cysteine peptidases including two with homology to sequences in the C14 MEROPS family. C14 contains cytosolic endopeptidases termed caspases that have strict specificity for the hydrolysis of aspartyl bonds. Caspases are important enzymes in apoptosis (Earnshaw *et al*., 1999). Genes homologous to the family of caspase peptidases have been identified in all but one (*Kluyveromyces lactic)* of the fungi listed in the database. A homologue to the C44 family of peptidases was identified in one of the OF48 UPTs and two of the *O. piliferum* (OP) UPTs. Homologues were also present in all of the fungi contained within the MEROPS database. C44 peptidases

have been identified as amidophosphoribosyltransferase precursors (synonymously called glutamine phosphoribosylpyrophosphate amidotransferase, ATase, PRAT or GPAT) and function as glycosyltransferases in the biological process of purine biosynthesis in *S. cerevisiae* (Swiss-Prot P04046).

Three OF48 UPTs aligned to bacterial peptidases and had no homologues in any of the *Ophiostoma* datasets. One of these was identified as being homologous to the metallo protease M41. However, homologues to this family of peptidases have previously been identified in all of the 19 fungal species listed within the MEROPS database. M41 peptidases are generally membrane proteins. The eukaryote homologues include the mitochondrial m- and i-AAA peptidases (M41.003, M41.004) which are considered to be essential for the breakdown of uncomplexed components (Tatsuta *et al.*, 2006), and in yeast the m-AAA protease have a chaperone function important for the correct assembly of protein complexes in the mitochondrion, including elements of the respiratory chain and ATP-dependent enzymes. Three homologues to peptidase families, one homologous to the M74 family and two homologous to the S11 family were identified in OF48 UPTS that were not identified in other *Ophiostoma* datasets, and had no homologues present in any of the fungal datasets contained within MEROPS. These homologues were associated with bacterial peptidases. Both families of peptidases are associated with bacterial cell walls. M74 peptidases are involved in the synthesis and lysis of bacterial cell walls and S11 peptidases are mainly involved in the synthesis of bacterial cell walls, cleaving the D-Ala-D-Ala crosslinks in the cell wall peptidoglycans.

In addition to M41 and M74, two other putative metallo proteases were identified. M20D had homologues in both the *O. calvigerum* (CV) dataset and the *O. piliferum* (OP) dataset as well as homologues in all fungi in the MEROPS database with the exception of *Candida giabrata*. The M20D peptidase family contains exopeptidases: carboxypeptidases, dipeptidases and a specialised aminopeptidase. In general, this group of peptidases hydrolyses the late products of protein degradation so as to complete the conversion of proteins to free amino acids. The M38 family of peptidases contain beta-aspartyl dipeptidase a cytosolic enzyme best characterised in bacteria. Its function is the release of iso-aspartate residues from peptides which accumulate during

the stationery phase of bacterial growth and may be toxic (Gary & Clarke, 1995). Homologues of this metallo protease were identified in all of the *Ophiostoma* datasets with the exception of *O. piceae* (PIC) and in 12 of the 19 fungal species present in the database.

The SO9X prolyl oligopeptidase family was the most abundant peptidase identified within the *Ophiostoma* EST datasets. Homologues of this family were identified in all fungi in the databases. Many peptidases in the SO9X family are believed to be important for the degradation of biologically active peptides. For example, DPP-IV metabolises the insulinotropic hormone, glucagon-like peptide 1 (Holst & Deacon, 1998; Deacon & Holst, 2002). Members of the family are found in different cellular locations; prolyl oligopeptidase in intracellular, DDP-IV and fibroblast activation protein α (S09.007) are expressed on cell surfaces; and oligopeptidase B is secreted to the bacterial periplasm (Tsuru, 1998). Serine peptidases have been identified previously in *Ophiostoma* species including *O. floccosum, O. novo-ulmi, O. ulmi* and *O. piceae* (Abraham *et al.*, 1995; Abraham and Breuil, 1996; Hoffman and Breuil, 2002; Wu *et al.*, 2007).

One threonine peptidase family, T01A was identified in the OF48 dataset. Homologues were also identified in all *Ophiostoma* datasets and all but 3 fungi in the database. T01A is a proteasome family of peptidases. The proteasome is involved in the turnover of intracellular proteins, including proteins specifically targeted for degradation by polyubiquitination.

In addition to the MEROPS analysis, annotation of OF48 using TargetIdentifier identified 1 UPT (OF4807_A8-BX) with possible protease function. This UPT aligned based on sequence homology to a UPT identified in the HN EST dataset to form the contig (MX234) (Table 4.3.4A). Both of the UPTs identified in the alignment originated from cDNA libraries specific for blastospore growth.

Cluster ID	Length	Non-redundant Genbank Putative Function	Accession	Score	*E*-value
MX234	894	Patatin-like serine hydrolase, XP_746486, 206, 2e-73	XP_746486	206	2.00E-73

Table 4.3.4.1C Annotation of contig MX234 containing an *O. novo ulmi* yeast specific EST fragment and *O. floccosum* yeast specific EST fragment.

To the best of knowledge, these findings represent the first reported findings of putative peptidases in the aspartic, cysteine, glutamic, and threonine peptidases families in *Ophiostoma* species.

4.3.4.2 Codon Usage

As was discussed in Section 1.2.5.4, it has been established in some organisms that mRNAs for highly expressed proteins preferentially use some codons encoding the same amino acids. The codon usage of predicted genes in up-regulated transcripts identified as being high frequency within the *Ophiostoma* EST datasets, was compared with transcripts identified as singlet sequences within a dataset, therefore, more likely to be lower frequency. Codon frequency was calculated using the following equation:

$$(x / y)1000$$

where x was the total number of codons predicted within all ORFs in a dataset, and y was the sum of all codons identified within the dataset. This equation gave the frequency of a particular codon that codes for an amino acid per thousand peptides.

A summary of codon frequency within the *Ophiostoma* datasets is given in Table 4.3.4.2A, wherein HF represents ORFs predicted in up-regulated or High Frequency transcripts and LF represents ORFs predicted in Low Frequency transcript. HFC for *O. floccosum* represents the codon frequency of all contigs containing 3 or more aligned fragments within the dataset as opposed to greater than 10 aligned fragments in the HF dataset. To determine codon frequency, the single best hit ORF from each UPT investigated was determined using OrFinder (NCBI) and BLASTx analysis.

Amino Acid	Codon	*O. floccosum*			*O. piliferum*		*O. novo-ulmi*		*O. calvigerum*	
		HF	HFC	LF	HF	LF	HF	LF	HF	LF
Phe (F)	TTT	66.7	30.5	13.3	7.2	7.7	11.2	13.1	10.7	13.3
	TTC	12.0	17.7	14.0	14.0	13.6	15.7	17.7	18.3	17.2
	7	7.9	4.8	2.7	2.1	2.1	2.7	3.1	2.9	3.1
Leu (L)	TTA	90.7	37.9	2.1	1.7	1.6	4.1	3.2	1.3	3.2
	TTG	6.6	6.6	7.7	14.3	14.3	9.4	14.0	4.9	15.4
	CTT	12.0	21.7	26.0	20.5	13.6	27.2	16.2	15.6	12.5
	CTC	0.0	14.0	31.0	28.4	23.4	39.6	22.2	38.2	23.6
	CTA	4.4	2.9	5.4	4.1	2.9	2.1	3.9	3.0	5.2
	CTG	8.7	23.9	34.9	26.4	26.2	39.3	37.0	40.5	29.4
		12.2	10.7	10.7	9.5	8.2	12.2	9.6	10.3	8.9
Ile (I)	ATT	29.5	23.4	10.7	5.1	5.9	14.0	13.4	4.7	11.8
	ATC	8.7	21.4	21.0	17.7	14.9	27.4	21.9	25.7	19.3
	ATA	59.0	24.2	4.7	2.6	2.5	2.6	4.5	1.9	5.6
		9.7	6.9	3.6	2.5	2.3	4.4	4.0	3.2	3.7
Met (M)	ATG	31.7	21.9	17.7	16.5	15.4	14.4	17.8	14.9	18.4
		3.2	2.2	1.8	1.7	1.5	1.4	1.8	1.5	1.8
Val (V)	GTT	23.0	24.2	19.1	13.6	9.6	21.9	12.2	11.2	14.0
	GTC	1.1	31.9	31.9	32.5	25.9	38.6	24.5	44.8	27.5
	GTA	15.3	10.3	11.6	9.0	5.7	8.2	7.7	7.1	9.0
	GTG	2.2	7.7	16.1	16.2	15.6	16.0	19.5	27.3	18.9
		4.2	7.4	7.9	7.1	5.7	8.5	6.4	9.0	6.9
Ser (S)	TCT	25.1	15.7	6.3	11.6	10.6	11.4	8.8	3.3	13.3
	TCC	7.7	11.1	9.1	16.0	15.7	17.5	13.3	13.5	16.7
	TCA	23.0	10.3	6.3	10.3	9.3	4.3	12.2	3.2	8.6
	TCG	9.8	9.4	16.7	20.9	26.6	11.1	17.4	18.5	18.4
	AGT	30.6	14.0	7.0	7.0	6.5	7.8	7.1	2.5	7.1
	AGC	2.2	9.7	20.9	18.0	20.5	18.8	20.8	21.8	23.4
		9.8	7.0	6.6	8.4	8.9	7.1	8.0	6.3	8.7
Pro (P)	CCT	19.7	10.5	7.7	13.6	16.5	5.1	8.6	4.4	11.0
	CCC	5.5	15.7	14.9	17.9	14.8	14.7	19.0	12.1	17.0
	CCA	10.9	8.0	11.6	18.0	17.2	7.3	16.0	8.9	16.7
	CGC	5.5	8.3	14.9	19.8	27.3	5.6	21.0	23.5	20.2
		4.2	4.2	4.9	6.9	7.6	3.3	6.5	4.9	6.5
Thr (T)	ACT	19.7	15.4	5.3	7.7	7.6	11.9	8.8	2.4	9.4
	ACC	5.5	21.1	22.3	24.0	15.7	26.2	17.8	24.9	16.1
	ACA	23.0	9.1	8.2	9.2	11.9	7.1	14.6	3.3	12.5
	ACG	7.7	9.7	15.8	16.8	19.9	10.6	15.2	20.1	16.5
		5.6	5.5	5.2	5.8	5.5	5.6	5.6	5.1	5.4
Ala (A)	GCT	29.5	25.1	19.3	18.4	18.7	19.0	17.5	21.1	21.5
	GCC	6.6	35.6	37.5	37.8	40.9	38.9	39.7	68.9	37.1
	GCA	18.6	11.1	14.2	15.2	19.4	11.4	20.7	8.9	20.8
	GCG	3.3	10.0	21.6	24.5	36.9	15.7	23.4	26.6	20.6
		5.8	8.2	9.3	9.6	11.6	8.5	10.1	12.6	10.0
Tyr (Y)	TAT	41.5	18.0	3.9	2.7	2.3	4.1	5.3	4.9	8.6
	TAC	5.5	14.3	16.5	9.0	8.8	13.5	16.0	12.7	13.7
		4.7	3.2	2.0	1.2	1.1	1.8	2.1	1.8	2.2
Ter (end)	TAA	3.3	3.1	1.8	1.4	0.6	2.6	2.4	1.0	0.6
	TAG	2.2	0.9	1.4	2.0	2.6	0.8	1.5	1.1	2.1
	TGA	2.2	2.6	1.4	7.8	7.9	1.5	1.4	1.1	1.9
His (H)	CAT	10.9	10.8	12.3	14.4	10.7	8.4	9.5	6.0	10.7

Amino Acid	Codon	*O. floccosum*			*O. piliferum*		*O. novo-ulmi*		*O. calvigerum*	
		HF	HFC	LF	HF	LF	HF	LF	HF	LF
	CAC	2.2	11.7	17	17.9	14.3	15.2	15.8	17.9	15.5
		1.3	2.3	2.9	3.2	2.5	2.4	2.5	2.4	2.6
Gln (Q)	CAA	13.1	7.7	6.5	12.6	26.9	4.1	7.8	4.6	10.3
	CAG	0	20.2	36.1	11.8	23.4	29.2	26.4	34.3	24.2
		1.3	2.8	4.3	2.4	5	3.3	3.4	3.9	3.4
Asn (N)	AAT	49.2	27.9	14.4	5.4	4.9	11.5	8.6	4.9	10.1
	AAC	7.7	19.7	21.7	14	13.8	28.7	18.6	16.5	18.9
		5.7	4.8	3.6	1.9	1.9	4	2.7	2.1	2.9
Lys (K)	AAA	41.5	23.1	8.6	6.6	5.8	8.2	10.3	6.9	15.4
	AAG	5.5	30.5	35.9	20.3	16.2	50.8	31.5	25.7	29.6
		4.7	5.4	4.5	2.7	2.2	5.9	4.2	3.3	4.5
Asp (D)	GAT	23	24.2	25.2	16.4	10.6	21	17.2	14.4	21.5
	GAC	6.6	25.4	36.6	31.8	28.6	40.3	29.4	46.4	32.6
		3	5	6.2	4.8	3.9	6.1	4.7	6.1	5.4
Glu (E)	GAA	24	16.5	14.2	13.6	10.1	14.2	14.3	11.3	17.6
	GAG	5.5	22.5	31.6	28.7	25.8	35.1	34.1	41.1	35.8
		3	3.9	4.6	4.2	3.6	4.9	4.8	5.2	5.3
Cys (C)	TGT	6.6	4.8	3.5	8.6	11.9	4.8	7.8	2.4	7.1
	TGC	5.5	7.1	10.2	17.4	23	10.1	12.5	9.2	11.4
		1.2	1.2	1.4	2.6	3.5	1.5	2	1.2	1.9
Trp (W)	TGG	3.3	7.4	11.9	15	16.3	10.1	14.6	11.1	14
		0.3	0.7	1.2	1.5	1.6	1	1.5	1.1	1.4
Arg (R)	CGT	1.1	10.8	13	16.5	21.3	11.9	12.1	12.4	9
	CGC	0	13.1	19.1	24.8	39.8	19.3	26.3	30.1	16.9
	CGA	2.2	4.6	11	18.2	22.8	5.4	9.2	8.7	9.2
	CGG	2.2	3.4	15.4	17.4	27.5	6.9	19.9	18.1	16.3
	AGA	19.7	9.7	9.3	10.2	9.4	9.1	10.7	3.8	10.1
	AGG	9.8	5.1	4.4	12.4	14.4	5.4	12.1	4.2	9.2
		3.5	4.7	7.2	10	13.5	5.8	9	7.7	7.1
Gly (G)	GGT	37.2	32.8	23.5	20.2	14.7	25.6	19.5	15.9	13.1
	GGC	7.7	29.6	44	40.6	40.3	43.1	35.2	57.9	40.8
	GGA	30.6	16	11.6	14.1	12.9	9.6	8.5	7.3	11
	GGG	5.5	6.6	11	16.7	12	13.5	10.7	9.9	11.6
		8.1	8.5	9	9.2	8	9.2	7.4	9.1	7.7

Table 4.3.4.2A Codon frequency of predicted ORFs in low frequency (LF) and high frequency (HF) transcripts between different *Ophiostoma* EST datasets.

A list of the preferred codons in up-regulated *Ophiostoma* datasets compared to select organisms is given in Table 4.3.4.2B. The organisms selected for comparison included

two yeast species (*Pichia pastoris* and *Saccharomyces cerevisiae*) both commonly used as recombinant gene expression hosts, the filamentous Pezizomycotina *Aspergillus nidulans* and up-regulated genes in humans.

Amino Acid	*O. floccosum*	*O. piliferum*	O. novo-ulmi	O. clavigerum	*P. pastoris*	*A. nidulans*	*S. cerevisiae*	*H. sapiens*
Phe	TTT	TTC	TTC	TTC	TTC	TTC	TTC	TTC
Leu	TTA	CTC	CTC/CTG	CTG	TTG	CTC/CTG	TTG	CTG
Ile	ATA	ATC	ATC	ATC	ATC	ATC	ATC/ATT	ATC
Val	GTT	GTC	GTC	GTC	GTT/GTC	GTC	GTT/GTC	GTG
Tyr	TAT	TAC	TAC	TAC	TAC	TAC	TAC	TAC
His	CAT	CAC	CAC	CAC	CAC	CAC	CAC	CAC
Asn	AAT	AAC	AAC	AAC	AAC	AAC	AAC	AAC
Lys	AAA	AAG	AAG	AAG	AAG	AAG	AAG	AAG
Asp	GAT	GAC	GAC	GAC	GAC	GAC	GAC	GAC
Glu	GAA	GAG	GAG	GAG	GAG	GAG	GAA	GAG
Ser	AGT	TCG	AGC	AGC	TCT/TCC	TCC	TCT/TCC	TCC/AGC
Pro	CCT	CGC	CCC	CGC	CCA	CCA/CCC	CCA	CCC
Thr	ACA	ACC	ACC	ACC	ACC	ACC	ACC/ACT	ACC
Ala	GCT	GCC	GCC	GCC	GCT/GCC	GCC	GCT	GCC
Arg	AGA	CGC	CGC	CGC	AGA	CGC	AGA	AGA/AGG
Gly	GGT	GGC	GGC	GGC	GGT	GGC	GGT	GGC
Gln	CAA	CAG	CAG	CAG	CAA/CAG	CAA	CAA	CAG
Cys	TGT/TGC	TGC	TGC	TGC	TGT/TGC	TGC	TGT	TGC

Table 4.3.4.2B List of preferred codons in up-regulated *Ophiostoma* data sub-sets and from select organisms taken from the literature (Sinclair and Choy, 2002).

Codon usage was consistent for up-regulated transcripts in all of the *Ophiostoma* datasets with the exception of *O. floccosum* and conferred with the previously reported codon usage in the hydrophobic cerato-ulmin (CU) protein in *O. ulmi* (Bowden *et al.*, 1994). The calculated codon usage of predicted *O. floccosum* genes in the HF sub-set showed no similarity with the other *Ophiostoma* HF transcripts analysed with the exception of cysteine. However, the codon usage of predicted LF genes was the same as for all *Ophiostoma* HF genes with the exception of proline and serine amino acids.

The GC% of high frequency and low frequency subsets was compared and listed in Table 4.3.4.2C. Also the positioning of guanine and cytosine nucleotides within the codon triplet nucleotide sequence was plotted (Figure 4.3.4.2).

Organism	High Freq.		Low Freq.	
	Genes	GC %	Genes	GC%
O. floccosum	7	30.2	34	58.5
O. piliferum	25	62.6	16	64.9
O. novo-ulmi	36	57.2	50	59.2
O. clavigerum	44	65.3	22	58.3

Table 4.3.4.2C The number of genes and percentage of guanines and cytoseines within high frequency and low frequency *Ophiostoma* data sub-sets.

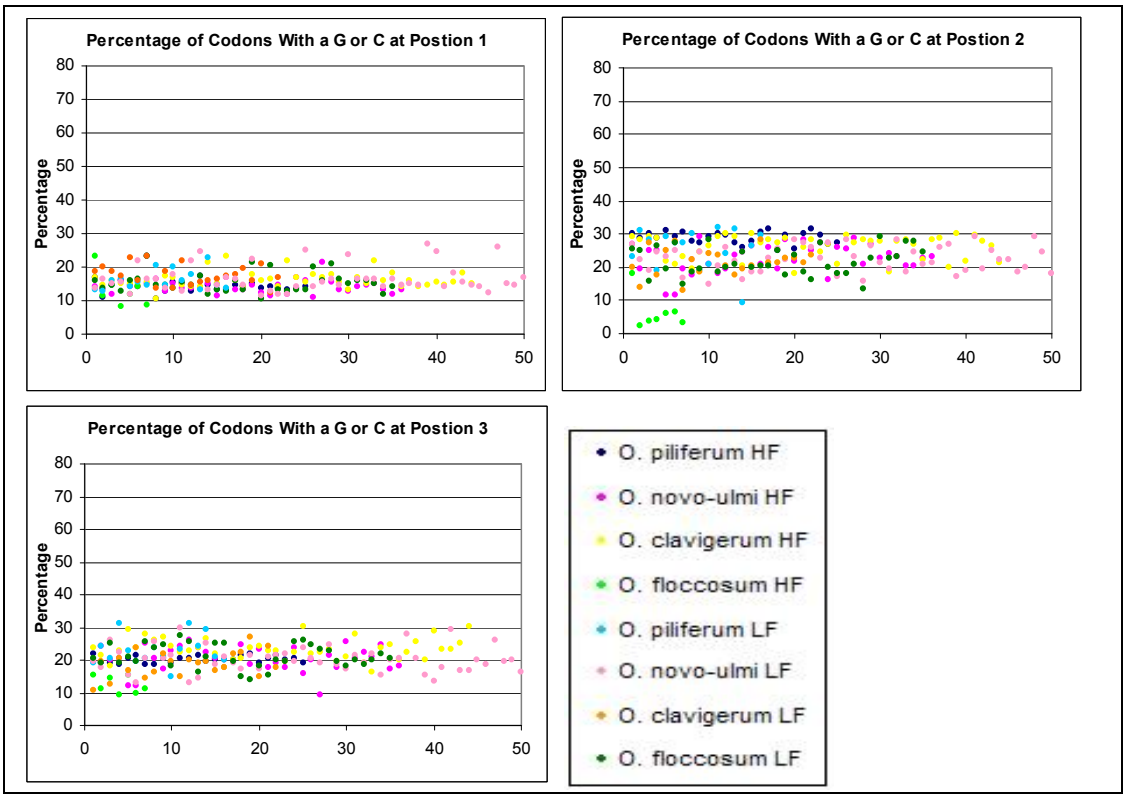


Figure 4.3.4.2 The positioning of guanine and cytosine nucleotides within the codon triplet for high frequency and low frequency *Ophiostoma* data sub-sets.

In the *O. floccosum* HF predicted genes, the GC content of this sub-set was markedly different compared to the other *Ophiostoma* species analysed. The low GC content of these ORFs is likely to be due to the lack of G or C in the second position. This is somewhat surprising as the wobble position or the third base in each codon is where much of the degeneracy of the genetic code resides (Gustafsson, 2004). For example in organisms with a high G/C content, G/C are preferred in the third position. The reverse is true as organisms with a high A/T content which avoid G/C in the third position. The codon usage for the hydrophobic cerato-ulmin (CU) gene identified in *O. ulmi* was reported to have a significant C bias in the wobble position (Bowden *et al*., 1994). This data indicated that this was not the case for *O. calvigerum, O. novo-ulmi, O. piliferum* or *O. floccosum*.

Codon usage bias may be influenced by many factors including the interactions of mutation, selection, random drift, effective population size, evolutionary history, biased gene conversion, mRNA secondary structure, and translational initiation (Kamatani and Yamamoto, 2006). An explanation for the marked difference in the *O. floccosum* HF sub-set may be that many of the transcripts identified as being up-

regulated were mitochondrial. In *S. cerevisiae* the mitochondrial genetic code has been shown to differ from the universal code (Jukes and Osawa, 1990; Swire *et al*., 2005). Mitochondrial genes are often AT rich (Knight *et al.*, 2001). It is likely then that with so few genes analysed in the *O. floccosum* HF sub-set, the presence of these mitochondrial genes may not reflect the codon usage of nuclear DNA in this species.

5 Molecular Cloning and Expression Profiling of Select Abundantly Transcribed *Ophiostoma floccosum* and *Ophiostoma piliferum* Genes

5.1 Introduction

Specific goals listed for this PhD thesis research included identifying up-regulated transcripts in the *O. floccosum* OF48 EST dataset and the *O. piliferum* OP EST dataset, and further elucidating these genes to determine the nucleotide sequence and if possible any transcriptional elements upstream and downstream of the predicted ORF, in order to further the understanding of gene expression and regulation in *Ophiostoma* sp. A central premise to this thesis is that up-regulated transcript is fairly indicative of transcriptional control by a strong promoter. To fully exploit the *Ophiostoma* genus as a recombinant expression host, it would be desirable to identify an array of gene promoters for the different stages and somatic forms of growth.

Wu *et al.* (2006) identified a number of highly secreted proteins in *Ophiostoma floccosum* produced after five days of mixed culture growth in minimal media containing 3% soluble starch. One of these proteins, an α-amylase was further investigated to identify potential transcriptional elements. A 6.5 kb fragment was amplified containing the 5' promoter region, the α-amylase enzyme coding region which contained three introns, and the 3' terminator region. Within the promoter region, a few key transcriptional factors were identified including a possible TATA box, a conserved CCAAT box, reported to be a transcription factor binding site in *A. nidulans* and *S. cerevisiae*, and five putative CAAT motifs (Wu *et al*., 2006).

A significant biotechnological advantage of many *Ophiostoma* species is their ability to grow as both a yeast-like and filamentous fungi on a large scale in a stirred tank bioreactor. Therefore, in this PhD thesis research, an investigation of up-regulated transcripts in early stage yeast-like growth in *O. floccosum* was conducted to likely identify a promoter suitable for eventual use in a recombinant expression system. It is hoped that such a promoter would instigate gene expression in early stage growth in either shaken liquid cultures or in a stirred tank bioreactor.

Non-normalised EST analysis is a quantitative method of determining mRNA abundance at a given time. This method, therefore, readily identifies up-regulated transcripts within a dataset. The most up-regulated transcripts from both *O. floccosum* and *O. piliferum* were identified from datasets OF48 and OP as discussed throughout Chapter 4. These up-regulated transcripts were as follows:

- OF48 contig OF0064, putatively identified as a NADH-ubiquinone oxidoreductase subunit 4; and
- OP contig OP0411, identified as a putative heat shock protein from both *O. floccosum* and *O. piliferum* EST data sets.

Genome 'walking', a method of elucidating flanking genomic segments was used to 'walk' upstream and downstream of the UPTs to identify ORFs and transcriptional elements such as promoter regions.

Promoters may be constitutive or they may be either active or inactive depending on the developmental stage and environmental stimuli such as nutrient availability. As mycelial and blastospore growth in liquid media in *O. floccosum* and *O. piliferum* is not exclusive to just one somatic form and that the level of one somatic form in relation to another can be influenced by cultivation conditions, the levels of expression of OF480064 and OP0411 in blastospore compared to mycelial forms throughout growth were investigated.

Within this chapter differences in the level of expression of OF480064 and OP0411 were determined in blastospore and mycelial states at 24 hour intervals throughout growth in liquid media using RT-PCR. This resulted in the construction of 7 blastospore-specific, and 7 mycelial specific cDNA libraries, constructed from cells in culture between 24 and 192 hours for both *O. floccosum* and *O. piliferum*. The RT-PCR results were comparatively analysed to identify any differences in the levels of OF480064 and OP0411 mRNA transcripts present in the cDNA libraries.

5.1.1 Specific Objectives For Molecular Cloning and Expression Profiling

Two specific objectives were addressed in this part of the thesis research and are discussed in this chapter, as follows:

1. To further elucidate the ORFs and transcriptional elements in the 5' promoter regions of *O. floccosum* UPT OF0064, predicted by sequence homology to code for a NADH-ubiquinone oxidoreductase subunit 4, and *O. piliferum* UPT OP0411, predicted by sequence homology to code for a heat shock protein.
2. To evaluate the level of OF0064 and OP411 expression throughout different stages in two somatic forms of growth, yeast-like and mycelia, in order to identify possible differences in transcript abundance as a function of somatic form and/ or culture stage.

5.2 Construction of *O. floccosum* and *O. piliferum* GenomeWalker™ Libraries

The GenomeWalker™ DNA walking protocol involved constructing four adaptor-ligated genomic DNA fragments called D1-D4 generated from the ligation of supplied adaptor sequences to blunt digested *Ophiostoma* DNA. The resulting libraries were then used to PCR amplify upstream, and in the case of HSP, downstream sequence using two sets of nested gene specific primers. Refer to Section 2.4 of this thesis for a detailed description of the GenomeWalker™ protocol.

Ophiostoma DNA was isolated in duplicate, respectively, from 1.5 grams of *O. floccosum* and from 1.5 grams of *O. piliferum* cultures shaken at 180 rpm for 72 hours as described in Section 2.4.2.1. The DNA was assessed for the level of degradation by gel electrophoresis before being used in the genomic DNA library construction. These gels, with the DNA visualised by Ethidium Bromide fluorescence are shown in Figures 5.2A(A) and 5.2A(B). The DNA was judged from the fluorescence not to be significantly degraded as it did not produce significant smearing to a low molecular weight. Using spectrophotometric analysis, each genomic DNA isolation contained between 2062 – 5653 ng/µl of DNA in a final volume of 50 µl .

Eight GenomeWalker™ libraries were constructed from *O. floccosum* and *O. piliferum* using purified, RNAase treated, high-molecular weight genomic DNA that was digested with blunt ended restriction enzymes (refer to Section 2.4.2 for GenomeWalker™ library construction). According to the methods described in

Section 2.4.2.2, four different restriction enzymes were used to construct four libraries from each species. Multiple restriction enzymes were used in order to increase the likelihood of obtaining a long amplicon. Each restriction enzyme digestion reaction (100 µl volume) contained 2.5 µg of DNA. Gel electrophoresis using 1% agarose revealed that all enzymes completely digested the *Ophiostoma* DNA, as is shown in Figures 5.2B(A) and 5.2B(B). Each digested library was purified and ligated to adaptor sequences provided by Clonetech according to the manufacturers instructions.

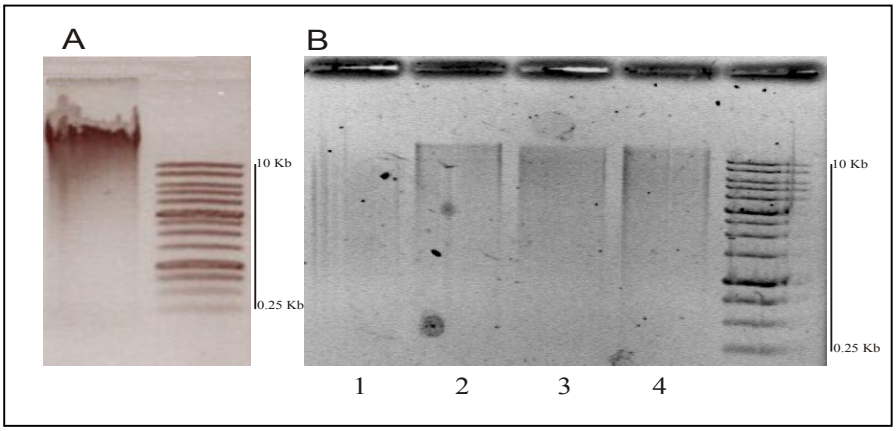


Figure 5.2A (A) *O. floccosum* genomic DNA used in genome walking (4 µl at 2613 ng/µl).
(**B**) Blunt restriction enzyme digests of *O. floccosum* genomic DNA (Library 1 (*Dra* I), Library 2 (*Eco*R V), Library 3 (*Pvu* II), Library 4 (*Stu* I)).

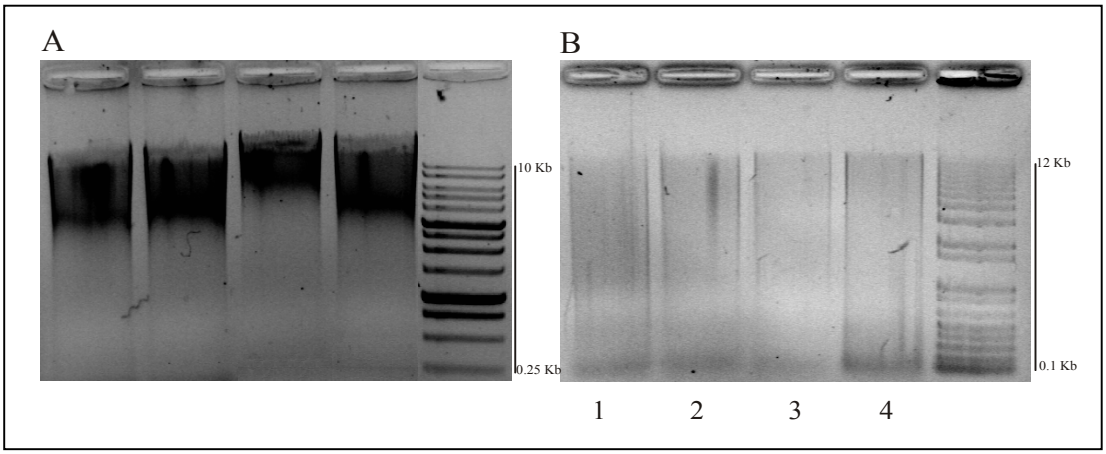


Figure 5.2B (A) *O. piliferum* genomic DNA used in genome walking (4 µl at 4986-5717 ng/µl).
(**B)** Blunt restriction enzyme digests of *O. piliferum* genomic DNA (Library 1 (*Dra* I), Library 2 (*Eco*R V), Library 3 (*Pvu* II), Library 4 (*Stu* I)).

5.3 *Ophiostoma floccosum* NADH dehydrogenase

5.3.1 Introduction

OF0064 was identified in the *O. floccosum* EST dataset OF48 as having the greatest frequency. The contig consisted of 59 aligned EST fragments representing approximately 4% of the OF48 dataset. BLASTx analysis of OF0064 identified it as being most homologous to a NADH-ubiquinone oxidoreductase (NADH dehydrogenase) subunit identified in various ascomycetes, as shown in Table 5.3.1A.

Name/Function Of Aligned Proteins Identified By BLASTx	Score	E-value
NADH dehydrogenase subunit 4 -Verticillium dahliae (YP 667826)	270	5e-71
NADH dehydrogenase subunit 4 - Podospora anserina (NP_074953)	266	7e-70
NADH dehydrogenase subunit 4 - Fusarium oxysporum (AAW67491)	265	2e-69

Table 5.3.1A Top 3 hits identified by BLASTx alignment of OF0064.

In total 135 OF48 EST fragments aligned as 4 contigs and 4 singlet sequences, were identified using TargetIdentifier software as being most homologous to various NADH-ubiquinone odxidoreductase subunits (Table 5.3.1B). This represented 11 % of the total library strengthening the argument that genes involved with metabolism likely dominate gene expression in early phase blastospores.

Sequence Identity	No. Fragments	Name/Function Of The Highest Score Pair in BLASTX	E-value
OF Contig_14	38	(Q8SHP7) NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3)	0
OF Contig 32	4	(P15959) NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3) (NADH dehydrogenase subunit 6)	4.00E-35
OF Contig 35	30	(P05510) NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3) (NADH dehydrogenase subunit 5)	5.00E-29
OF0064	59	(P15582) NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3) (NADH dehydrogenase subunit 4)	9.00E-72
OF 07_A1-GX	1	(Q02854) NADH-ubiquinone oxidoreductase 21 kDa subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-21KD) (CI-21KD)	7.00E-60
OF 4F-y	1	(P15578) NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3) (NADH dehydrogenase subunit 2)	6.00E-49
OF 05_2H-y	1	(P15578) NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3) (NADH dehydrogenase subunit 2)	9.00E-56
OF 05_3A-y	1	(P15578) NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3) (NADH dehydrogenase subunit 2)	4.00E-61

Table 5.3.1B Summary of OF48 contigs and singlets that aligned most readily with NADH-ubiquinone oxidoreductase subunits using TargetIdentifier software.

5.3.2 GenomeWalker™ PCR Amplification

GenomeWalker™ methodologies were employed to sequence upstream of the 5' end of OF0064 to identify the promoter region associated with this contig, and presumptively the NADH-ubiquinone oxidoreductase gene. A downstream region was also amplified using Genome Walker methodologies, however, the segment produced was within the length of the starting UPT and, hence, not included in the results. Further attempts at sequencing downstream of the 3' end of OF0064 failed. Primary and secondary PCR reactions were conducted with NAD 1 and NAD 2 primers specific for the 5' end of OF0064 (see Table 2.3.1 Section 2.3.1). *O. floccosum* genome walking genomic DNA libraries 1, 2, 3 and 4 were used as template for each reaction.

Analysis of the primary PCR results revealed an amplicon from the 5'end of approximately 500 bp and the appearance of smearing in the 3 other libraries (Figure 5.3.2(A)). Analysis of the secondary PCR results revealed amplification of DNA fragments in one of the libraries of approximately 500 bp (named NH730_3) in the 3'-5' direction from the 5' end OF0064 (as evidenced by the PCR reaction bands visualised by Ethidium Bromide fluorescence and shown in Figure 5.3.2(B)).

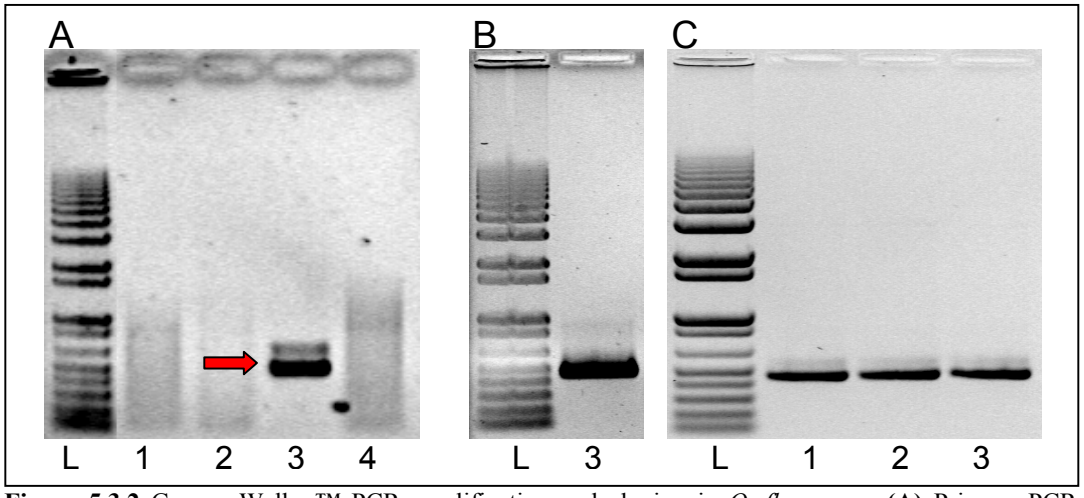


Figure 5.3.2 GenomeWalker™ PCR amplification and cloning in *O. floccosum*. **(A)** Primary PCR reaction results (L= ladder). The red arrow indicates a 500 bp amplicon in library 3. **(B)** Secondary PCR reaction results indicating 500 bp amplicon in library 3. **(C)** PCR analysis of 500 bp NADH-ubiquinone Oxidoreductase insert cloned in pCR®4-TOPO vector.

5.3.3 Cloning of PCR Product Into pCR®4-TOPO

The NH730_3 amplicon (~500 bp) was excised and cleaned using the E-Gel® CloneWell system as described in Section 2.5.1 and cloned into pCR®4-TOPO according to the ligation methods described in Section 2.5.2. Diagnostic digests with *Eco*RI and PCR amplification using NAD 2 gene specific and Clonetech AP2 primers revealed that a 500 bp amplicon successfully ligated into pCR®4-TOPO. The resulting clones were subsequently named NH730_3 X (X= 1-10). The PCR reaction bands visualised by Ethidium Bromide fluorescence and shown in Figure 5.3.2(C)).

5.3.4 Sequencing Analysis

Forward and reverse sequencing of NH730_3 1, 2, 3 further revealed that a fragment upstream of OF0064 had successfully been cloned into pCR®4-TOPO. Sequences were screened and trimmed accordingly, as described in Section 2.5.6.1, to remove contaminating vector and poor quality sequence. Screened sequence data was then aligned and assembled using Vector NTI (Invitrogen) to form a consensus sequence NADH-UR4; the alignment of these sequences is given in Figure 5.3.4A. The blue portion of the consensus scale bar represented consensus sequence derived from the alignment of 6-7 fragments. Areas of cross hatching indicated alignment between 3-4 fragments and single lines indicated only a single fragment used to form part of the consensus sequence.

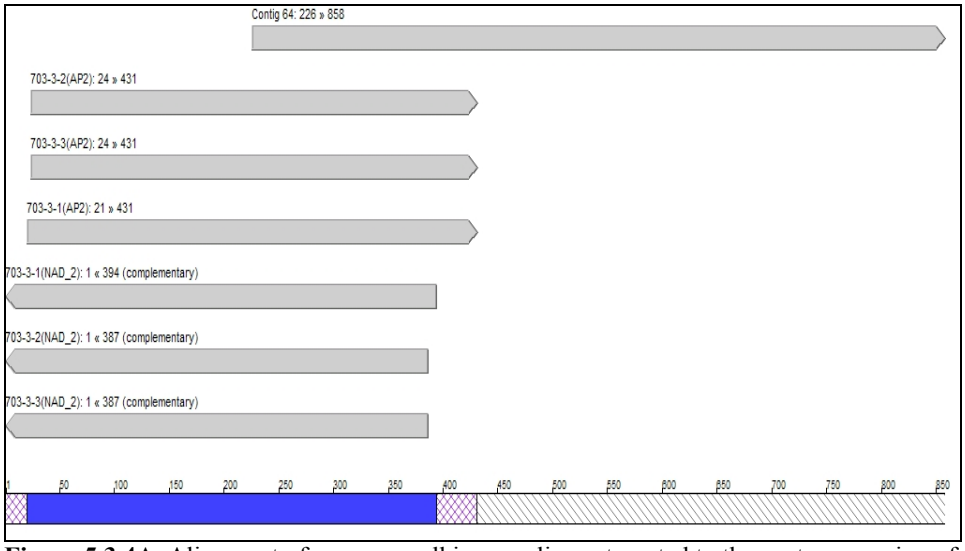


Figure 5.3.4A Alignment of genome walking amplicons targeted to the upstream region of *O. floccosum* UPT OF0061.

The entire nucleotide sequence elucidated, and subsequent predicted amino acid sequence are given in Figure 5.3.4B. No stop signal was predicted indicating that further sequencing downstream of the 3' region was likely required to elucidate the full ORF for NADH-UR4.

```
CGACGGCCCGGGCTGGTAAAACAAATTTCGATTATATAACACAATTATTTTTATTTGGAGGAGTATTT
TTATCTTTTGCTGTAAAAACTCCTATATGAGGACTAAATAATTGATTATTAAAGGCTCACGTTGAATC
TCCTTTAGGTGGAAGTATAGTATTAGCGGCTATAGTGTTAAAAACTAGTTTATATGGTATATGTAGAC
TAATTTTACCTATATGCCTAAAGCTTCAATTAATTTTACTTATATAGTTTATGTAATAGGTGTTGTT
                M  P  K  A  S  I  N  F  T  Y  I  V  Y  V  I  G  V  V
ACAATAATTTATGCTAGTTTTAGTACATTAAGAACTACAGATATTAAAGAATTAATAGCTTATAGTTC
T  I  I  Y  A  S  F  S  T  L  R  T  T  D  I  K  E  L  I  A  Y  S  S
TGTATCTCATGCTGCAGTATATTTAATAGGAGTGTTTAGTAATACAATACAAGGTATAGAAGGAAGTA
 V  S  H  A  A  V  Y  L  I  G  V  F  S  N  T  I  Q  G  I  E  G  S  I
TAATTTTAGGTTTAGCTCACGGATTTGTTTCTAGTGGTTTATTTATATGTGCAGGTGGTATATTATAT
   I  L  G  L  A  H  G  F  V  S  S  G  L  F  I  C  A  G  G  I  L  Y
GATAGATCAGGAACTAGATCTATTTATTTTTATAAAGGTATAGCTCAAATTATGCCATTATTTTCTAT
D  R  S  G  T  R  S  I  Y  F  Y  K  G  I  A  Q  I  M  P  L  F  S  I
ATTATTCTTTATATTATCTTTAGGTAATTGTGGTGTACCTTTAACATTAAATTTTATAGGTGAATTTA
 L  F  F  I  L  S  L  G  N  C  G  V  P  L  T  L  N  F  I  G  E  F  M
TGTCTCTTTATGGAGTATTTGAGAGATTACCTTTATTAGGTGTATTTGCAAGTTCTTCTATTATATTA
   S  L  Y  G  V  F  E  R  L  P  L  L  G  V  F  A  S  S  S  I  I  L
TCTGCAGCATATACAATGTATATGTTTAATAGAATAGGTTTTGGAGGTACATTTAGTAAATTTTTTAA
S  A  A  Y  T  M  Y  M  F  N  R  I  G  F  G  G  T  F  S  K  F  F  K
AGAAAACATAATTGATGTTACAAAAAGAGAATTCGGATCCGATATCGCCATGGCCTTGTCGTCGTCGT
   E  N  I  I  D  V  T  K  R  E  F  G  S  D  I  A  M  A  L  S  S  S  S
CGGTACCCAGATCTTGGTATTCTATAGTGTCACCTAAATCAA
   V  P  R  S  W  Y  S  I  V  S  P  K  S
```

Figure 5.3.4B NADH-UR4 nucleotide and predicted amino acid sequence

BLASTx analysis of NADH-UR4 resulted in higher scoring hits as shown in Table 5.3.4A compared to previously identified homologous proteins, which are shown in Table 5.3.2.

Accession Number	BLASTx Hit	Score	*E*-value
YP_667826	NADH dehydrogenase subunit 4 [Verticillium dahliae]	371	4.00E-101
AAW67491	NADH dehydrogenase subunit 4 [Fusarium oxysporum]	365	2.00E-99
AAX21834	NADH dehydrogenase subunit 4 [Fusarium oxysporum]	365	2.00E-99

Table 5.3.4A BLASTx analysis of NADH-UR4

The enzyme NADH-ubiquinone oxidoreductase oxidises NADH, transferring two electrons to ubiquinone with concurrent translocation of four protons across the mitochondrial membrane (Clason *et al.*, 2007). Complex 1, a large multi subunit, mitochondrial enzyme, is the most synonymous and commonly identified enzyme responsible for electron transfer from NADH to the ubiquinone pool of the respiratory chain in bacteria, plants and fungi (Joseph-Horn *et al.*, 2001; Kerscher *et*

al., 2001). Complex 1 examined in fungi to date have been composed of at least 35 subunits. Seven of these subunits are thought to be mitochondrially encoded subunits in eukaryotes. The remaining subunits are predicted to be encoded by nuclear genes which are synthesised in the cytoplasm and imported into the mitochondria. Without a full 3'sequence the use of prediction based software to determine the likely location of the gene is prevented. BLASTn analysis of NADH-UR4 nucleotide sequence identified it as being most homologous to a 778 bp stretch of *Verticillium dahliae* mitochondrion complete genome (DQ351941) as shown in Table 5.3.4B. The homology of NADH-UR4 to mitochondrially encoded sequence is an indication that NADH-UR4 is likely to be mitochondrially encoded rather than a nuclear gene.

Accession Number	BLASTn Hit	Score	*E*-value
DQ351941.1	Verticillium dahliae mitochondrion, complete genome	754	0
DQ364632.1	Gibberella zeae mitochondrion, complete genome	645	0
AY916130.1	Epidermophyton floccosum mitochondrion, complete genome	601	2.00E-168

Table 5.3.4B BLASTn analysis of NADH-UR4

5.3.5 Identification of Transcriptional Elements

Transcriptional elements upstream of a 5' partial ORF identified as a putative NADH-ubiquinone oxidoreductase protein in NADH-UR4 were investigated *in silico.* Transcriptional elements were identified based on conserved motifs identified from literature. A number of possible transcriptional elements and potential factor binding sites were identified, as shown in Figure 5.3.5 and listed in Table 5.3.5. Within table 5.3.5 the column headed Motif indicated the conserved motif identified within the sequence, the column headed No. referred to the number of these conserved motifs identified within the 5' promoter region and Position indicated the position of the conserved motif from the start of the nucleotide sequence. Highlighted in yellow, the cap site, defined as +1, is a transcription initiation sequence or start point at which the transcription process actually starts; upstream, towards the 5' end, the base pairs are connotated with a negative (-) number, and downstream, towards the 3' end, they are connotated by continuing positive (+) number. Three possible TATA boxes were identified (bordered and in red text in Figure 5.3.5). Similar to other promoter regions investigated for *Ophiostoma* genes, including the *O. floccosum* α-amylase (Wu *et al.*, 2006), *O. ulmi* CU gene (Bowden *et al.*, 1994), and the *O. novo-ulmi COL1* gene (Pereira *et al.*, 2000), a predicted CAAT box was

identified at -170 bp from the predicted start site. In many eukaryotes The CAAT box has been demonstrated to have an important role in increasing promoter strength (Bezhani *et al.*, 2001). In plants it is replaced by a consensus sequence called an AGGA box (CTGGWWW). A sequence matching an AGGA box was also identified at -198 bp (written and underlined in blue in Figure 5.3.5). Using TESS analysis many other putative transcriptional factors were identified. Protein production in eukaryotes is typically controlled by a multiplicity of factors. In many cases the role of activators is to recruit transcription factors to the promoter site rather than directly recruit the polymerase (Jaillon, *et al.*, 2008). Those putative transcription factors identified by TESS analysis that were found to have homologues in fungal species are described in Table 5.3.5 and shown in Figure 5.3.5. All putative transcription factors identified by TESS analysis are listed in Appendix 4.

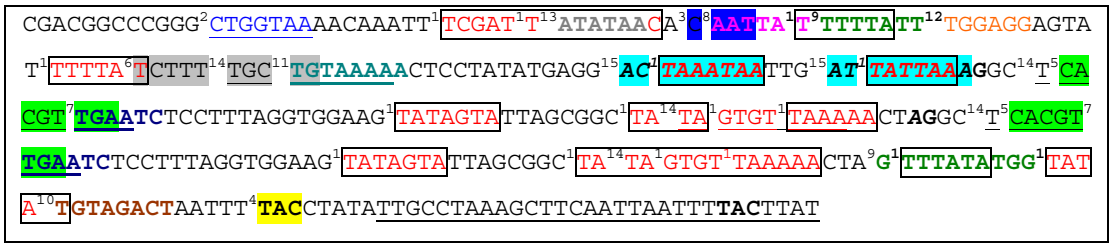


Figure 5.3.5 Putative transcriptional factors identified in the 5' region of NADH-UR4 sequence. The start of the predicted ORF is underlined in black. Descriptions of the identified transcription factors are given in Table 5.3.5.

.

	Motif(s)	Description	No.	Position	TESS Factor
1	TATA, TATAT, TATTAA, TTTTTA, TATATA, TATAAC, ATTAAA, TAAAAA, TTTATA, TAAATA, GTTAAA, TCGATTATATA, TATAGTA, GTGTTAAAA, TAAATAA, TATATAA	TATA box. A conserved motif found in a number of eukaryotes including *S.cerevisiae* and *S. pombe*. Activator of all three polymerases in *S. cerevisiae*	19	33, 35, 48, 66, 82, 103, 115, 116, 167, 173, 175, 185, 194 28, 152, 171, 33, 103	TFIID, TBP
2	CTGG (AAA/TTT)	AGGA box. A consensus sequence found in plants. It replaces the CAAT box which is close to -80 bp from the start point. It plays an important role in promoter efficiency, by increasing its strength and it seems to function in either orientation.	1	12	
3	CAAT	CAAT box	1	42	
4	TAC	CAP site. A transcription initiation sequence or start point at which the transcription process actually starts			
5	CACGTTGA	An ARS-binding protein, identified in S. cerevisiae involved in sorting of mtDNA, mitochondrial matrix protein, and mtDNA recombination	1	126	ABF-2
6	TCTTTTGCTG	A conserved motif identified in *S. cerevisiae*	1	72	CUP2
7	**TGAATC**	A conserved motif identified in *S. cerevisiae*	1	131	GCN4
8	**AATTATT**	Homeobox motif found within a number of genes. Involved in morphogenesis of fungi, animals and plants	1	43	Homeo-box
9	**GTTTATATGG TTTTATTTGG**	MADS-box. They are transcription factors present in all multi-cellular eukaryotes from fungi to plants and humans where they regulate developmental pathways	2	49, 184	MADS
10	**TGTAGACT**	A conserved motif identified in *S. cerevisiae*	1	198	PHO2
11	**TGTAAAAA**	A conserved motif identified in *S. cerevisiae*	1	80	SEF1
12	**TGGAGG**	Positive regulator of peroxisomal protein genes identified in *S. cerevisiae* . It is required in yeast for ADH2 activation and glycerol metabolism	1	56	T00011 ADR1
13	ATATAA	Activator, mediates galactose response, repressed by GAL80. It binds to DNA and to nucelosome assembled DNA. It binds selectively to human and yeast replication factor (RPA)	1	34	T00302 GAL4
14	TGCTGTAAAA TCACGTTGAA TAGTGTTAAA	A repressor of a-specific genes in alpha cells in *S. cerevisiae*. Along with a1 it represses haploid-specific genes in diploid cells.	3	77, 125, 169	T00487 MATalp-ha2
15	***ATTATTAAA ACTAAATAA***	A transcription factor identified in humans that has homologues (MATa1) in *S. cerevisiae*	3	101, 113	T01481 Pbx-1a

Table 5.3.5 A description of the transcription factors identified in the 5' region of NADH-UR4.

5.3.6 Profile Of Expression At Different Stages Of Growth Using RT-PCR

The relative levels of NADH-UR4 expression throughout growth in liquid media, and in different growth forms, blastospore and mycelial, were determined using RT-PCR.

5.3.6.1 Culture Growth and RNA Isolation

Nineteen 400 ml conical flasks, each containing 100 ml of YM media, were inoculated with 2 x 10^5 *O. floccosum* strain J2122 blastospores (refer to Sections 2.1.7 and 2.1.8 for details of the methodology). Duplicate cultures were sampled every 24 hours for 8 days with the exception of 168 hours when no sample was taken for either culture. Given the inherent variability between fungal cultures, the samples were analysed to determine blastospore count (refer to Section 2.1.3), and viability (refer to Section 2.15). This data was then compared to previously determined growth data produced under identical conditions to establish if cultures were representative of standard growth. The number of blastospores present per ml and the average viability measured in CFUs of culture sampled at various time points was plotted against a standard curve (Figure 5.3.6.1A). In terms of blastospore count and viability, the levels detected in cultures used for RT-PCR analysis were comparable to those numbers used to construct the standard curve.

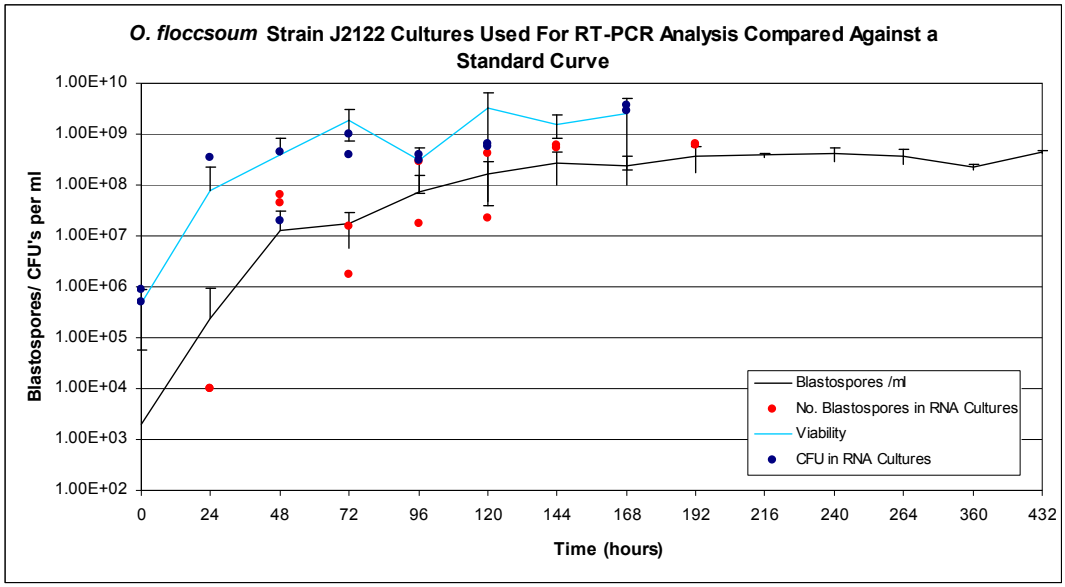


Figure 5.3.6.1A Graph showing the number of blastospores and CFUs determined in duplicate cultures used in RT-PCR analysis, harvested at different time points compared against standard curves.

RNA was spectrophotometrically quantified (as given in Table 5.3.6.1) and assessed visually for quality using agarose gel electrophoresis (as shown in Figure 5.3.6.1B). Assessment was based on the amount of degradation (smearing) as well as the visibility and density of ribosomal RNA bands. RNA samples found to have very

weak or no apparent ribosomal RNA bands, or that had been severely degraded were not used for the construction of cDNA.

O. floccosum Cultures			
Sample	Time Harvested (hours)	Concentration (ng/µl)	260/280
Mycelium	24	787	1.94
Spore	24	-3	1.55
Mycelium	48	162	1.78
Spore	48	3115	2.00
Mycelium	72	3426	1.99
Spore	72	1452	2.03
Mycelium	96	3225	2.01
Spore	96	1010	2.02
Mycelium	120	3317	1.97
Spore	120	338	1.86
Mycelium	144	3538	1.95
Spore	144	264	1.85
Mycelium	192	79	1.87
Spore	192	158	1.89

Table 5.3.6.1 Spectrophotometric analysis of *O. floccosum* RNA isolations used for RT-PCR analysis.

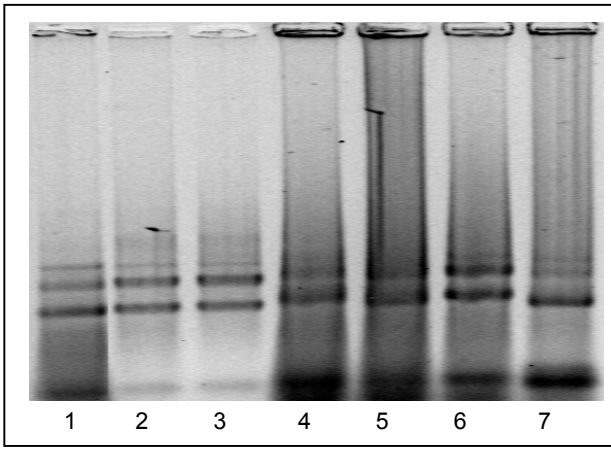


Figure 5.3.6.1B Ethidium bromide stained electrophoresis image of RNA isolations (1 = 48 hour *O. piliferum* strain 97 mycelium, 2,3 = 48 hour *O. piliferum* strain 97 blastospores, 4,5 = 72 hour J2122 mycelium, 6,7 = 72 hour J2122 blastospores).

5.3.6.2 Construction Of cDNA

cDNA was constructed from RNA samples immediately following RNA isolation to ensure minimum degradation of the samples. All cDNA was constructed according the protocol listed in section 2.2.5. To normalise samples, 1 µg of RNA was used in each reaction to construct 20 µl of cDNA. Two *O. floccosum* samples, 24 hours blastospores and 192 hours mycelium were not sufficiently concentrate despite repeated attempts to isolate RNA, with no significant RNA isolated from 24 hour spores and a total of 632 ng used to construct cDNA from 192 hour mycelium.

5.3.6.3 RT-PCR Amplification

RT-PCR was conducted using NADH-UR4 gene specific primers NADH F and NADH R. The primers were designed with a melting temperature of between 50 and 65 ºC (Table 5.3.6.3). Amplification produced an amplicon of 119 bp in length.

Gene	Primer	Nucleotide Sequence	Tm (NN) ºC	Amplicon Size
NADH	NADH F	5'- GCTCACGGATTTGTTTCTAG -3'	57.4	119
NADH	NADH R	5'- GAAAATAATGGCATAATTTGAG -3'	55.5	

Table 5.3.6.3 NADH-UR4 specific oligonucleotides used in RT-PCR analysis.

RT-PCR reactions were run according to the method described in Sections 2.6.5 and 2.6.6. Dissociation curve analysis was performed after every reaction to identify the presence of contaminating sequence. RT-PCR product was also analysed by electrophoresis on a 0.7% agarose gel to confirm amplicon size and purity.

5.3.6.4 Determination of RT-PCR Efficiency

As real-time quantification is based on the relationship between initial template amount and the cycle threshold (C_T) value obtained during amplification, an optimal PCR assay was required for accurate and reproducible comparative results. The efficiency of RT-PCR amplification of *O. floccosum* NADH gene was determined by creating a standard curve from duplicate, and in some instances triplicate, tenfold serial dilutions of cDNA template. The cDNA template was constructed from RNA isolated from blastospore and mycelium growth cultivated for 96 hours.

The amplification curves of the serial dilutions (Figures 5.3.6.4A and 5.3.6.4B) show the exponential increase in fluorescence detection as a function of the thermocycling process. PCR cycle number is shown on the x-axis and the fluorescence from the amplification reaction (Delta Rn), which is proportional to the amount of amplified product in the tube, is shown on the y-axis. During the exponential phase of the amplification plot, the amount of PCR product should approximately double in each cycle. However, as the reaction proceeds, components are consumed and ultimately one or more of the components becomes limiting. At this stage the reaction slows and enters a plateau phase.

The C_T of a reaction is determined mainly by the amount of template present at the start of the amplification reaction. The C_T value was measured in the exponential phase. at threshold 0.4 for all NADH-UR4 RT-PCR analysis.

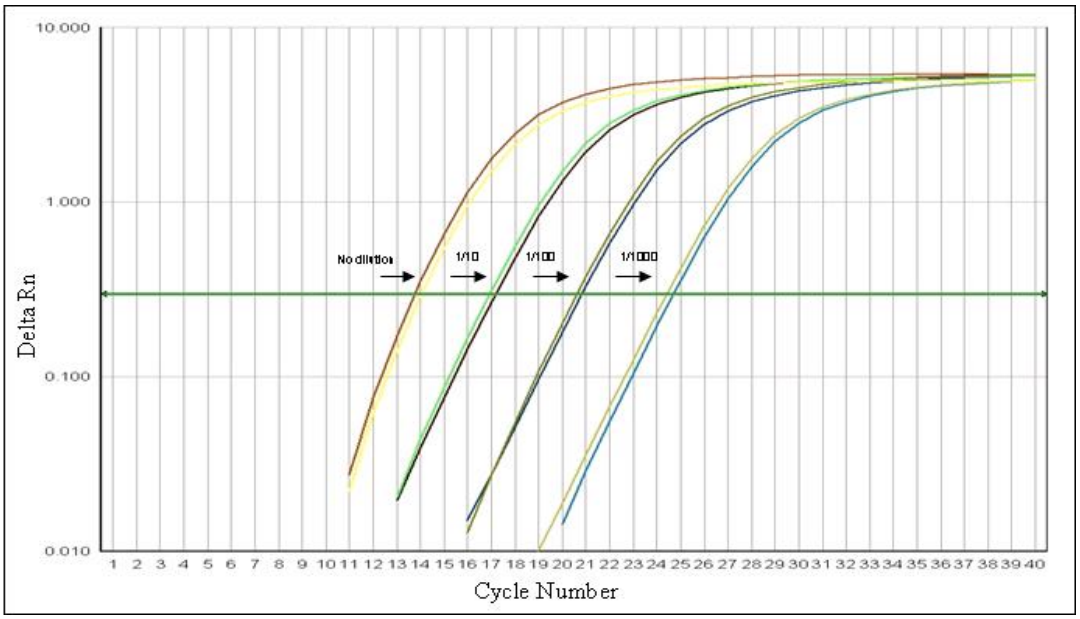


Figure 5.3.6.4A Amplification plot showing Delta Rn against Cycle Number for duplicate RT-PCR analysis of serial dilutions of cDNA template constructed from mycelium cultivated for 96 hours.

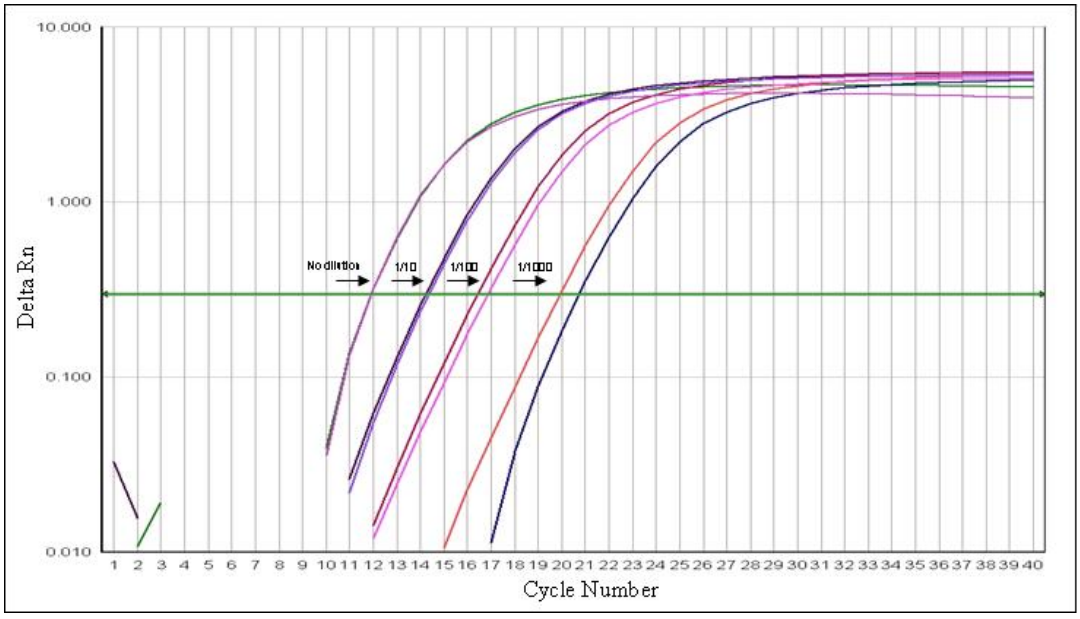


Figure 5.3.6.4B Amplification plot showing Delta Rn against cycle number for duplicate RT-PCR analysis of serial dilutions of cDNA template constructed from blastospores cultivated for 96 hours.

Dissociation curve (melt curve) analysis involved the fluorescent monitoring (470nm/510nm) while 'melting' the dsDNA into single strands. The dissociation of

dsDNA was coupled with a loss in detected fluorescence due to the release of bound SYBR® Green I. The point at which the amplification products dissociated was labeled the melting temperature TM and was a function of the product sequence properties (i.e. GC %) and length.

The fluorescence data obtained was subsequently translated into melting curves by removing background fluorescence and any confounding influence of temperature on fluorescence, then plotting as the negative derivative of fluorescence, with reference to temperature (ºC) (Ririe *et al.*, 1997). The construction of dissociation curves allowed for the detection in most cases of contaminating and/or non specific product such as primer dimers. Dissociation curve analyses of RT-PCR NADH-UR4 amplicons had a melting temperature of approximately 73ºC and are illustrated in Figures 5.3.6.4C and 5.3.6.4D.

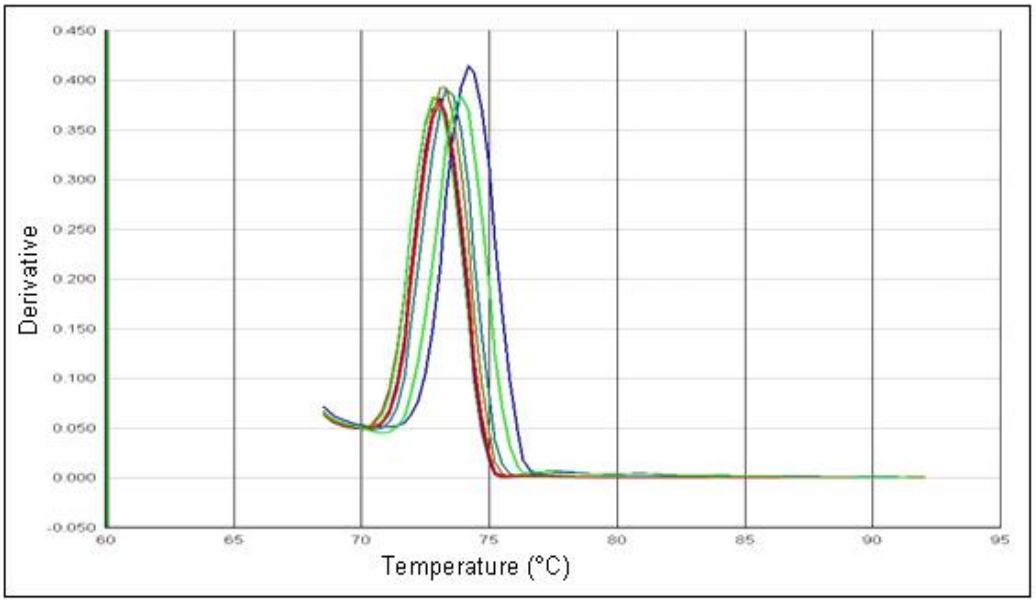


Figure 5.3.6.4C Dissociation curve showing Derivative against Temperature for duplicate RT-PCR analysis of serial dilutions of NADH-UR4 cDNA template constructed from mycelium cultivated for 96 hours.

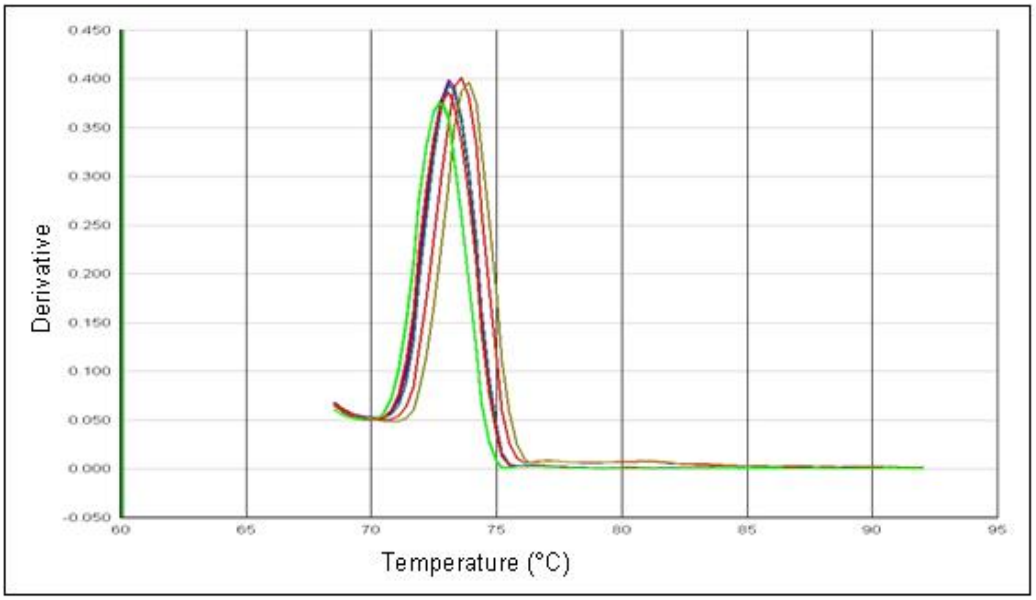


Figure 5.3.6.4D Dissociation curve showing Derivative against Temperature for duplicate RT-PCR analysis of serial dilutions of NADH-UR4 cDNA template constructed from blastospores cultivated for 96 hours.

To further confirm the purity of amplicons produced during RT-PCR analysis, resulting product from the assay was analysed by electrophoresis on a 0.7% agarose gel (Figure 5.3.6.4E).

Dissociation curve analysis of the NADH-UR4 cDNA serial dilutions and electrophoretic analysis did not indicate the presence of any contaminating sequence.

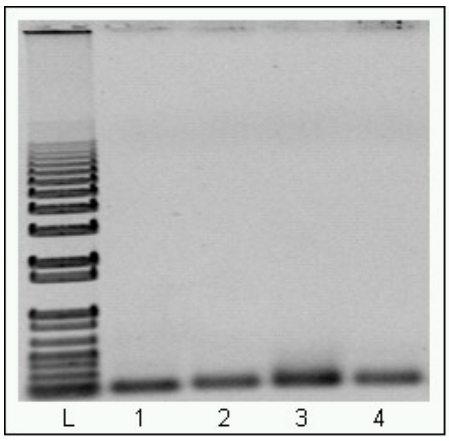


Figure 5.3.6.4E Ethidium bromide stained electrophoresis image of RT-PCR amplicons produced using NADH gene specific primers run on a 0.7% agarose gel. L = ladder, 1 = 96 hour blastospore standard concentration of cDNA template (2 µl), 2 = 96 hour blastospore 1/10 dilution of cDNA template, 3 = 96 hour mycelium standard concentration of cDNA template (2 µl), 4 = 96 hour mycelium 1/10 dilution of cDNA template.

The standard curves were constructed by plotting the log of the dilution factor of template against the averaged C_T value obtained during amplification of each

dilution. The coefficient of determination (R^2) was calculated using the following equations and used to evaluate the level of RT-PCR assay optimisation.

$Y = mx + b$ where m is the slope and b is the intercept

$R^2 = 1 - SSE/SST$ where $SSE = \sum(Y_j - \tilde{Y}_j)^2$ and $SST = (\sum Y_j^2) - (\sum Y_j)^2 / n$

An optimised RT-PCR reaction should show a linear standard curve ($R^2 > 0.980$ and consistency across replicate reactions.

RT-PCR analysis of *O. floccosum* cDNA constructed from 96 hour mycelia and blastospore growth produced R^2 values of 0.9979 and 0.9818 respectively, indicating relatively consistent amplification (Figure 5.3.6.4F).

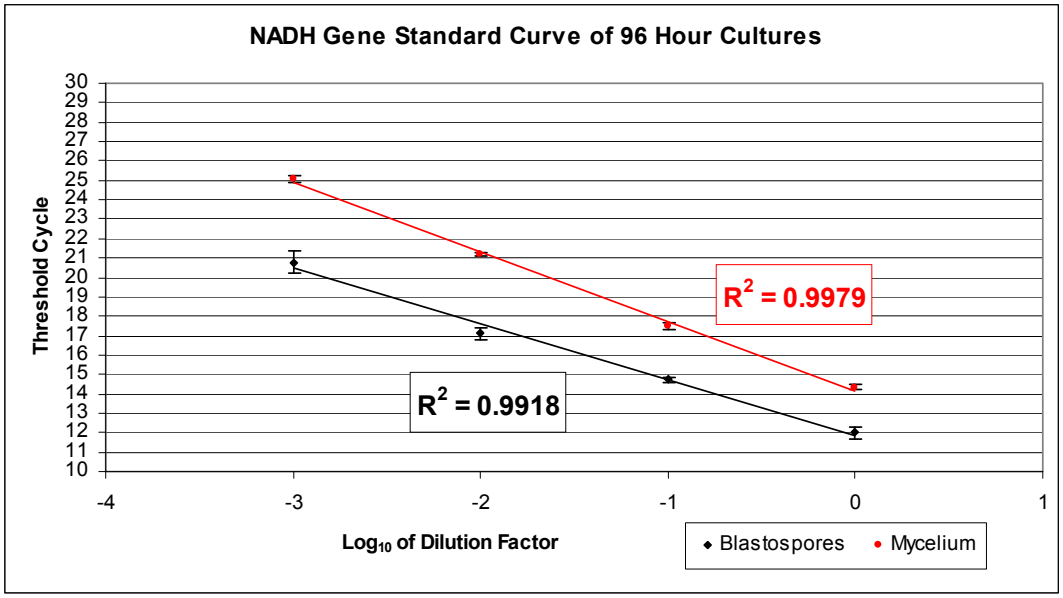


Figure 5.3.6.4F Standard curve used to calculate the efficiency of NADH-UR4 RT-PCR.

The PCR efficiency was calculated using the following equation:

PCR efficiency = $10^{\wedge(1/S)} - 1$

PCR efficiency % = $(10^{\wedge(1/S)} - 1)100$

Where S = the slope of all standard curve samples assayed using NADH-UR4 gene specific primers.

$S_{(NADH\text{-}UR4)} = -3.19$

PCR efficiency $\%_{(NADH\text{-}UR4)} = 105\ \%$

A PCR efficiency of 105% indicated that at the end of each cycle, the amplicon copy number increased by 1.105-fold or 105% of the template was amplified. Efficiency close to 100% is the best indicator of a robust and reproducible assay. Ideally one should strive for an amplification efficiency of 90 – 105 %.

5.3.6.5 NADH-UR4 Expression In Different Somatic States Throughout Growth

The relative levels of expression in the mycelial and blastospore somatic forms were compared in 14 cDNA samples constructed over seven time points. All samples were run in duplicate and in some instances triplicate RT-PCR assays under identical conditions.

Dissociation curve analysis of NADH-UR4 mycelial and spore cDNA samples did not indicate the presence of any contaminants as shown in Figures 5.3.6.5A and 5.3.6.5B respectively.

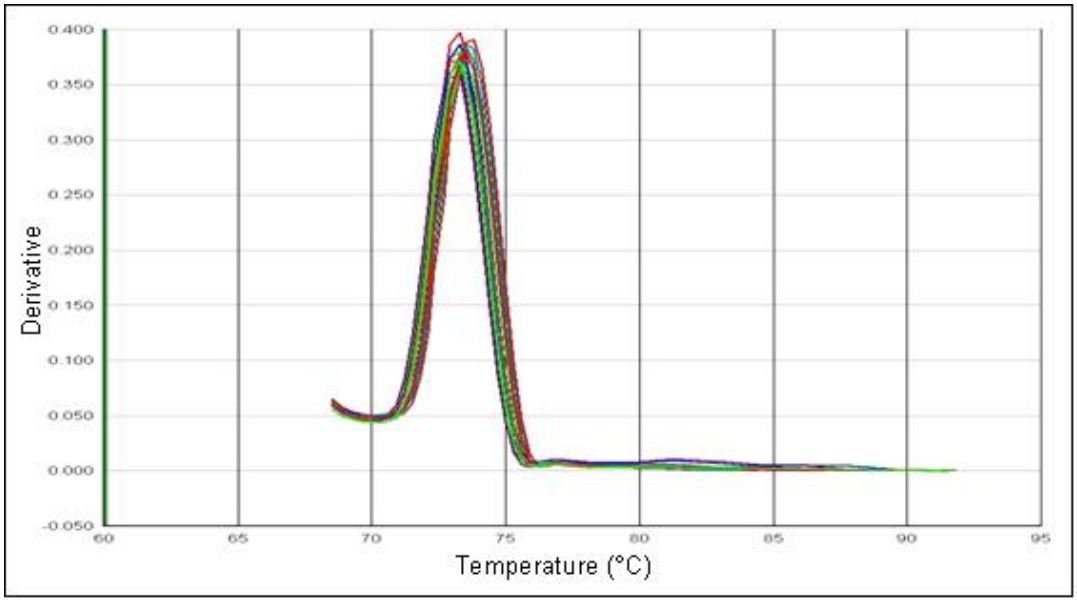


Figure 5.3.6.5A. Dissociation curve showing Derivative against Temperature for NADH-UR4 mycelial samples at different stages of growth.

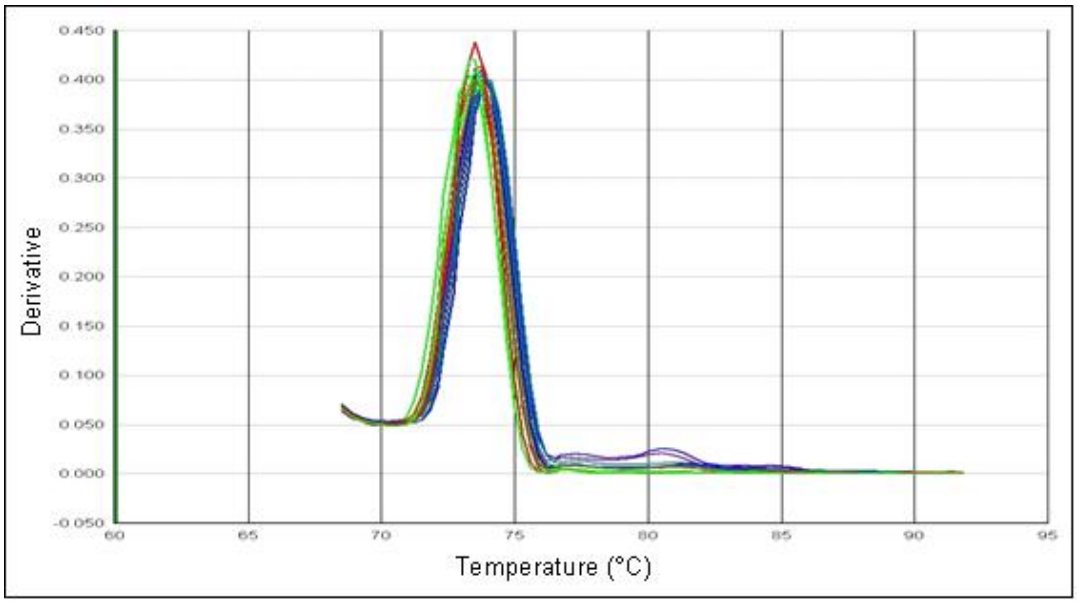


Figure 5.3.6.5B. Dissociation curve showing Derivative against Temperature for NADH-UR4 blastospore samples at different stages of growth.

The average C_T values calculated for between 2-4 PCR replicates at given time points and in both forms of growth are presented in Table 5.3.6.5. Lower C_T values were indicative of higher levels of mRNA transcript in the sample. The data indicated that NADH-UR4 expression was detected in both blastospore and mycelial growth at all stages of growth sampled. [Though average C_T values for blastospores sampled at 24 hours and mycelia at 192 hours were included in the summary table to demonstrate that expression was detected, given that the samples did not contain sufficient RNA, they were not comparative.]

	Blastospore		Mycelia	
Sample Time	Average C_T	Standard deviation	Average C_T	Standard deviation
24	18.659	0.389	13.057	0.053
48	10.848	0.261	14.661	0.138
72	10.500	0.519	14.711	0.347
96	11.989	0.314	14.353	0.140
120	12.116	0.170	12.737	0.191
144	10.089	0.312	11.987	0.067
192	11.477	0.108	11.961	0.252

Table 5.3.6.5 Table of the average C_T value and standard deviations produced by RT-PCR analysis of NADH-UR4 at different time points throughout growth in blastospore and mycelial forms.

The relative expression of NADH-UR4 in mycelium compared to blastospore growth was compared using the following equations and is presented in Figure 5.3.6.5C.

$$\text{Ratio}_{(\text{blastospores / mycelium})} = E^{C_{T\,(\text{mycelium})} - C_{T\,(\text{blastospores})}}$$

Where $E = 10^{\wedge(1/S)}$

And S = the slope of all standard curve samples assayed using NADH-UR4 gene specific primers.

$E = 10^{\wedge(1/3.19)}$
$= 2.05$

$Ratio_{(blastospores\ /\ mycelium)} = 2.05^{\Delta C_T}$

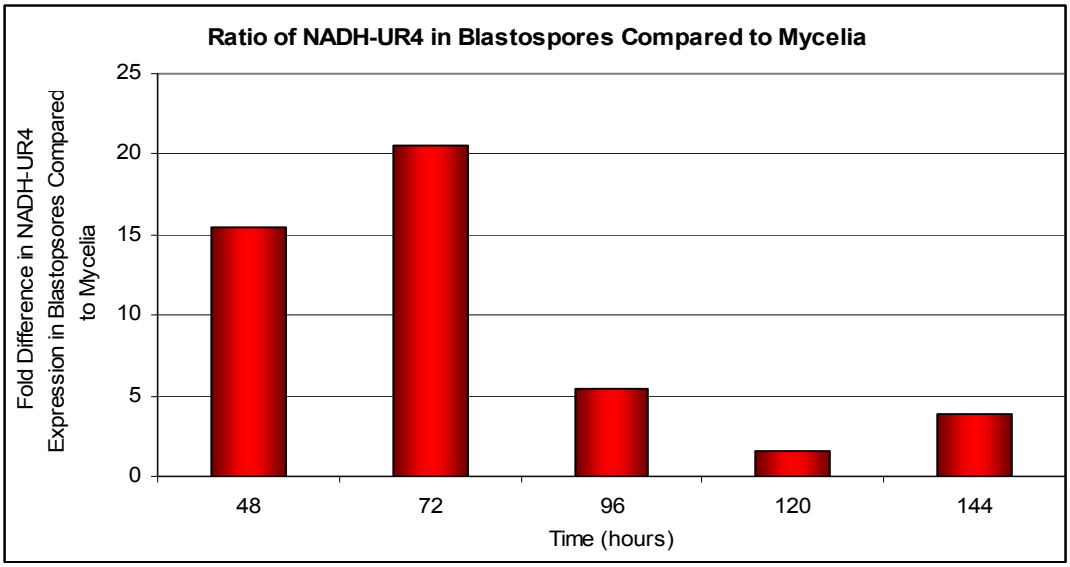


Figure 5.3.6.5C The ratio of NDH-UR4 expression in blastospores compared to mycelia at different time points in *O. floccosum* strain J2122 cultures.

From the bar graph of Figure 5.3.6.5C it is clear that NADH-UR4 expression is up to 20-fold higher in blastospores compared to mycelia, and that expression was higher in blastospores compared to mycelia at all time points measured.

5.4 *Ophiostoma piliferum* heat shock protein

5.4.1 Introduction

OP0411 UPT was identified in the *O. piliferum* EST dataset OP as being the most abundant transcript. The contig consisted of 178 aligned EST fragments representing approximately 1.9 % of the OP dataset.

BLASTx analysis revealed similarity of OP0411 to a number of unknown fungal proteins. Although not identified as being the most homologous, some homology was shown between OP0411 and a *N. crassa* 30 kDa heat shock protein (Table 5.4.1).

Name/Function Of Aligned Protein Identified By BLASTx	Score	E-value
hypothetical protein MGG_00766 - Magnaporthe grisea (XP 368478)	219	1.00E-54
hypothetical protein FG10050.1 - Gibberella zea (XP 290226)	216	5.00E-54
hypothetical protein NCU04635 - Neurospora crassa (XP 959178)	214	2.00E-53
30 kDa heat shock protein - Neurospora crassa (XP_963774)	134	3.00E-29

Table 5.4.1 The results of BLASTx alignment of OP0411

5.4.2 GenomeWalker™ PCR Amplification

Primary and secondary PCR reactions were conducted with HS1, HS2, HS3 and HS4 primers specific for the 5' and 3' end of OP0411 (see Table 2.3.1 Section 2.3.1). *O. piliferum* genome walking genomic DNA libraries 1,2,3 and 4 were used as template for each reaction.

Analysis of the primary PCR results revealed an amplicon in library 4 from the 5' end of approximately 1500 bp. Amplicons ranging from approximately 200 to 900 bp from the 3' end were detected from amplification of libraries 1, 2 and 4 (Figure 5.4.2A). Analysis of the secondary PCR results, termed HSP_(x) with x = 1-8, revealed the amplification of DNA fragments ranging from the 5' end of approximately 1000 and 1500 bp from libraries 3 and 4, respectively. Amplicons from the 3' end of 200, 900 and 300 bp were detected from libraries 1, 2 and 4 and were visualised by Ethidium bromide fluorescence as shown in Figure 5.4.2B.

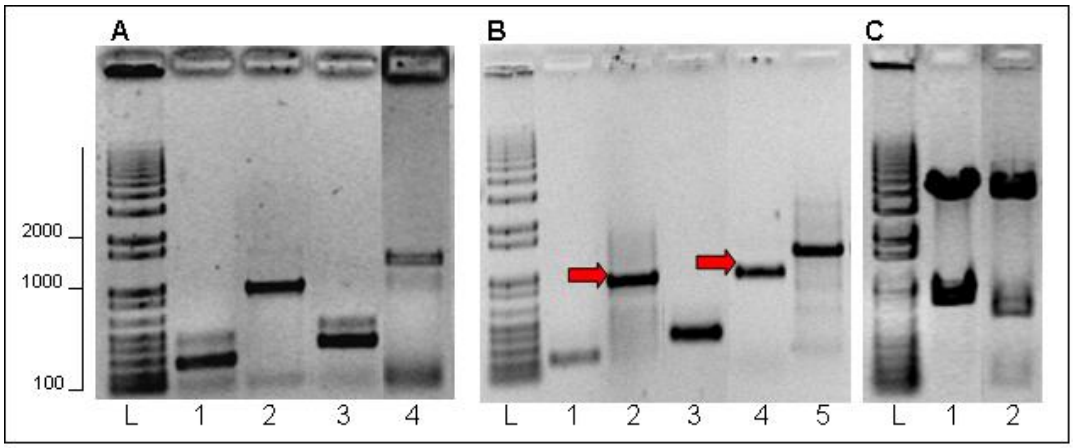


Figure 5.4.2 GenomeWalker™ PCR amplification and cloning in *O. piliferum*. **(A)** Primary PCR reaction results (1 = amplification of 3'end using library 1, 2 = amplification of the 3'end using library 2, 3 = amplification of the 3'end using library 3, 4 = amplification of the 5'end using library 4). **(B)** Secondary PCR reaction results. The red arrows indicate amplicons used in cloning reactions (1 = amplification of 3'end using library 1(HSP_1) , 2 = amplification of the 3'end using library 2 (HSP_2), 3 = amplification of the 3'end using library 4 (HSP_4), 4 = amplification of the 5'end using library 3 (HSP_7), 5 = amplification of the 3'end using library 4(HSP_8). **(C)** *Eco*RI digestion of HSP clones (1= HSP_71, 2= HSP_21). (L= ladder).

5.4.3 Cloning of PCR Product Into pCR®4-TOPO

HSP_2 and HSP_7 amplicons were excised and cleaned using E-Gel® CloneWell system as described in Section 2.5.1 and cloned into pCR®4-TOPO according to the ligation methods described in Section 2.5.2. Diagnostic digests with *Eco*RI revealed that inserts of approximately the correct size had successfully ligated into pCR®4-TOPO. The resulting clones were subsequently named HSP_2(x) and HSP_7(x) (Figure 5.4.2C).

5.4.4 Sequencing Analysis

Forward and reverse sequencing of HSP_7 clones 1, 2 and 3 further revealed that a fragment upstream of OP0411 was successfully cloned into pCR®4-TOPO. Attempts at forward and reverse sequencing of HSP_2(1,2 and 3) were unsuccessful. Rather than repeating the cloning process, successful forward and reverse sequencing of the 3' end of OP0411 was achieved by directly sequencing the PCR product HSP_2. Sequences were screened and trimmed accordingly to remove contaminating vector and poor quality sequence. Screened sequence data was then aligned and assembled using Vector NTI (Invitrogen) as shown in Figures 5.4.4A. 5.4.4B and 5.4.4C, to form a consensus sequence HSP. The blue checkered and hatched portions of the consensus scale bars represented consensus sequence derived from the alignment of 2 fragments. Portions illustrated with diagonal lines indicated where only a single fragment was used to form part of the consensus sequence.

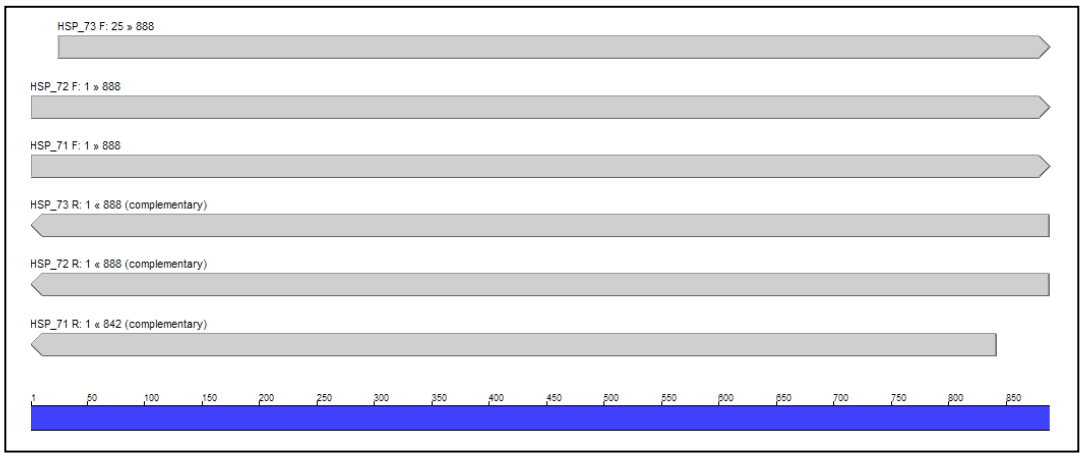


Figure 5.4.4A Alignment of genome walking amplicons targeted upstream of the 5' region of *O. piliferum* UPT OP0411.

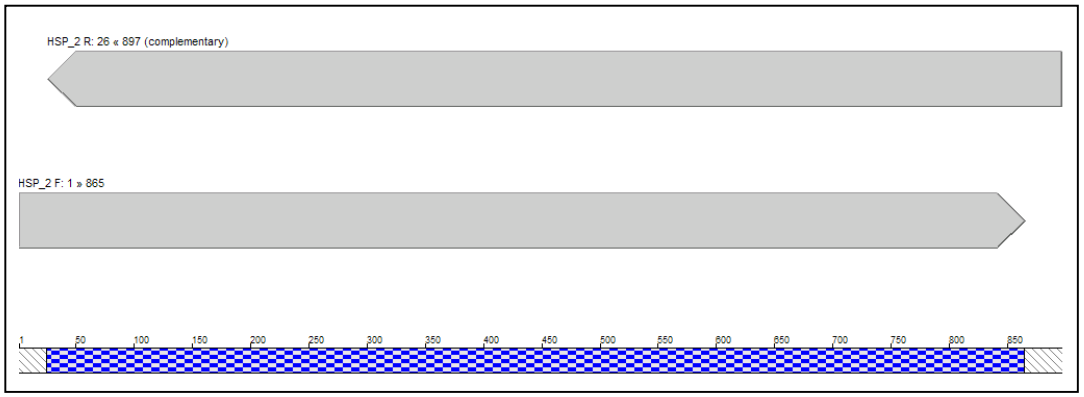


Figure 5.4.4B Alignment of genome walking amplicons targeted downstream of the 3' region of *O. piliferum* UPT OP0411.

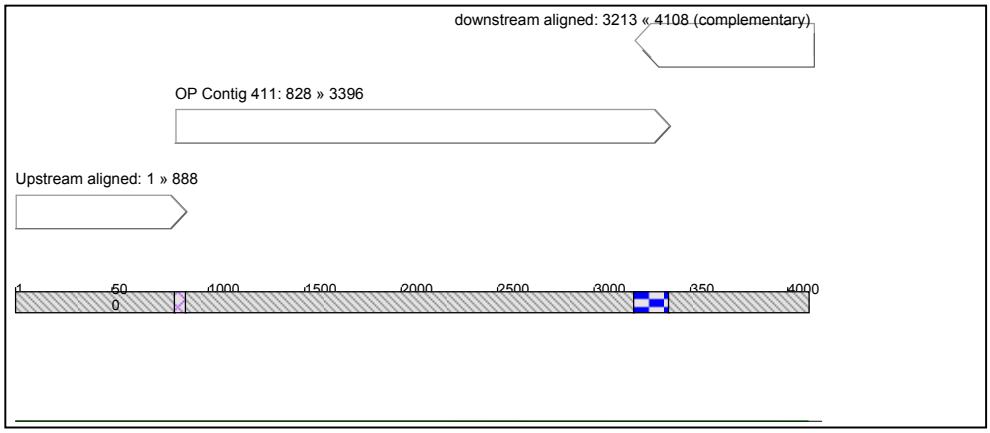


Figure 5.4.4C Alignment of all genome walking amplicons to form HSP.

Alignment of all sequenced fragments resulted in a 4108 bp sequence shown in Figure 5.4.4D. The predicted CAP site was highlighted in yellow and a putative 3' polyadenylation site found in other filamentous fungi (Gurr *et al.*, 1987), was boxed in bold and written in italics

CGACGGCCCGGGCTGGTATCCTGCCCGAGCTGTTCTGGTTCCGTCAGCAGTGTGCCGCGGAGGGCGTC
AACATTCCCTTCTTCTTCCATGCGGGCGAGTGTCTGGGCGACGGTGACGACACGGACAACAACGTCTT
TGATGCCGGGCTGCTGGGGACGCGGCGCATTGGTCACGGCTTCAGTCTGTACAAGCACCCGCTGGTGA
TTGACATGGTCAAGGACAAGAAGATTTTGATCGAGTCCTGCCCGATCAGCAACGAAGTGCTGCGTCTC
TGTGGCAGTATCCTGTCGCACCCTCTGCCGGCTCTCATTGCACGCGGTGTGCCATGTGGCCTGAACAA
CGACGACCCGGCCATCCTGGGCCAGGACACCGAGGGCCTGACGCACGACTTCTGGCAGGCCCTGCAGG
GCTGGGAGAACCTGGGCCTGGCCGGCCTGGGCTCGCTGGCCGAGAACAGCGTGCGGTGGTCTGCTTTC
GAAGACGAGACCAACGAGGAGTGGATGCAGCACATCAAGGAGGCGTCGCTGGGCAAGAGCATCAAGGC
ACGGCACCTCAAGGAATGGAGCATCCAGTGGGAGAAGTTCTGCCTGTGGATCGTGACGGAGTACGGCG
AGAAGTACGGCGATGAGGAGTAGGTATATACGGAGCCCAGGTATTTTCAGGTTTGATATTGTTCTTTC
TTATTTGGTATATTATATTATTTTTCCTTTTTGGTTCTACTCAACTTTATACAACACATTGGTTATGG
AGGTCAGGGCATCACGACATTCGCATTCAACATCAGGGTGCAATTCAACATTAGGGATTTTATCAAAT
TTAACATGAAAAGGGATTTGGAAATTTAGCTAGCTATGAAATAAGATATGTCCACTGCAAATTGAGCG
GCTGTGGTGTTCCTCCTGGTGGTATCATTTTGGGTACAAATAAACAAATACATCAATCGCCTTTTTGA
CCGCAGGCTAGTTTACTCCCGGCGGGAGTCCTGGGCCTCGGCCTGCTTCTGGATATCGTGCTCAGCGC
CGCGCTGCTGGGTCTTGCCAGTATCGTGCTGCTTCTGGTACTCGGCCTCGGCCTGGGGGTTGCCTGTC
AGGTTGGCAAAGCCACTGCCGACAGCGCCCTTGACGCGGTCGGATGCACCCGACACAAAGTCGGTGAC
CTGGCCACGGGCCTCCTGCTCCTGGCCGGACAGGTTCTGCTCACGGCCGGCCTGCTTGAGAGACGCGG
AGCCGACAAGACCGCCGAGGGTCTCCTTGGCAGAGCCGACGGTCTGGTTCCAGTTGCCAGCCGAGCGG
TTGGGGTCATCGCGGGTGACGGCACCGGACGAGGATGCCGTGAAGCCAGGGCCCTTGATGGTAGCGTG
TGATGCGTCGTACTCCTGCTGGGCGTTGGCCTTCTTGGCCTCGCCAGCGCCCTCGTGGCCAGTGCTGC
CGGTTAGGCTGCCAATAGCGCTCTGGACGGTGCCGACGGCGGAGTCGATGTACGACTTGGCAGTCGAG
GGCTGGTTGGAGGTGTTCTCGGTAGACATGATGGTGACTGGGAGGTTGGTTGTTTTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTACACACCACACACATTGGCTTCACAGCTTCAGTTCCACTGCATCTCTACGAG
CAAGATTAAAAAACACTTGGCTTATAGCTTCACACACTCACACACATCACAATGTCTGCCATTGCATC
AAAAGCCGTTGGACCCCTCGTCTCTCTCGGCGCAGCGTCCGCCGTCGTCGGCTACGTCTACACGCAAC
TGTACGCCGAGTCGGCCACCATGGACCGCTACTTCTCCCAATACAACACGCCCCAGAGCGAAGCCTCG
CGCCAGCGCGTCTTTGACGGTGCCATTGAAGACCCCCGGAGAAACATCCTCAACTTTCTCAGCTGGAA
GTGATGTGAATTAGAGGCTTTGTTTTCTCTCTCTCTCTCTCTCTCTATCGTTTTTTTTTTTTTTGCT
TGTTTTCAATAATGCGCTGGGCCAAATACCAGCAGCCTTGCCGAGGCTGCCTAATTAGCATTTTCTAT
TTCCAATTGGATAACCTATTGATATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAC
AAGAATTCAAGAACAATCAAACACTCGGCTTTCAAAGTCAAAACAACATCTCTCACATCTTCACCATT
CCTCTCTTTTCAAACAATACTCCCCCAACAACAACATTCAAAATGCCTTTCTTCCAGCCCGCATTCAT
M P F F Q P A F I
CGCCCCGGTTCAGGCTGATGCCAGCGCCTTTGCCAACCTCTTCCAGCTGGTCAACGACATTGACGG
A P V Q A D A S A F A N L F Q L V N D I D G
CTACCAGCGCGAGGCTTCCTGTGAATCTGCTTGCAAGCCCAAGCAGCAGCAGCAGCGTCAGCAGGC
Y Q R E A S C E S A C K P K Q Q Q Q R Q Q A
CCAGCCCAAGGCTCAGCCCAAGCCCCAGCAGAAGTCCCAGCAATCCCAGCAGGCTCAGGCCAAGCC
Q P K A Q P K P Q Q K S Q Q S Q Q A Q A K P
CCAGCAGACCCAGCAGAAGCGCGCGTCTTTCCGCCACCCATTCGAGGAGTTCTTCAACAGCCTCCA
Q Q T Q Q K R A S F R H P F E E F F N S L Q
GCAGGCTGCTGCCGAGACCAGCCGCGCTGCTTCTCTCCCCACGTTCCACCCGCGCTTCGATGTGCG
Q A A A E T S R A A S L P T F H P R F D V R
CGAGACTGAGAACACCTACGAGCTCCACGGCGAGCTGGCCGGTGTCGACCGCAACAACATCTCGCT
E T E N T Y E L H G E L A G V D R N N I S L
CGAGTTCACCGAGCCCCAGACTCTGGTCGTGAGCGGCAAGGTCGAGCGCAACTACCAGTCCCAGAA
E F T E P Q T L V V S G K V E R N Y Q S Q K
GAGCACTCCTCTCCAGGCCGTGCCTGACGTCGAGAACACGCCTGAGGCGATTGCCGACACCCACTC
S T P L Q A V P D V E N T P E A I A D T H S
CGAGCCCGACCACCATGACGACAACGCCTCTACCAGCGGCGCCCGCACCCCCGTCGACGAGGAGTT
E P D H H D D N A S T S G A R T P V D E E F
CACAGAGATCAACTCGCCGCGCTCGCCTTCGCCGGCCCGCAGCCACCGTGCCACTGTGACTGACGA
T E I N S P R S P S P A R S H R A T V T D E
GGAGACCGAGGAGGCTCTGGAGCGCGGCCTGGCCATCGCACAGGAGGAGGCGAAGAAGGCCGTGGT
E T E E A L E R G L A I A Q E E A K K A V V
GGCTACTCCTGCCCCTGTCGCGCCGGCTCCCGCCCAGGCGCCCTCTGCCCCTGCTGAGCGCTACTG
A T P A P V A P A P A Q A P S A P A E R Y W
GTACCAGGAGCGCGCGGTGGGCCAGTTCAAGCGGGTGTTCGACTTCCCCGTGCCTGTCGACGAGGC

```
 Y  Q  E  R  A  V  G  Q  F  K  R  V  F  D  F  P  V  P  V  D  E  A
CAATGTGCGCGCCAGCCTTGACAACGGCATCCTGAGCGTGTCTGTCCCCAAGGTCAAGCGCGAGGT
 N  V  R  A  S  L  D  N  G  I  L  S  V  S  V  P  K  V  K  R  E  V

TCGCCGCATTGTCGTCTTTTAAGCTTACATTGTGCTTTGTGCTTGATACTTGCTCGACAGCTTGAC
  R  R  I  V  V  F  ter
CACCACTTGACTTTCTTTACCACTTCATCGCAAGGCGTTTTTAGGATTGATTTTCGCTTGTTTTCACT
TTATGGAATTTGCTTGTTATCGCTTTCTGGAATTCACAGCTTGAGGAGAATTGCATAATATGATAGTT
TCTAAATCGATATTTTAAATAATCAATATTGCATGAAGTGATCAGATAAATAGACGCCATTAAAAATT
AGACAAAGTCCTAGAGCTACTATTCTTTTAAATTCTTTTAAATTCTTTTAAATATACAGCTGTTTTAC
CAACATAATGTTTCTCCGAGAAAGCACATCGGTGCATGTTAAGACCGGCCCCGAAATCATCCGCTATT
GAAATGGCCCGTAGGACTTTAGGAGCCCGTATTAAGCAGCATCCAGCACTACTATAAATACGTGATTG
CGTGCCCGTGTCCATCTCCCATCAGCGATAGCTCCAGACATAAATAACCACAACGACGAGCACCCGAC
TGACCAGCCAAGACAAACCCGCAATGTCCCATCTTGCAAAGGTCGCCGACAGCGGCTTTCACGATGCC
AGTGCCTATGATGCGCACCGGCCATCATACCGCCCCGAGGTCGTCAGTGCATTCCTTGGGCGGCTCAA
CGTCCTCGGGAACACAGCAAGCGGACCTCTGCAGATTGTCGAGGTCGGCGCTGGCACCGGCAAGTTTA
CCGAGATCCTGTCCCGGAAGGCCGCCGAGGATAACGCCTCGCTCAATGTCATCGCTGTCGAGCCGCAC
GACCAGATGCGCGCACAGCTGACGGCCAAGAACCTGCCGCACGTCACCGTCGTCGACGGCCACGGCGC
CGATCTCCGCAGTGTCGGGGACAGCACCGCCGATGCCGTCGTCATGCCCAGACCAGCC
```

Figure 5.4.4D HSP nucleotide and predicted amino acid sequence

Analysis of HSP using NCBI's ORFinder program identified 11 possible ORFs greater than 300 bp in length. One of these ORFs, with a predicted start site at 2219 bp and a predicted stop site at 3190, was also predicted by OrfPredictor as being the most likely ORF. The predicted ORFs were analysed using BLASTp to identify homologous proteins. ORFs that resulted in a significant alignment with proteins in the database were listed in Table 5.4.4. HSP ORFs producing a significant alignment were also analysed against the Pfam database to identify possible conserved domains as shown in Table 5.4.4B.

ORF	Length	Frame	Accession	BLASTp -Best Hit	*E* - value	Score
2219..3190	324	+2	XP_963774	30 kDa heat shock protein Neurospora crassa	8.00E-34	147
965..1525	187	-1	XP_001910788	unnamed protein product Podospora anserina	1.00E-61	240
210..635	142	+3	XP_959185	hypothetical protein NCU04642 Neurospora crassa	2.00E-59	232
3710..4108	133	+2	XP_380633	hypothetical protein FG00457.1 Gibberella zea	1.00E-18	97.4

Table 5.4.4A List of HSP predicted ORFs showing significant homology by BLASTp analysis to proteins in the non redundant (nr) database. ORF indicates the predicted ORF start site followed by the predicted ORF stop site.

ORF	Pfam Conserved Domains	Start/ Stop	Bit Score	*E*-value
2219..3190	HSP20	117..154	5.4	0.79
	HSP20	240..323	0.4	0.0021
965..1525	CsbD	78..130	29.8	9.70E-06
210..635	A_deaminase	5..100	64.7	2.20E-01
3710..4108	Methyltransf_11	59..129	18.7	0.00024

Table 5.4.4B List of conserved domains in significant ORFs in HSP identified by Pfam analysis. Start/Stop represents the predicted start followed by stop site in amino acids of the conserved domain.

Interestingly, none of the significant ORFs identified in HSP overlapped. The largest ORF predicted with a length of 324 amino acids (aa) aligned most significantly with a

N. crassa heat shock protein. Analysis using Pfam also indicated two putative HSP20 conserved domains. Prokaryotes and eukaryotes respond to environmental stress such as heat shock by inducing the synthesis of proteins collectively termed heat-shock proteins. HSP20 heat shock proteins have an average molecular weight of 20 Kd and are thought to act as chaperones to protect other proteins from heat-induced denaturation and aggregation.

Within the same frame, a 133 aa unknown putative gene was identified. This gene contained a predicted methyltransferase conserved domain. Methyl transfer from the ubiquitous S-adenosyl-L-methionine (SAM) to either nitrogen, oxygen or carbon atoms occurs in many organisms including bacteria, plants, mammals as well as fungi (Dekkers *et al.*, 2006). The reaction is catalyzed by methyltransferases (Mtases) and modifies DNA, RNA, proteins and small molecules, for regulatory purposes. DNA methylation in fungi is thought to be involved in state-specific control of gene expression. It has been well studied in *N. crassa* (Selker *et al.*, 2003). Within the *N. crassa* genome half of the methylation occurs in repeated DNA including centromeric DNA and transposon relics that were subject to RIP mutation.

5.4.5 Identification of Transcriptional Elements

Transcriptional elements upstream of an ORF identified as a putative heat shock protein in HSP were investigated *in silico*. Transcriptional elements were identified based on conserved motifs identified from literature and using TESS, a computational tool used to search for transcription factor binding sites in eukaryotes. A number of possible transcriptional elements and potential factor binding sites were identified (Table 5.4.5 and Figure 5.4.5). Within Table 5.4.5 the positions indicated refer to the position from the start of the nucleotide sequence Highlighted in yellow, the cap site defined as +1, is a transcription initiation sequence or start point at which the transcription process actually starts. A predicted TATA box was identified (bordered and in red text). In many eukaryotes it tends to be surrounded by GC rich sequences. In HSP this does not appear to be the case with no GC boxes with a conserved motif of GGGCGG identified. .

Two possible CCAAT binding complex sites were identified. CCAAT sequences are found in the 5' regions or approximately 30% of eukaryotic genes. They are regulatory elements found in the promoters of genes encoding plant cell wall degrading enzymes and transcriptional activators or factors modulating expression. Mutation of the CCAAT motif in promoters or deletions of the subunits has been shown to reduce the expression level of a gene. Reduction was shown to occur either at the basal expression level, or in the response to specific stimulatory signal, indicating that the CCAAT motif cooperates with other specific elements to induce transcription (Aro *et al*., 2005). Two possible E2Fs were identified (highlighted in blue in Figure 5.4.5). E2Fs as transcription factors bind to the TTTCGCGC consensus binding site in the target promoter sequence. E2F is a group of genes that codes for a family of transcription factors in higher eukaryotes. Three of them are activators: E2F1,2 and E2F3a. Six others act as suppressors: E2F3b, E2F4-8. In mammalian cells all of them have demonstrated involvement in cell cycle regulation and synthesis. Two possible *PacC* binding sites were identified (highlighted in green in Figure 5.4.5). *PacC* is a zinc finger transcription factor demonstrated in *A. nidulans* to act as an activator for alkaline-expressed genes and prevent expression of acid-expressed genes. *PacC* contains a DNA-binding domain with three Cys_2His_2 zinc fingers that bind to promoter sites containing the core hexanucleotide sequence 5'GCCARG (Aro *et al*., 2005). Six possible *Cre* binding sites were identified (highlighted in pink in Figure 5.4.5). 5'SYGGRG consensus motifs are present, either singular or closely spaced in duplicate in all functional CREI/CreA binding sites characterised so far. CreA/CREI is a Cys_2His_2 type transcription factor that mediates glucose repression in many ascomycetous fungi (Aro *et al*., 2005). Five possible *AreA* binding sites were identified (highlighted in grey in Figure 5.4.5). Gata factors are a class of eukaryotic transcriptional activators or repressors characterised by a highly conserved DNA binding motif comprising a Cys(4) zinc finger followed by a basic domain. This motif mediates binding to DNA elements that have the core sequence 5'GATA. In filamentous fungi they regulate nitrogen catabolic enzymes during the use of secondary nitrogen sources such as nitrate, nitrite and purines (Aro *et al*., 2005).

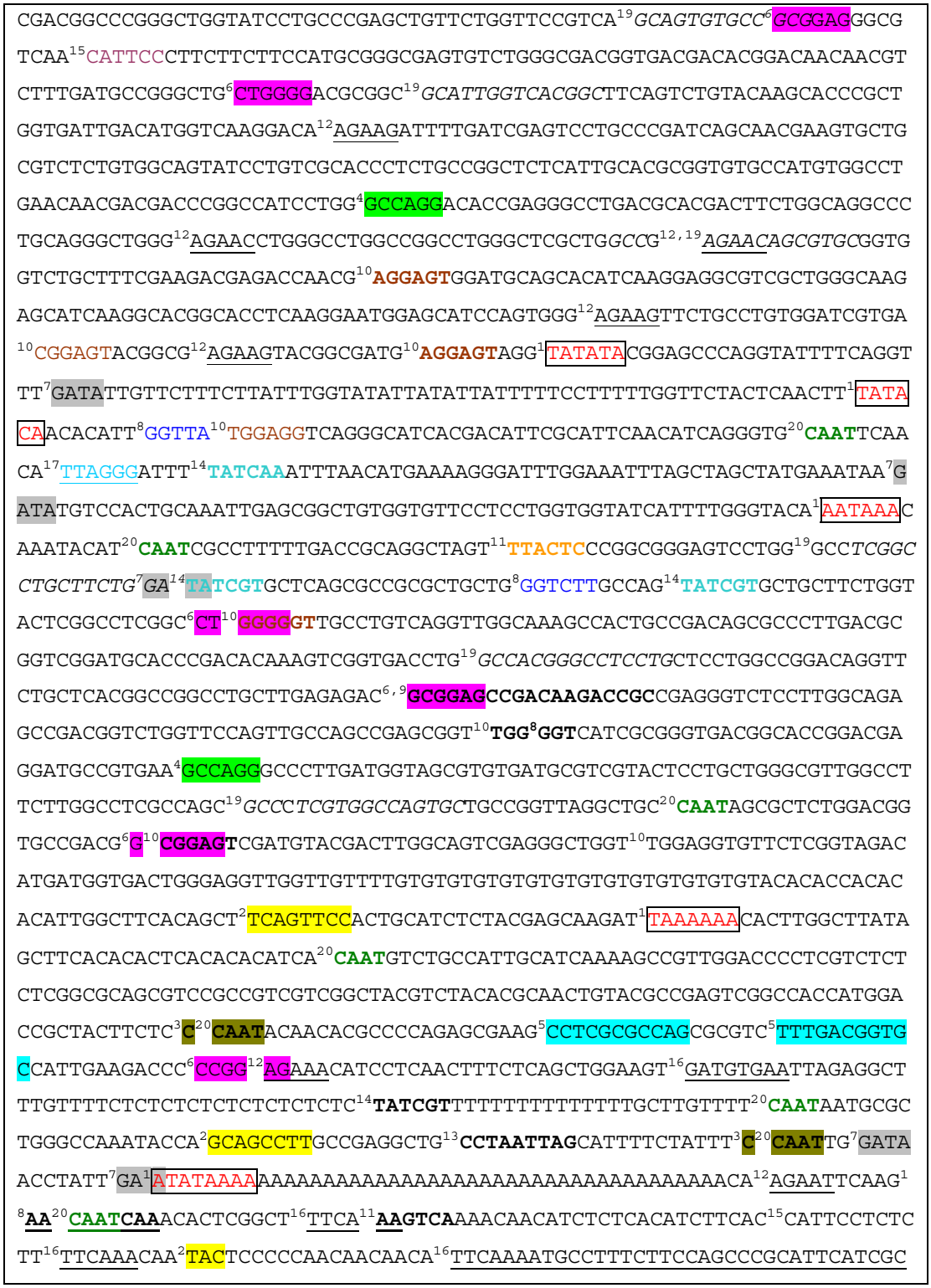


Figure 5.4.5 Putative transcriptional factors identified in the 5' region of HSP sequence. The start of the predicted ORF is underlined in black. A description of the identified transcription factors in given in Table 5.4.5.

	Motif(s)	Description	No.	Position	TESS Factor
1	TATATA TATAAA TATACA AATAAA TATATAA TAAAAAA	TATA box. A conserved motif found in a number of eukaryotes including *S. cerevisiae* and *S. pombe*. Activator of all three polymerases in *S. cerevisiae*	6	637, 728, 923, 2063, 1639, 2063	TFIID, TBP
2	TAC GCAGCCTT TCAGTTCC	Cap signal for transcription initiation	3	1609, 2004, 2194	CAP
3	CCAAT	CCAAT binding complex. Binds regulatory elements found in the promoters of genes encoding plant cell wall degrading enzymes and transcriptional activators or factors modulating expression	2	1806, 2043	CP1
4	GCCARG	PacC. A zinc finger transcription factor that acts as an activator for alkaline expressed genes and prevent expression of acid-expressed genes	2	361, 1337	
5	TTTGACGG-TGC CCTCGCGC-CAG	E2F initiation factor. It acts as both and activator and suppressor involved in cell cycle regulation and synthesis	2	1832, 1849	
6	SYGGRG	Cre transcription factor. It mediates glucose repression	6	57, 150, 1073, 1221, 1467, 1872	MIG1
7	GATA	AreA transcription factor. Regulates nitrogen catabolic enzymes during the use of secondary nitrogen sources	5	667, 861, 1004, 2050, 2061	
8	GGTCAT GGTCTT GGTTAT	Transcriptional element identified in *S. cerevisiae* associated with HAP1. Has been linked with regulation of yeast iso-1-cytochrome C expression	3	741, 1031, 1298	RC2
9	**GCGGANN-NNNNNNNt-CCGC**	Transcriptional element in *S. cerevisiae* that Activates the CHA1 gene for L-serine (L-threonine) deaminase responsible for the utilisation of serine/threonine as nitrogen sources	1	1221	T02848 CHA4
10	**CGGAGT, AGGAGT, GGGGGT, TGGAGG, TGGGGT, TGGAGG**	Positive regulator of peroxisomal protein genes identified in *S. cerevisiae*. It is required in yeast for ADH2 activation and glycerol metabolism	8	493, 601, 628, 746, 1075, 1294, 1468, 1504	T00011 ADR1
11	**TTACTC, AAGTCA**	Activator of genes involved in protein and purine biosynthesis identified in *S. cerevisiae*	2	965, 2143	T00321 GCN4
12	AGAAG AGAAT AGAAC AGAAA	An activator identified in *S. cerevisiae* and *Kluyveromyces lactis*. It is essential for growth, mediating expression of heat-shock genes in response to elevated temperatures or to other stress conditions such as glucose starvation.	7	223, 415, 451,577, 613, 1876, 2110	T00385 HSF1
13	**CCTAATT-AG**	An activator of DNA replication in *S. cerevisiae*. Presumably it regulates genes involved in cell cycle control in synthesis of cell wall/membrane structures, cell metabolism, and heat-shock-inducible secreted glycoprotein.	1	2022	T00500 MCM1
14	**TATCGT, TATCAA**	Activator of nitrogen-regulated genes; NIT2 can partially complement for lack of AREA in Aspergillus nidulans. NIT2 is also major nitrogen regulatory gene in N. crassa. It encodes a protein with a putative zinc finger DNA-binding domain	4	809, 1006, 1042, 1950	T00627 NIT2

	Motif(s)	Description	No.	Position	TESS Factor
15	CATTCC	TCS-like and PRE- like elements. In yeast they have been shown to play a major role in the regulation of genes during the filamentation and invasive response through binding of TEC1 and STE12 transcription factors. TCS-like elements in *A. nidulans* are necessary for spore differentiation and binding to developmentally regulated genes	2	71, 2173	T01085 abaA
16	TTCAAA, GATGTGA-AaT	Transcriptional element identified in *S. cerevisiae*. It is required for activation of phospholipid synthesis.	4	1907, 2139, 2185, 2213	
17	TTAGGG	A conserved motif identified in *S. cerevisiae*	1	799	T01246 TBF1
18	AACAATC-AA	A transcription factor identified in humans that has homologues (MATa1) in yeast	1	2120	T01481 Pbx-1a
19	*GCANNN-NNNNNG-GC, TCGNNN-NNNNNN-NGGA, GCCNNNN-NNNNNNN-TGC*	Transcriptional factor identified in *A. nidulans* required for acetate induction of acetamidase and the acetate utilisation enzymes acetyl-CoA synthase, isocitrate lyase and malate synthase.	10	652, 769, 69, 162, 56, 1271, 314, 285, 1188, 19	T02841 FACB
20	**CAAT**	CAAT box	8	796, 944, 1448, 1688, 1813, 2053, 2131, 2200	

Table 5.4.5 Summary of putative transcriptional elements in the 5' region of the predicted HSP gene in *O. piliferum*.

5.4.6 Profile of expression at different stages of growth using RT-PCR

The relative levels of HSP expression at different stages of growth in liquid, shaken cultures, and in different growth forms, were determined using RT-PCR.

5.4.6.1 Culture Growth and RNA Isolation

Nineteen 400 ml conical flasks, each containing 100 ml of YM media were inoculated with 2 x 10^5 *O. piliferum* strain 97 blastospores (refer to Sections 2.1.7 and 2.1.8 for details of the methodology). Duplicate cultures were destructively sampled every 24 hours for 8 days with the exception of 168 hours when no sample was taken for either culture. Given the inherent variability between fungal cultures, the samples were analysed to determine blastospore count (refer to Section 2.1.3), biomass (refer to Section 2.1.4) and viability (refer to Section 2.1.5) to establish if cultures were

representative of previously determined growth curves for both species (Figure 5.4.6.1A).

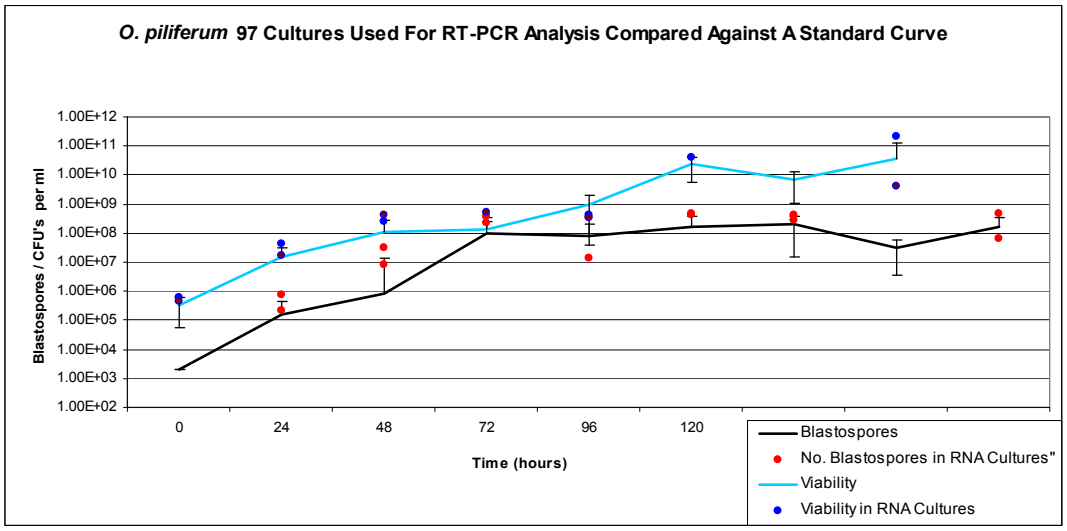


Figure 5.4.6.1A Graph showing the number of blastospores and CFUs in *O. piliferum* 97 determined in duplicate cultures used for RT-PCR analysis, harvested at different time point compared against standard curves.

In terms of blastospore count and viability, the levels detected in cultures used for RT-PCR analysis were comparable to those numbers used to construct the standard curve.

RNA was spectrophotometrically quantified (Table 5.4.6.1) and assessed visually for quality using agarose gel electrophoresis as described in Section 5.3.6.1 and as observed in Figure 5.3.6.1B).

***O. piliferum* Cultures**			
Sample	**Time Harvested (hours)**	**Concentration (ng/µl)**	**260/280**
Mycelium	24	2744	2.02
Spore	24	146	1.71
Mycelium	48	3070	2.03
Spore	48	148	1.87
Mycelium	72	1755	2.03
Spore	72	565	2.00
Mycelium	96	1421	2.01
Spore	96	537	1.93
Mycelium	120	975	1.97
Spore	120	405	1.89
Mycelium	144	463	1.92
Spore	144	946	1.97
Mycelium	192	2089	2.00
Spore	192	162	1.79

Table 5.4.6.1 Spectrophotometric analysis of *O. piliferum* RNA isolations used for RT-PCR analysis.

5.4.6.2 Construction of cDNA

cDNA was constructed from *O. piliferum* strain 97 RNA samples as described in Sections 5.3.6.2 and 2.2.5.

5.4.6.3 RT-PCR Amplification

RT-PCR was conducted using HSP gene specific primers HSF and HSR. The primers were designed with a melting temperature between 50 and 65 ºC (Table 5.4.6.2). The RT-PCR assay was designed to produce an amplicon 106 bp in length.

Gene	Primer	Nucleotide Sequence	Tm (NN) ºC	Amplicon Size
Heat Shock	HSF	5'- GAGTTCTTCAACAGCCTCCAG -3'	61.4	106 bp
Heat Shock	HSR	5'- CGTAGGTGTTCTCAGTCTCG -3'	58.4	

Table 5.4.6.2 Forward and reverse oligonucleotides designed to amplify a 106 bp fragment of HSP.

RT-PCR reactions were run according to the method described in Sections 2.6.5 and 2.6.6. Dissociation curve analysis was performed after every reaction to identify the presence of any contaminating sequence and/or non specific products. RT-PCR product was also analysed by electrophoresis on a 0.7% agarose gel to confirm amplicon size and purity.

5.4.6.4 Determination of RT-PCR Efficiency

The efficiency of RT-PCR amplification using *O. piliferum* HSP gene specific primers was determined using the methods described in Section 5.3.6.4.

The amplification curves of the HSP serial dilutions are shown in Figures 5.4.6.4A and 5.4.6.4B. C_T was measured at a threshold 0.4 for all HSP RT-PCR reactions.

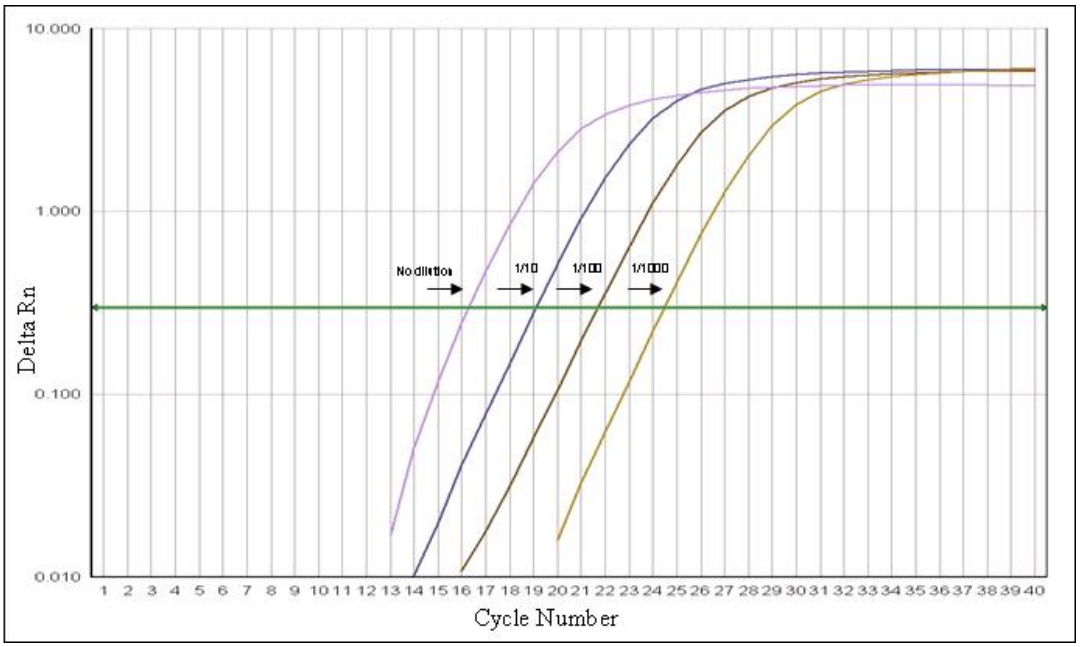


Figure 5.4.6.4A Amplification plot showing Delta Rn against Cycle Number for representative RT-PCR analysis of the HSP gene in *O. piliferum*. The cDNA template consisted of serial dilutions constructed from mycelium cultivated for 96 hours.

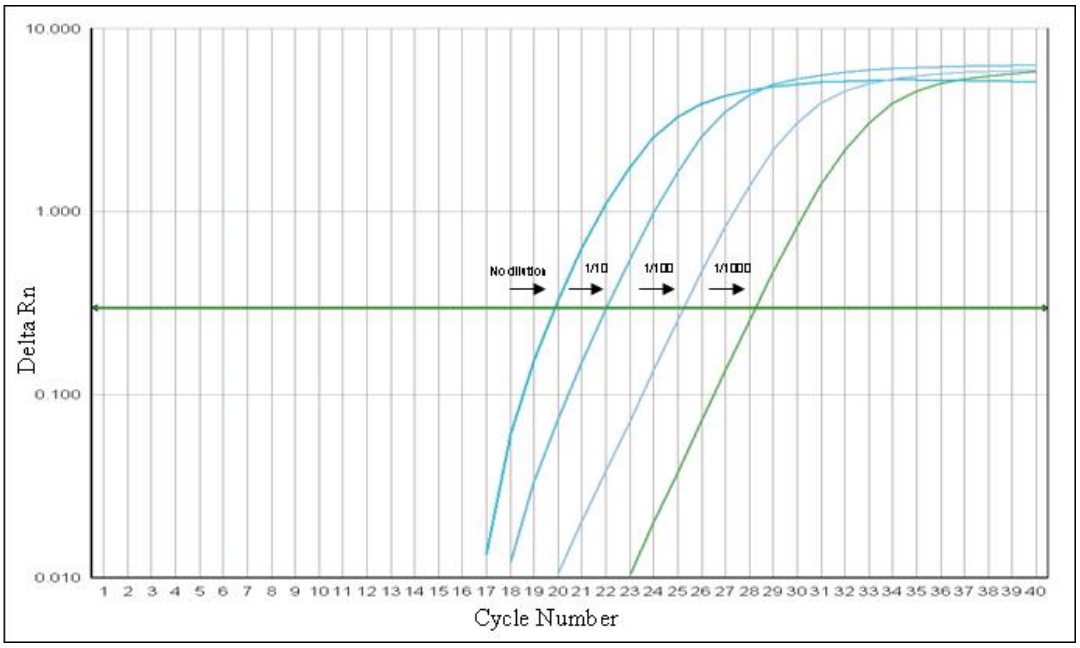


Figure 5.4.6.4B Amplification plot showing Delta Rn against Cycle Number for representative RT-PCR analysis of the HSP gene in *O. piliferum*. The cDNA template consisted of serial dilutions constructed from blastospores cultivated for 96 hours.

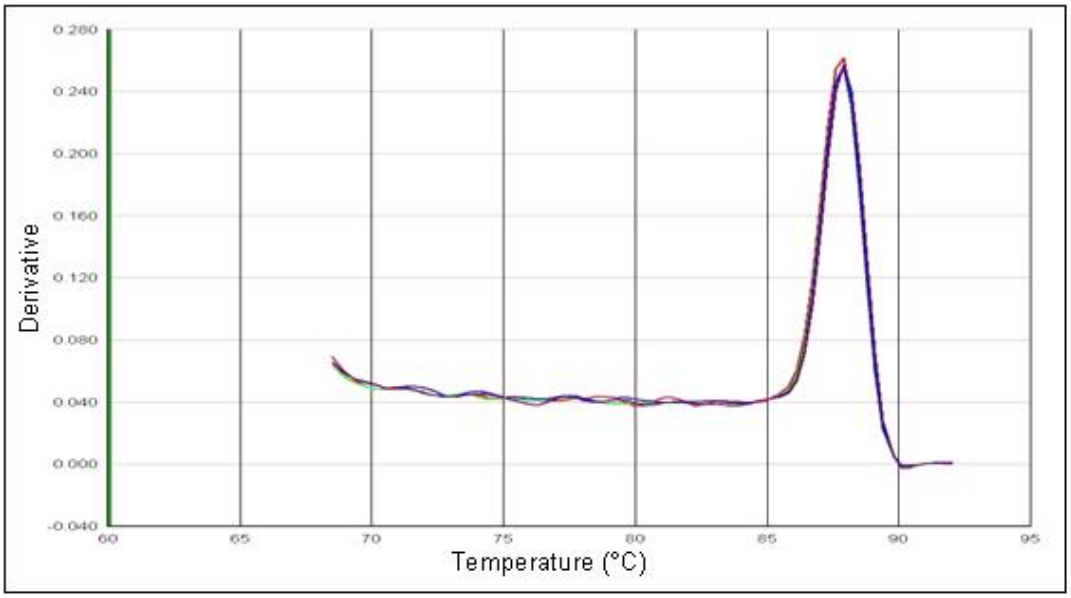


Figure 5.4.6.4C Dissociation curve showing Derivative against Temperature for representative RT-PCR analysis of HSP gene in *O. piliferum*. The cDNA template consisted of serial dilutions constructed from mycelium cultivated for 96 hours.

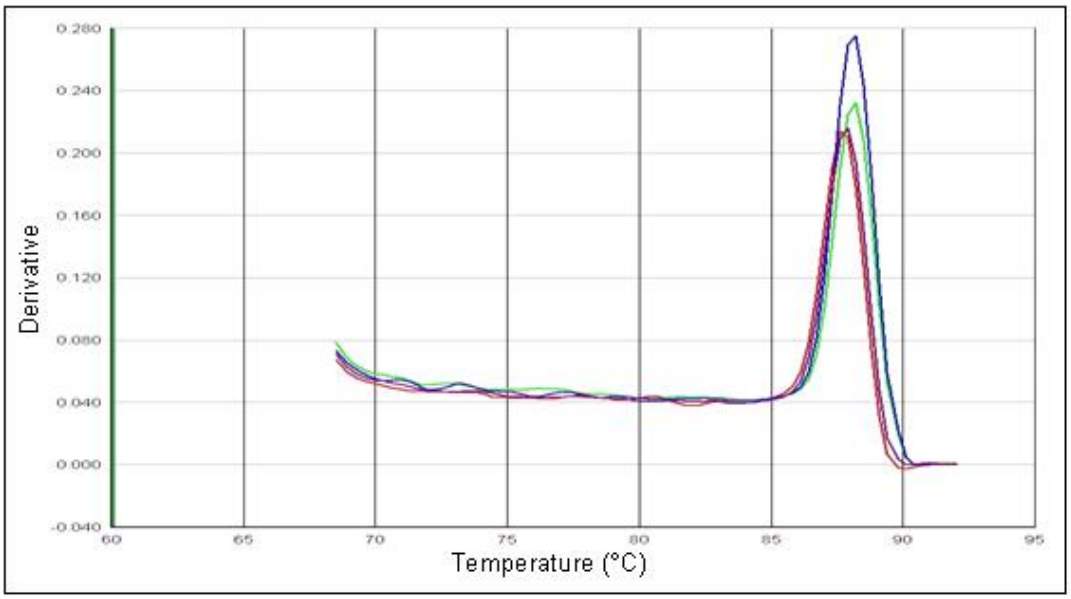


Figure 5.4.6.4D Dissociation curve showing Derivative against Temperature for representative RT-PCR analysis of HSP gene in *O. piliferum*. The cDNA template consisted of serial dilutions constructed from blastospores cultivated for 96 hours.

Dissociation curve analysis (as shown in Figures 5.4.6.4C and 5.4.6.4D) of sequence verified HSP RT-PCR product had a melting temperature of approximately 88ºC. The presence of a contaminating sequence was indicated in some of the HSP RT-PCR reactions. An example of this is given in Figure 5.4.6.4E. Samples containing the contaminating sequence were easily detected by the presence of a peak at a melting

temperature of approximately 77ºC (Figure 5.4.6.4F) and data was easily removed from the results before further analysis.

Electrophoretic analysis of the HSP RT-PCR amplicons (Figure 5.4.6.4G) did not indicate the presence of any contaminating sequence highlighting the methods limited use when dealing with such small DNA fragments. Sequence analysis of the contaminating product failed on several occasions. Repeat RT-PCR analysis of the same sample template did not indicate the presence of any contaminant. All other factors were also kept constant, such as the master mix used. Therefore, the source and likely explanation of the contamination was not ascertained.

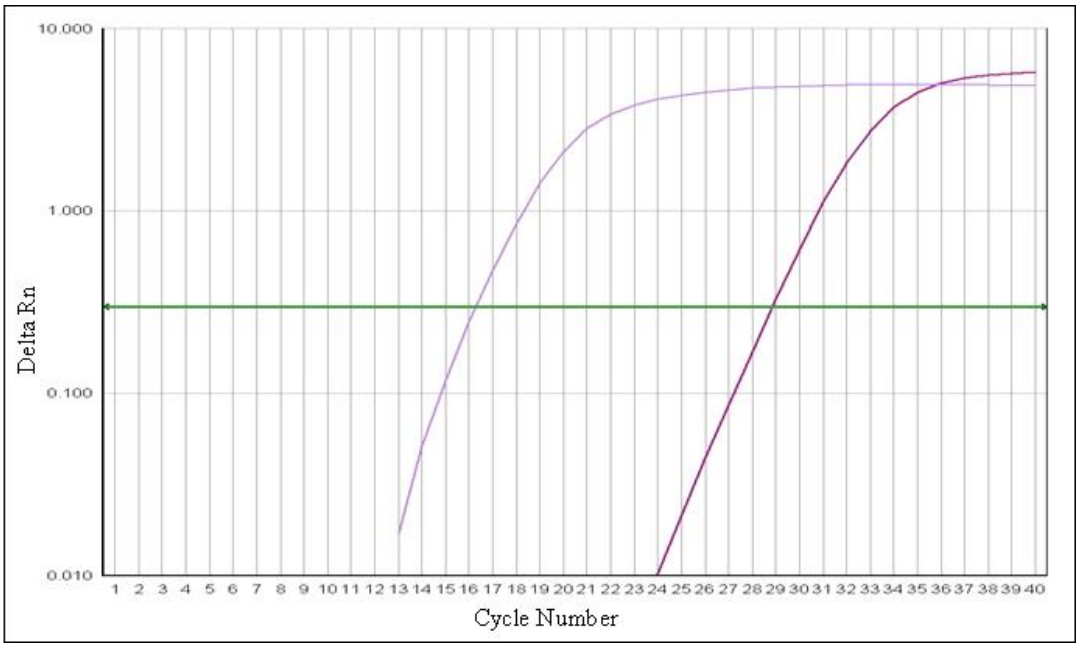


Figure 5.4.6.4E Amplification plot showing Delta Rn against Cycle Number for duplicate RT-PCR analysis of the HSP gene in *O. piliferum*. The reaction template consisted of undiluted cDNA constructed from mycelium cultivated for 96 hours.

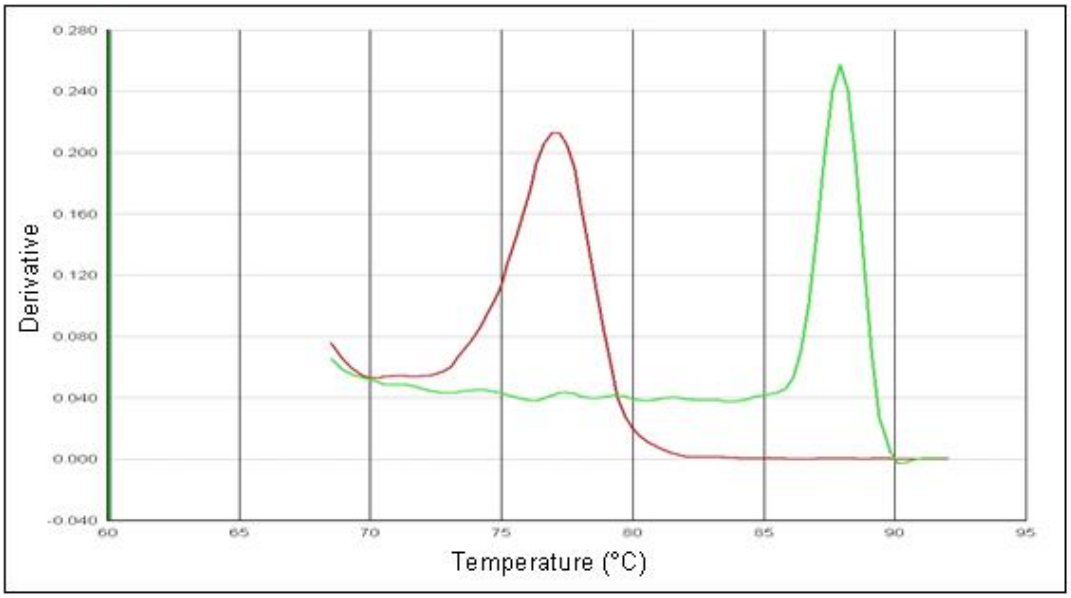


Figure 5.4.6.4F Dissociation curve showing Derivative against Temperature for duplicate RT-PCR analysis of the HSP gene in *O. piliferum*. The reaction template consisted of undiluted cDNA constructed from mycelium cultivated for 96 hours.

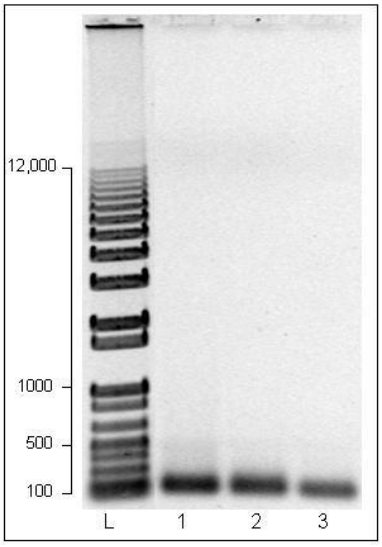


Figure 5.4.6.4G Ethidium bromide stained electrophoresis image of RT-PCR amplicons produced using HSP gene specific primers run on a 0.7% agarose gel. L = ladder, 1 = 96 hour blastospore standard concentration of cDNA template (2 µl), 2 = 96 hour blastospore 1/10 dilution of cDNA template, 3 = 96 hour mycelium standard concentration of cDNA template (2 µl).

The standard curves were constructed according to the method described in Section 5.3.6.4. RT-PCR analysis of *O. piliferum* cDNA constructed from 96 hour mycelial and 96 hour blastospore growth produced R^2 values of 0.9998 and 0.9993, respectively, indicating relatively efficient amplification (Figure 5.4.6.4H).

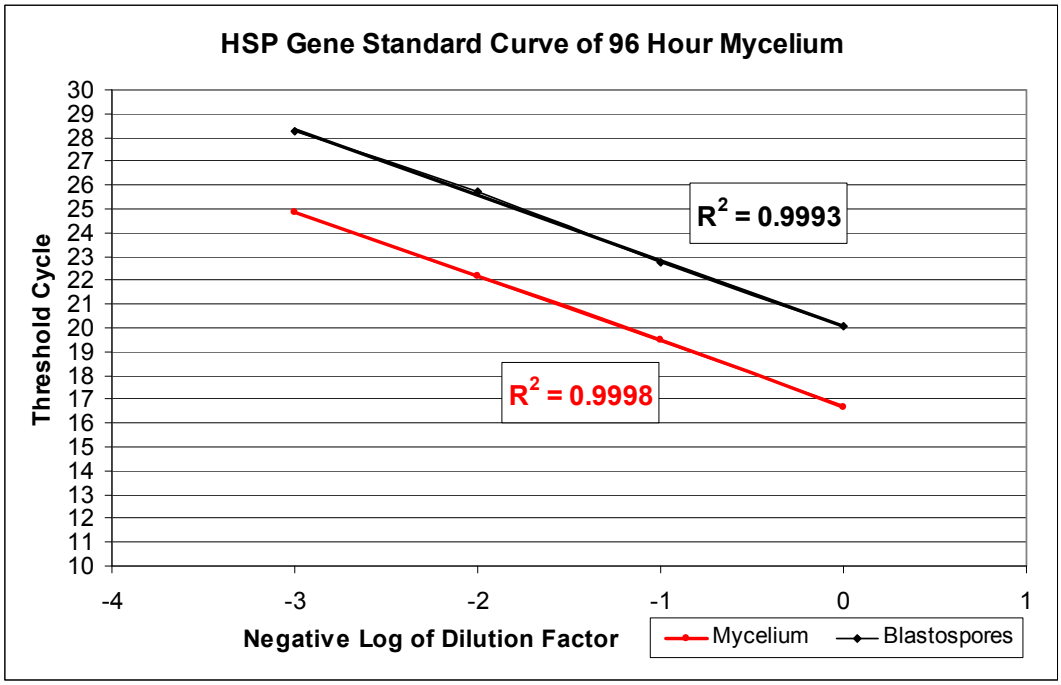


Figure 5.4.6.4H Standard curve used to calculate the efficiency of HSP RT-PCR.

The PCR efficiency was calculated as described in Section 5.3.6.4.

Where $S_{(HSP)} = -2.73$ the PCR efficiency $_{(HSP)} = 132$ %

RT-PCR reaction efficiencies of greater than 105 % may indicate pupating error in the serial dilutions or co-amplification of non-specific products such as primer dimers. Also, the presence of an inhibitor can result in the increase of PCR efficiency. Given that the R^2 values for HSP RT-PCR amplification were so good, such a high PCR efficiency is surprising. Possibly the presence of the contaminant identified in some of the samples has influenced the results somewhat, or that the very high concentrations of cDNA used in the PCR reactions inhibited some of the reactions.

In general, the data produced from samples identified as not containing the contaminant were highly reproducible and it was interpreted that if inhibition due to high concentrations of cDNA was the cause of the high calculated efficiency, this inhibition would be applied somewhat universally to all samples.

5.4.6.5 HSP Expression In Different Somatic States Throughout Growth

The relative levels of expression in the mycelial and blastospore somatic forms were compared in 14 cDNA samples constructed over 7 time points. All samples were run in duplicate, and in some instances triplicate, RT-PCR assays. Dissociation curve analysis of HSP mycelial and spore cDNA samples indicated the presence of a contaminating replicon in some of the samples as was seen in the standard curve analysis. This was evident by the curves produced at approximately 77ºC, as shown in Figures 5.4.6.5A and 5.4.6.5B. Data produced from samples containing the contaminant were disregarded and repeated to produce at least duplicate consistent C_T values with the correct melting temperature.

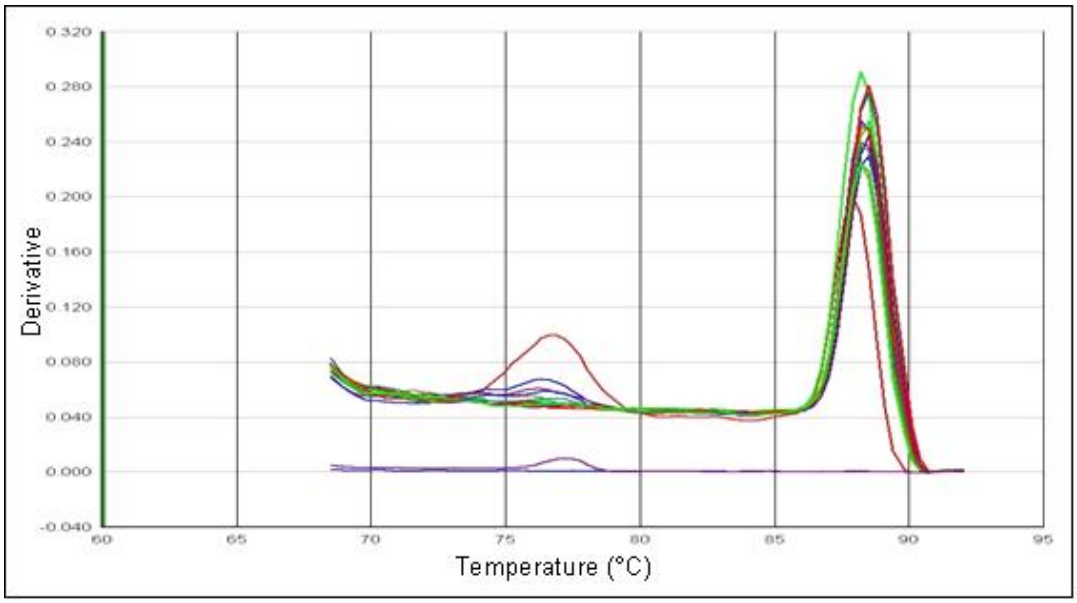


Figure 5.4.6.5A Dissociation curve showing Derivative against Temperature for duplicate HSP RT-PCR analysis of cDNA template constructed from mycelium cultivated for specific lengths of time.

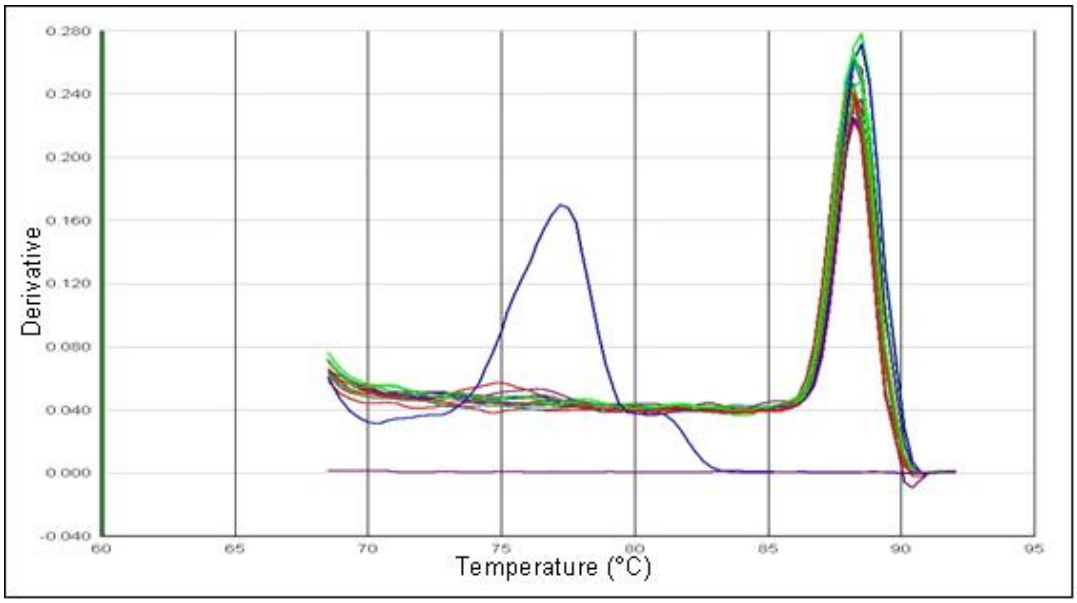


Figure 5.4.6.5B Dissociation curve showing Derivative against Temperature for duplicate HSP RT-PCR analysis of cDNA template constructed from blastospores cultivated for specific lengths of time.

The average C_T values calculated for between 2-4 PCR replicates at given time points and in both forms of growth is presented in Table 5.4.6.5. Lower C_T values were indicative of higher levels of mRNA transcript in the cDNA template. The data indicated that HSP expression was detected in both blastospore and mycelia at all stages of growth sampled and is likely to be constitutively expressed throughout growth.

	Blastospore		Mycelia	
Sample Time	Average C_T	Standard deviation	Average C_T	Standard deviation
24	20.777	0.753	20.316	0.659
48	18.074	0.355	18.984	0.517
72	18.567	0.187	17.312	0.136
96	19.872	0.393	16.973	0.315
120	19.290	0.163	16.764	0.466
144	19.132	0.222	15.898	0.440
192	18.286	0.632	17.747	0.485

Table 5.4.6.5 Average C_T value and standard deviations produced by RT-PCR analysis of HSP at different time points throughout growth in blastospore and mycelial forms.

The relative expression of HSP in mycelium compared to blastospore growth was compared as described in Section 5.3.6.5. A graph comparing the relative levels of expression based on this ratio is presented in Figure 5.4.6.5C.

Where $S_{(HSP)} = -2.73$

$Ratio_{(blastospores\ /\ mycelium)} = 2.44^{\Delta C_T}$

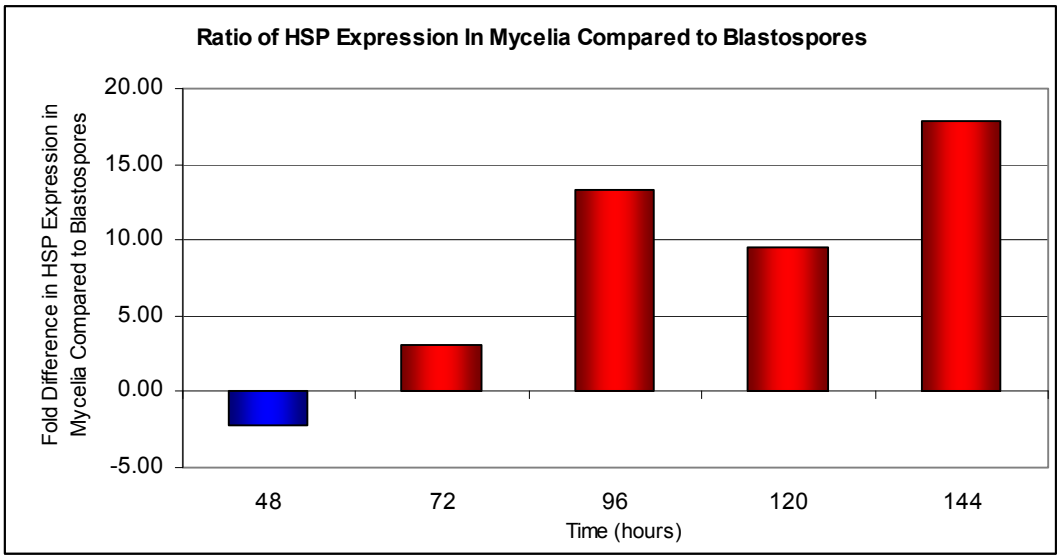


Figure 5.4.6.5C The ratio of HSP expression in mycelia compared to blastospores at different time points in *O. piliferum* strain 97 cultures.

From the data we determined that HSP expression was greater in mycelia compared to blastospores for all samples with the exception of 48 hours. Expression of HSP was up to 17-fold higher in mycelia than blastospores.

Very few qPCR-based studies on gene expression have been carried out in *Ophiostoma* species. The first paper on this was by Tadesse *et al.* (2003) who quantified the expression of the CU gene in both yeast-like and mycelial cultures of *Ophiostoma ulmi* and *O. novo ulmi* fungi. As was evident in this study, Tadesse *et al.* (2003) also reported differences in the expression of the CU gene between the two growth forms. More recently, papers describing the use of qPCR were published by Bouvet *et al*. (2008) and Tanguay *et al*. (2006, 2007) but they do not deal with the yeast/mycelium comparison. Quantitative analysis has previously been based on the use of the beta-tubulin gene as a housekeeper (Tanguay *et al.*, 2006,). This gene was also used in this investigation but was found to be inconsistently expressed in both mycelial and blastospore forms and multiple products were produced using the same primers reported in the paper.

6 Identification and Molecular Cloning of a Phospholipase A1 Gene in *Ophiostoma floccosum*

6.1 Introduction

Ophiostoma species have long been of interest in paper pulp manufacture for their ability to decrease the concentration of wood triglycerides and subsequently reduce pitch (Brush *et al.,* 1994). These fungi produce extracellular lipases that hydrolyse triglycerides into free fatty acids, resin acids (or whatever R group is attached through oxygen to the glycerol backbone) and glycerol during colonization (Gao and Breuil, 1998; Brush *et al.*, 1999; George *et al*., 1999). Although to date no purified lipase protein or sequence data has been published for *Ophiostoma floccosum*, the extracellular supernatant of *O. floccosum* was shown to contain esterase activity, as demonstrated by the ρ_1-nitrophenol assay adapted from Janssen *et al.* (1994), described in the MSc thesis of Sandra Wilcocks (2004). Given its identification in other *Ophiostoma* species (Gao and Breuil, 1998; Brush *et al.*, 1999; George *et al*., 1999), and the Wilcocks report of esterase activity, it could be extrapolated that lipase activity in *O. floccosum* would occur.

Lipases from fungi serve important roles in industry not only in pulp and paper manufacture, but also with applications ranging from yogurt and cheese fermentation to the use of recombinant lipases in laundry detergents and the production of bio-fuels. In addition to the possible biotechnological benefits of identifying novel lipases in *O. floccosum*, it is possible that abundantly transcribed lipase genes would be controlled by strong, inducible promoters. Therefore, it was expedient to the goals of this research to attempt to identify potential lipase genes within *O. floccosum*.

Lipases have been shown to have a common catalytic mechanism and structure but share little overall similarity with each other at the amino acid level (Jaeger *et al.*, 1994; Jaeger *et al.* 1999; Fojan *et al*., 2000; Pouderoyen *et al.*, 2001; Bell *et al.*, 2002; Bradner *et al*., 2003) Two relatively conserved regions including a serine active site and an oxyanion hole were identified in fungi (Herggard *et al*, 2000; Bell *et al.*, 2002).

Degenerate primers designed by Bradner *et al.*, (2003) were used to target conserved regions in lipases (see Table 2.3). These degenerate primers were used to prospect for a previously unidentified lipase in *O. floccosum*. The gene sequence of this putative lipase was then elucidated using genome walking methods and expression throughout growth in both the mycelial and blastospore forms was evaluated using RT-PCR.

6.1.1 Specific Aims and Objectives For The Identification and Molecular Cloning of a Phospholipase A1 Gene in *O. floccosum*

Three specific objectives were addressed in this part of the PhD thesis research, as follows:

1. To determine whether a lipase gene could be identified in *O. floccosum*, using PCR-based prospecting methods.

2. To elucidate the gene sequence, as well as upstream and downstream regions of potential *O. floccosum* lipase genes in order to investigate the promoter regions and identify possible transcriptional elements.

3. To map expression of any possible lipase genes throughout different stages in *Ophiostoma floccosum* yeast-like and mycelial growth, in order to contribute to the overall goal of the PhD thesis research "to further the understanding of protein transcription and transcriptional regulation in *Ophiostoma* species with particular focus on the species *O. floccosum* and *O. piliferum"* (as discussed in Section 1.3) .

6.2 Bioprospecting For Lipases in *O. floccosum* Using PCR-Based Methods

Degenerate primers designed to target two conserved regions in lipase genes, a serine active site and oxyanion hole (Bradner *et al*., 2003), were used to probe for possible lipase gene in *O. floccosum* genomic DNA, as described in Section 2.3. The degenerate primers were designed to amplify a short region of lipase gene. All

amplicons ranging between approximately 200 and 400 bp in length were considered possible targets.

Ethidium bromide fluorescence post gel electrophoresis of the amplicons revealed four bands within the approximate size range as designated by coloured arrows in Figure 6.2A, which were arbitrarily called LP1, LP2, LP3, and LP4. These bands were isolated from the agarose gel and cloned using a $TOPO_{TA}$ (Invitrogen) system (Figure 6.2B).

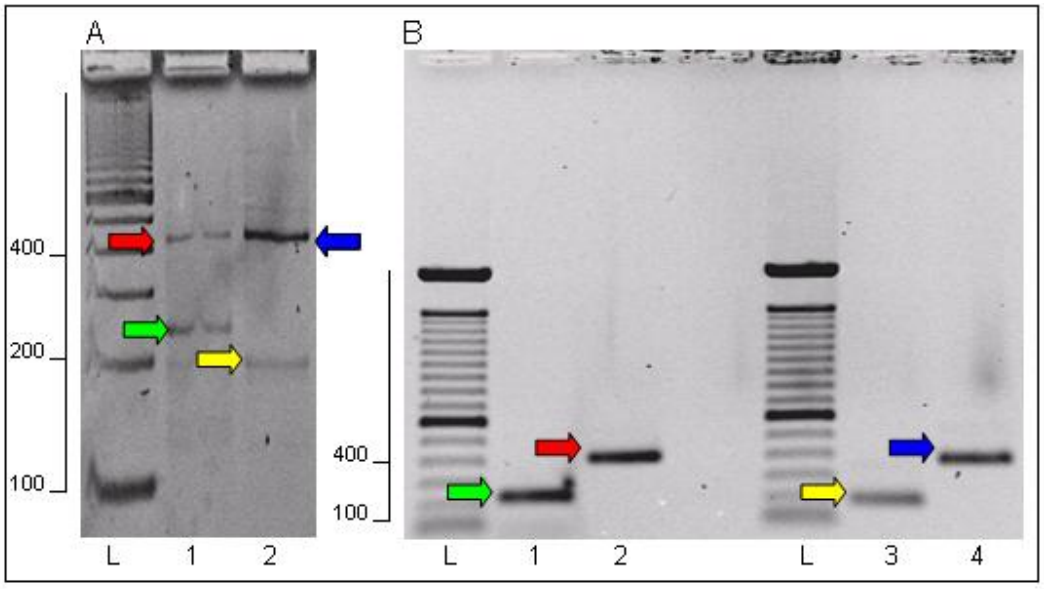


Figure 6.2 Amplicons produced from PCR using degenerate lipase specific primers. The arrows indicate amplicons of the approximate targeted size range that were isolated and cloned using $TOPO_{TA}$ (Invitrogen) system. Red = LP1, Green = LP2, Blue = LP3 and Yellow = LP4. B. PCR amplicons of cloned possible *O. floccosum* lipase genes.

Recombinant plasmids of each of the constructs with the LP fragment ligated into $TOPO_{TA}$ were then sequenced in the forward and reverse direction and analysed using NCBI BLASTx alignment. LP2 sequence was identified as having a conserved DDHD domain found in some phosphoesterases. Proteins are generally composed of one or more functional regions, commonly termed domains. Different combinations of domains give rise to the diverse range of proteins found in nature. The identification of domains that occur within proteins can therefore provide insights into their function. The DDHD domain is 180 residues long and contains four conserved residues that may form a metal binding site.

A partial open reading frame 66 aa in length from the 5' end of LP2 was identified using ORFinder (NCBI). None of the other sequenced products displayed any homology to known proteins or possessed any conserved domains of interest. The sequence data gained from LP2 was subsequently named PLIP and was used to initiate a genomic walking method of identifying sequence data both upstream and downstream of the replicon.

6.3 Elucidation Of Gene Sequence Using GenomeWalker™ PCR Amplification

The methodology used for the construction of *O. floccosum* GenomeWalker™ libraries was described in Section 5.2.

6.3.1 GenomeWalker™ PCR Amplification

Four consecutive rounds, consisting of both primary and secondary PCR reactions, were conducted with PLIP gene specific primers PL(1-11) listed in Section 2, Table 2.4.1 and presented here in Figure 6.3.1. These primers were designed to amplify upstream of the 5' end and downstream of the 3' end of PLIP. *O. floccosum* libraries 1, 2, 3 and 4 were used as templates for each reaction. PL11 PLIP specific oligonucleotide primer was designed to further clarify a region upstream of PL 10.

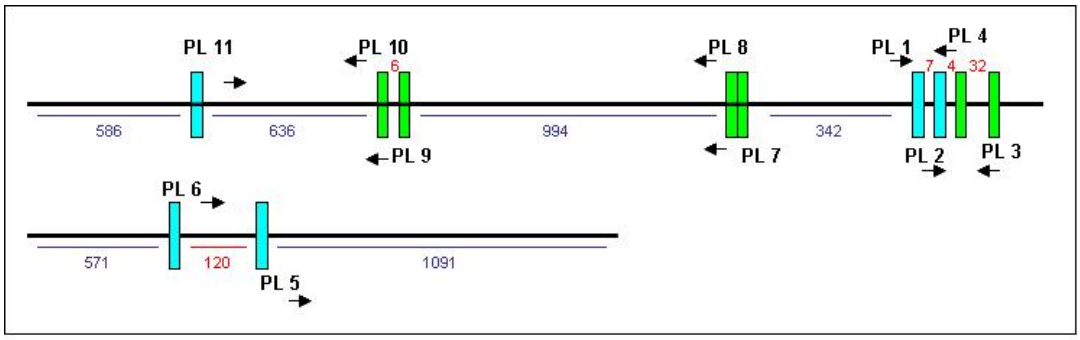


Figure 6.3.1 Illustration of PLIP gene specific primers used, and segments of DNA sequence amplified, cloned and sequenced during consecutive rounds of genome walking. Note, the diagram is not drawn to scale with regard to nucleotide base distances.

6.3.2 Cloning Of PCR Product Into pCR®4-TOPO

Amplicons produced in secondary PCR reactions from each subsequent round were purified, cloned into the vector pCR®4-TOPO and then sequenced as described in

Section 2.5. All amplicons targeting the PLIP gene that were successfully cloned or sequenced directly from PCR product are illustrated in Figure 6.3.1. In total, 6 amplicons were cloned and sequenced, ranging from 342 to 1222 bp in the 5' or 3' direction.

6.3.3 Sequencing Analysis

Clones containing genome walking-nested PCR products were sequenced in the forward and reverse directions. Sequences were screened and trimmed accordingly to remove contaminating vector and poor quality sequence identified from electropherogram data (Section 2.5.5). Screened sequence data was then aligned and assembled using Vector NTI (Invitrogen) to form a consensus sequence PLIP-F.

Three putative ORFs were detected in both the forward and reverse directions of PLIP-F that aligned with significant homology to any known proteins using both OrfPredictor (Min *et al.*, 2005) and ORFinder (NCBI) in a combined approach (Figure 6.3.3.A). The predicted ORFs were subsequently named PLIP-Lg, PLIP-St and HLC. The best matched ("hit") predicted proteins identified using BLASTp analysis of the predicted ORFs are listed in Table 6.3.3.

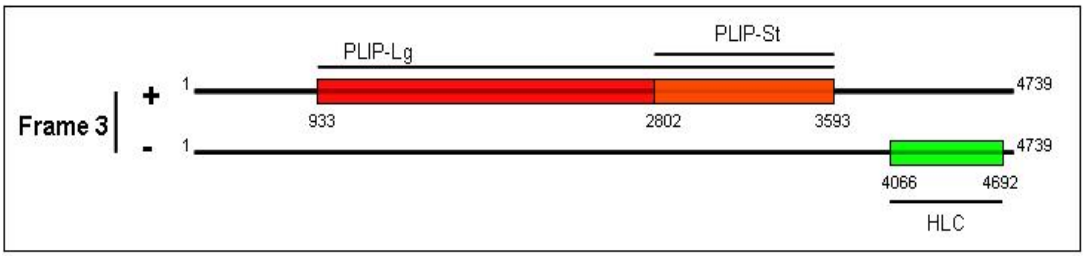


Figure 6.3.3A Putative ORFs detected in PLIP-F. Note, the diagram is not to scale

ORF I.D.	Name/Function Of Aligned Proteins Identified By BLASTX	Score	E-value
PLIP-Lg	Hypothetical protein NCU06812 from *Neurospora crassa* (XP_963653)	855	0
PLIP-St	Hypothetical protein FG04294.1 from *Gibberella zea* (XP_384470)	344	4E-93
HLC	DNA repair helicase RAD3 from *Neurospora crassa* (XP_956536)	355	1E-96

Table 6.3.3A BLASTp best hits for predicted ORFs

PLIP-Lg was 886 aa in length with a predicted molecular weight of approximately 96 kd. PLIP-St was predicted to be 263 aa in length with a calculated molecular weight of approximately 28 kd. Using TargetP, CBS prediction software it was predicted that PLIP-Lg was most likely a mitochondrial protein with a putative cleavage site at

position 38 in the amino acid sequence. No secretory signal was predicted, indicating that PLIP-Lg is an intracellular protein.

The predicted conserved domain architecture of PLIP-Lg was analysed using Pfam and Pfam SMART web based software. Using Pfam A and B searching, the resulting predicted domain architecture was homologous to a *Neurospora crassa* protein (Q8X0B7_NEUCR), thought to be related to a phosphatidic acid-preferring phospholipase A1 (Figure 6.3.3.B). The ORF PLIP-St was a shorter coding region predicted within PLIP-Lg. Hence, both genes had the same conserved domain architecture from 2802 bp to 3593 which included part of a conserved DDHD domain.

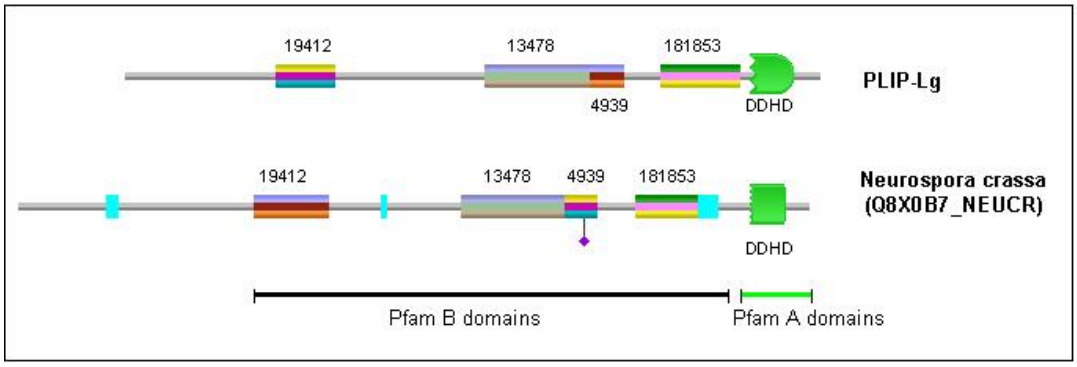


Figure 6.3.3.B. Pfam predicted conserved domain architecture of PLIP-Lg compared with an *N. crassa* predicted phospholipase gene identified in the database.

PLIP-Lg was aligned using ClustalW2 (EBI) with *Neurospora crassa* OR74A (Q8X0B7, XP_963653) and *Cryptococcus neoformans* JEC21 (XP_571372) phospholipase (Figure 6.3.3C). All three protein sequences contained predicted DDHD domains which were highlighted within the figure in yellow. In addition PLIP-Lg also contained a predicted abhydrolase 1 conserved domain (underlined and in bold text), and displayed weak homology to a WWE conserved domain (highlighted in blue). Lines in the sequence indicated gaps introduced by the program (ClustalW2) to optimize the alignment. Identical residues in all of the sequences were denoted by asterisks; conservative substitutions were denoted by dots. In both *Neurospora crassa* OR74A and PLIP-Lg, a consensus lipase sequence was detected and presented in boxed text in Figure 6.3.3C. This catalytic serine is considered to be part of the conserved nucleophilic elbow and can be identified by the conserved motif GXSXG in most lipases (Pleiss *et al.*, 2000).


PLIP-Lg
OR74A
JEC21

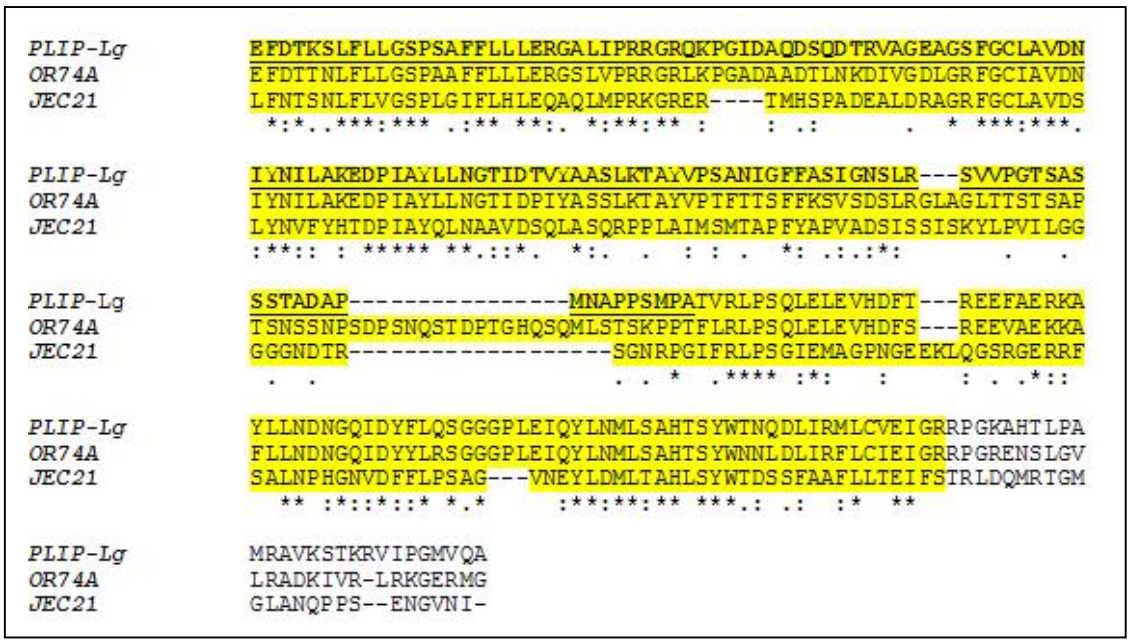


Figure 6.3.3C. Amino acid sequence alignment of the deduced sequences PLIP-Lg, OR74A from *N. crassa* (Q8X0B7, XP_963653) and JEC21 from *C. neoformans* (XP_571372).

The alpha/beta hydrolase fold is common to a number of hydrolytic enzymes. The core of each enzyme is an alpha/beta-sheet containing 8 strands connected by helices (OLLIS *et al.*, 1992). Abhydrolase is a 4-element fingerprint that provides a signature for members of this diverse family of enzymes. The WWE domain is named from its three conserved residues. It is thought to function in ubiquitin and ADP-ribosylation as well as having a predicted role in some proteins mediating specific protein-protein interactions (Aravind, 2001; Arrese *et al.*, 2006).

The presence of WWE and DDHD domains, as well as the conserved lipase sequence GXSXG was identified in a number of phospholipase genes including two insects, *Drosophila melanogaster* (fruit fly) and *Anopheles gambiae* (mosquito) and KIAA0725, a cytosolic PLA_1 ubiquitously expressed in human tissue. KIAA0725 is a mammalian member of the PA-PLA_1 protein family. Homologues of this family were found in *Saccharomyces cerevisiae, Caenorhabditis elegans* and various plant species (Kato *et al.*, 2002; Arrese *et al.*, 2006; Seo *et al.*, 2008). The biological function of proteins in this family is not known. In yeast, the PA-PLA_1 homologue (YORO22C) is a mitochondrial protein. It contains both the conserved lipase active site and a DDHD domain.

The amino acids that are crucial in the catalytic site, histidine, aspartic or glutamic acids, are difficult to find because they are not part of conserved patterns in lipase amino acid sequences. They can only be identified by mutation experiments, structure determination or homology to hydrolases with a known catalytic triad. As no sequence with known structure was identified as being significantly homologous to PLIP-Lg, the presence of this conserved catalytic triad was not able to be detected.

The codon usage of PLIP-Lg was investigated (Table 6.3.3) and compared to the codon usage determined in ORFs in up-regulated transcript and low frequency transcript identified in the *O. floccosum* EST dataset (OF48), as described in Chapter 4.

Amino Acid	Codon	Up-regulated Transcript	Low Frequency Transcript	PLIP-Lg
Phe	TTT	**66.7**	13.3	**57.3**
	TTC	12.0	14.0	21.0
Leu	TTA	**90.7**	2.1	0.0
	TTG	6.6	7.7	**71.0**
	CTT	12.0	26.0	12.1
	CTC	0.0	31.0	17.7
	CTA	4.4	5.4	6.5
	CTG	8.7	**34.9**	4.0
Ile	ATT	29.5	10.7	20.2
	ATC	8.7	**21.0**	0.0
	ATA	**59.0**	4.7	**33.1**
Met	ATG	31.7	17.7	16.1
Val	GTT	**23.0**	19.1	**45.2**
	GTC	1.1	**31.9**	0.0
	GTA	15.3	11.6	29.0
	GTG	2.2	16.1	0.0
Ser	TCT	25.1	6.3	5.6
	TCC	7.7	9.1	16.1
	TCA	23.0	6.3	11.3
	TCG	9.8	16.7	0.0
	AGT	**30.6**	7.0	**42.7**
	AGC	2.2	**20.9**	4.8
Pro	CCT	**19.7**	7.7	**25.8**
	CCC	5.5	14.9	9.7
	CCA	10.9	11.6	17.7
	CGC	5.5	14.9	15.3
Thr	ACT	19.7	5.3	**27.4**
	ACC	5.5	**22.3**	6.5
	ACA	**23.0**	8.2	0.0
	ACG	7.7	15.8	0.0
Ala	GCT	**29.5**	19.3	**14.5**
	GCC	6.6	**37.5**	6.5
	GCA	18.6	14.2	11.3
	GCG	3.3	21.6	11.3

Amino Acid	Codon	Up-regulated Transcript	Low Frequency Transcript	PLIP-Lg
Tyr	TAT	**41.5**	3.9	**19.4**
	TAC	5.5	**16.5**	14.5
Ter	TAA	3.3	1.8	20.2
	TAG	2.2	1.4	16.1
His	CAT	**10.9**	12.3	0.0
	CAC	2.2	**17.0**	**29.0**
Gln	CAA	13.1	6.5	15.3
	CAG	0.0	36.1	22.6
Asn	AAT	**49.2**	14.4	**15.3**
	AAC	7.7	**21.7**	0.0
Lys	AAA	**41.5**	8.6	**36.3**
	AAG	5.5	**35.9**	7.3
Asp	GAT	**23.0**	25.2	0.0
	GAC	6.6	**36.6**	**3.2**
Glu	GAA	**24.0**	14.2	**2.4**
	GAG	5.5	**31.6**	0.0
Cys	TGT	**6.6**	3.5	**4.0**
	TGC	5.5	**10.2**	0.0
Ter	TGA	2.2	1.4	12.1
Trp	TGG	3.3	11.9	25.0
Arg	CGT	1.1	13.0	0.0
	CGC	0.0	**19.1**	**42.7**
	CGA	2.2	11.0	13.7
	CGG	2.2	15.4	0.0
	AGA	**19.7**	9.3	0.0
	AGG	9.8	4.4	31.4
Gly	GGT	**37.2**	23.5	12.1
	GGC	7.7	**44.0**	0.0
	GGA	30.6	11.6	**29.0**
	GGG	5.5	11.0	10.5

Table 6.3.3B Comparison of the frequency of each codon per 1000 codons identified in PLIP-Lg to both up-regulated transcript and low frequency transcript in the *O. floccosum* OP48 EST dataset.

Interestingly, codon usage for PLIP-Lg did not fit particularly well with either model. However, codon usage within PLIP-Lg was more similar to up-regulated transcript identified within the dataset than transcript that did not show any homology to ORFs predicted within singlet sequences.

Primers Plip1 and Plip2 (Table 2.6.4) specific for PLIP-Lg were used to probe a genomic DNA library produced from *O. piliferum* strain 97. Amplification was not successful indicating that no regions homologous to the primers were present in *O. piliferum* genomic DNA. Alignment of the PLIP nucleotide sequence to all UPTs identified in the OF48, OP, ANU and PIC datasets resulted in no significant matches indicating that no homologues of PLIP existed within any of the EST datasets.

6.3.4 Recombinant Expression of the Phospholipase in *E. coli*

Expression of isolated genes in recombinant hosts is a standard method used to determine the biological function of predicted gene sequences and mRNA/protein expression profiling data (Kost, 1999). *E. coli* is one of the most widely used bacterial hosts for the production of recombinant proteins (Baneyx, 1999; Kost, 1999).To test whether PLIP-Lg and PLIP-St had lipase activity, both predicted genes were cloned in *E. coli* and assayed against a negative control for esterase activity.

As described in Section 2.7, PLIP-Lg and PLIP-St ORFs were amplified by PCR from *O. floccosum* genomic DNA using PLIP-Lg and PLIP-St specific primers (listed in Table 2.7.1) containing BamH1, Pstl and Hind III restriction sites. Amplified product of the correct size, determined by gel electrophoresis, was ligated into $pP_{RO}EX$ HTb vector containing a $(his)_6$ spacer region and TEV protease cleavage site. The recombinant ligate was transformed using electrocompetent methods into *E.* coli Top10 cells (Invitrogen). A negative control containing empty $pP_{RO}EX$ HTb vector was also cloned into *E.* coli Top10 cells in a separate reaction. Isolated recombinant transformants were cultivated overnight in antibiotic liquid media. Colony PCR using 5 µl of culture was performed using primers ProF and ProR (Section 2.7) targeted to either side of the multiple cloning site. Sequence analysis of the amplicons confirmed the presence of the PLIP-Lg and PLIP-St inserts. Given that antibiotic resistance was

conferred and the correct inserts were detected it can be assumed that the cloning reaction was successful. SDS polyacrylamide gel electrophoresis (PAGE) analysis of crude cell lysate and protein material purified by a nickel pull down method from PLIP-Lg, PLIP-St and the negative control cloning reactions did not indicate the presence of any expressed protein. In addition to SDS PAGE, the crude lysate and purified protein from all samples were assayed for esterase activity using a discontinuous p-nitrophenyl (*p*NP) assay (described in Section 2.7.3). No increase in the level of esterase activity was detected in the PLIP-Lg and PLIP-St samples compared to the negative control.

A number of factors may have cause or contributed to the lack of observed protein expression. Phospholipase is likely toxic to the cell, therefore selective pressures may have prevented expression of the protein (Arrese *et al.*, 2006). Also the difference in codon usage (discussed in Section 1.4.5.4) observed between *O. floccosum* and *E. coli* may have reduced or prevented expression. Given the difference between eukaryotic and prokaryotic protein production the protein may have been unstable due to incorrect folding or the absence of required chaperones among other factors.

6.3.5 Identification Of Transcriptional Elements

Transcriptional elements upstream of PLIP-Lg were investigated *in silico*. The promoter region of PLIP-Lg contained a number of consensus sequences corresponding to transcription factor binding sites (as shown in Table 6.3.5 and Figure 6.3.5). Transcription factors were identified using both TESS analysis and conserved motifs identified from the literature. All significant motifs identified with homology to motifs previously identified in other fungi were listed in Table 6.3.5. All motifs identified using TESS were listed in Appendix 4. A large number of putative transcription factors were identified in the 5' region of PLIP-Lg, including two predicted TATA boxes and two conserved CCAAT motifs. Interestingly, a number of transcription factor binding motifs potentially involved in restricting the expression of the phospholipase such as GATA factors were identified.

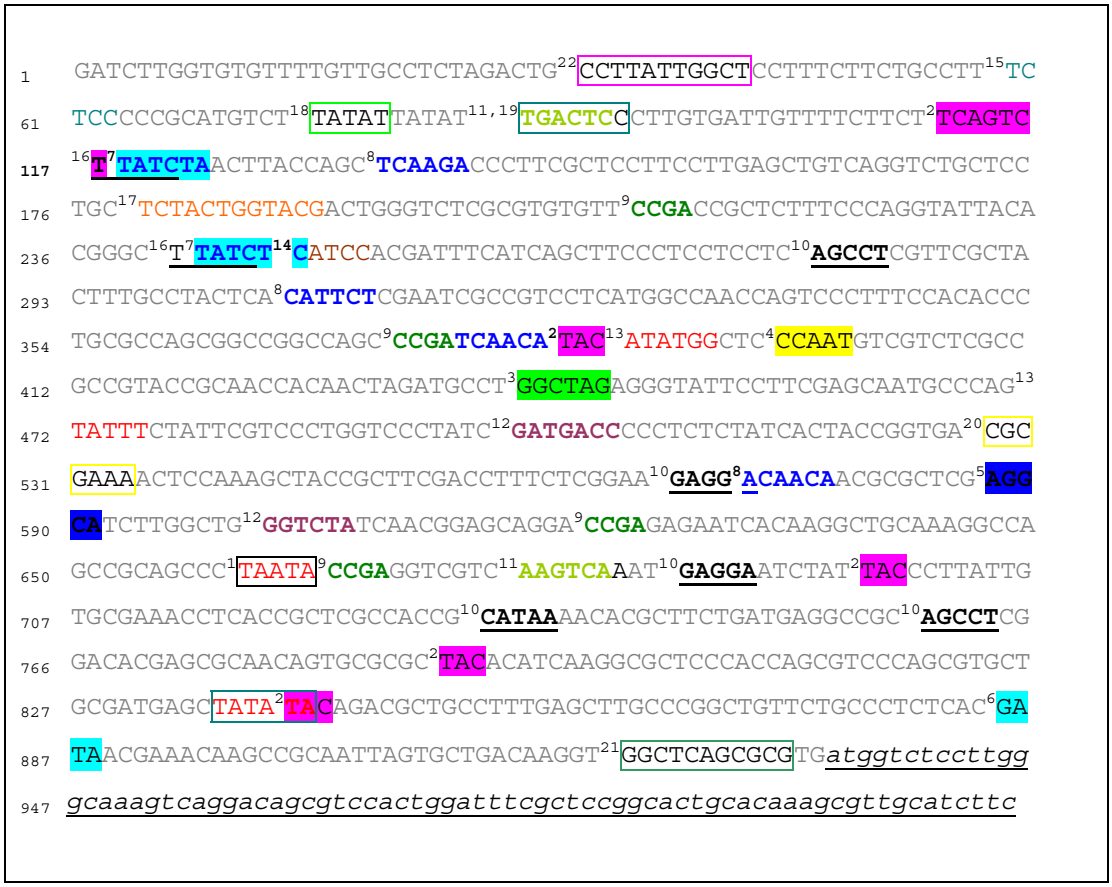


Figure 6.3.5 Upstream 5' region of PLIP-Lg ORF. Part of the predicted coding region is underlined and in lower-case letters. Highlighted areas differentiated also by a number written in superscript at the 5'end of the conserved motif represent predicted transcriptional elements described in Table 6.3.5.

	Motif(s)	Description	No.	Position	TESS Factor
1	**TATAA** **TATATA**	TATA box. A conserved motif found in a number of eukaryotes including *S. cerevisiae* and *S. pombe*	2	835	TFIID
2	TAC, TCAGTCTT	Cap signal for transcription initiation	5	110	CAP
3	GGCTAG	XlnR. A transcriptional activator that controls the expression of genes encoding xylanolytic and celluloytic enzymes. The conserved motif is found in most hemicellulase and cellulose gene promoter of *A. niger*.			
4	CCAAT ATTGG	CCAAT is a conserved motif found in the 5' regions of approximately 30% of eukaryotic promoters including many fungi. ATTGG is a CCAAT binding factor. It contains 2 subunits, both of which are required for DNA binding The B subunit is related to yeast HAP proteins.	2	36, 395	CP1 T00082 T00083 T00087 T00088 T00092 T00099 T00108 T00170 T00613 T00614
5	AGGCA	Conserved motif identified in *cbh1* promoter which binds ACEI.	1	586	
6	GATA	AreA. A conserved motif that binds GATA factors which along with Nit-2 in *A. nidulans* and *N. crassa* functions in the regulation of nitrogen catabolic enzymes during the use of secondary nitrogen sources such as nitrate, nitrite and purines.		883	
7	**TATCTA** **TATCTC**	NIT2. Activator of nitrogen-regulated genes in *A. nidulans* and *N. crassa*. NIT2 can partially complement for lack of AreA in *A. nidulans*.	3	117, 117, 241,	T00627 NIT2
8	**SATTCY,** **$WN_{(1,2)}AAN_{(1,2)}A$**	TCS-like and PRE-like elements. In yeast they have been shown to play a major role in the regulation of genes during the filamentation and invasive response through binding of TEC1 And STE12 transcription factors. TCS-like element in A. nidulans is necessary for spore differentiation and binding to developmentally regulated genes.	4	133, 305, 377, 571	T01085 abaA
9	**CCGA**	Is associated in some way with HAP1 activator or RC2 factor involved in binding to the upstream activation site UAS1 of the CYC1 gene in S. cerevisiae	4	209, 373, 620, 664	T00714 RAF
10	AGCCT GAGGA GATAA GGGG	Activator, mediates galactose response, repressed by GAL80. It binds to DNA and to nucelosome assembled DNA. It binds selectively to human and yeast replication factor a (RPA)	5	278, 567, 684, 758, 884	T00302 GAL4, T00794 TBP
11	**TGACTC,** **AAGTCA**	Activator of genes involved in protein and purine biosynthesis identified in *S. cerevisiae.*	2	84, 675,	GCN4, T00321 GCN4
12	**GATGACC,** **GGTCTA**	Transcriptional element identified in *S. cerevisiae* associated with HAP1. Has been linked with regulation of yeast iso-1-cytochrome C expression.	2	498, 600	T00724 RC2, RC2
13	ANATGG, TATTT	GL1 type zinc finger protein. Is a homolog of yeast global regulator that functions as a transcriptional repressor.	2	386, 471	T00865 YY1, T00915 YY1

	Motif(s)	Description	No.	Position	TESS Factor
15	TCTCC	Positive regulator of peroxisomal protein genes identified in *S. cerevisiae*. It is required in yeast for ADH2 activation and glycerol metabolism.	1	58	T00011 ADR1
16	TTATC	An ARS-binding protein, identified in *S. cerevisiae* involved in sorting of mtDNA, mitochondrial matrix protein, and mtDNA recombination.	2	116, 240	T01274 ABF2
17	TCNNNNNNNA-CG	Transcriptional element identified in *S. cerevisiae*. It contains two functionally essential C-terminal regions (CS1, CS2). The C-terminal region CS1 is responsible tor transcriptional silencing and/or repression in context-dependent manner, CS2 is required for all functions of Abf1p, namely transcriptional activation, stimulation of DNA replication and gene silencing. Functionally it is similar to RAP1.	1	178	T00056 ABF1
18	TATAT	Transcriptional element that has an interacting factor T00043; ARG80 in *S. cerevisiae. It functions to* control arginine metabolism by acting as a repressor of arginine synthesis and inducing Arginine catabolism.	1	79	T00765 SRF (504 AA)
19	TGACTCC	Transcriptional element identified in *S. cerevisiae* involved in response to oxidative stress/oxygen detoxification and metal resistance. It has similarities with GCN4. An upstream ORF allows 40S subunits to proceed via leaky scanning and re-initiation to the major ORF.	1	84	AP-1 - YAP1
20	CGCGAAA	Transcriptional element identified in *S. cerevisiae*. It functions as a cell cycle box factor by mediating cell-cycle dependent transcription of HO gene and START-specific transcription.	1	527	T00096 CCBF T00775 SWI4 T01013 SWI6
21	GGCTCAGCG-CG	A transcriptional element identified in Arabidopsis. It has a homolog in *S. pombe*. It is hypothesised to have a role in cell cycle regulation.	1	920	T02855 CDC5 -
22	CCWTNTTNNN-W	A transcription factor that binds to downstream elements in several polymerase II promoters.	1	32	T00278 delta factor T00915 YY1

Table 6.3.5 Summary of putative transcriptional elements in the 5' region of the predicted PLIP-Lg gene in *O. floccosum*.

6.4 Profile Of Putative Lipase Gene Expression At Different Stages Of Growth Using RT-PCR

The relative levels of PLIP-Lg expression throughout growth in liquid media, and in different growth forms, either blastospore or mycelial, were determined using RT-PCR.

6.4.1 Culture Growth and RNA Isolation

RNA was isolated in duplicate at specific time points in *O. floccosum* liquid cultures as described in Sections 5.3.6.1 and 2.2.2.

6.4.2 Construction of cDNA

cDNA was constructed from RNA samples as described in Sections 5.3.6.2 and 2.2.5.

6.4.3 RT-PCR Amplification

RT-PCR was conducted using PLIP-Lg gene specific primers PLIP 1 and PLIP 2. The primers were designed with a melting temperature of between 50 and 65ºC (Table 6.4.3). Amplification produced an amplicon of 90 bp in length.

Gene	Primer	Nucleotide Sequence	Tm (NN) ºC	Amplicon Size
PLIP	PLIP 1	5'- GATAGCCAGGATACCCGAGTAG -3'	61.2	90 bp
PLIP	PLIP 2	5'- GTCTTCTTTGGCAAGGATGTTG -3'	63.2	

Table 6.4.3 PLIP oligonucleotides used in RT-PCR

RT-PCR reactions were done according to the methods described in Sections 2.6.5 and 2.6.6. Dissociation curve analysis was performed after every reaction to identify the presence of contaminating sequence. RT-PCR product was also analysed by electrophoresis on a 0.7% agarose gel to confirm amplicon size and purity.

6.4.4 Determination of RT-PCR Efficiency

The efficiency of RT-PCR amplification in *O. floccosum* cDNA libraries using PLIP specific primers was assessed according to the methods described in Section 5.3.6.4.

The amplification curves of the PLIP serial dilutions are shown in Figures 6.4.4A and 6.4.4B. C_T was measured at a threshold of 0.4.

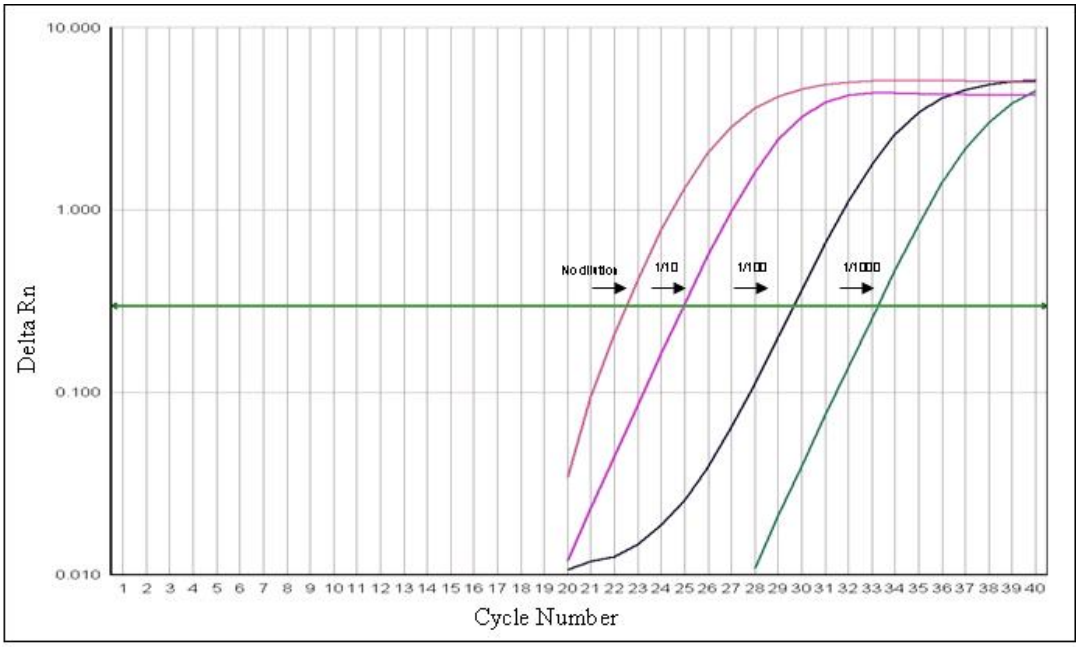


Figure 6.4.4A Amplification plot showing Delta Rn against Cycle Number for representative RT-PCR analysis of the PLIP gene in *O. floccosum.* The cDNA template consisted of serial dilutions constructed from mycelium cultivated for 96 hours.

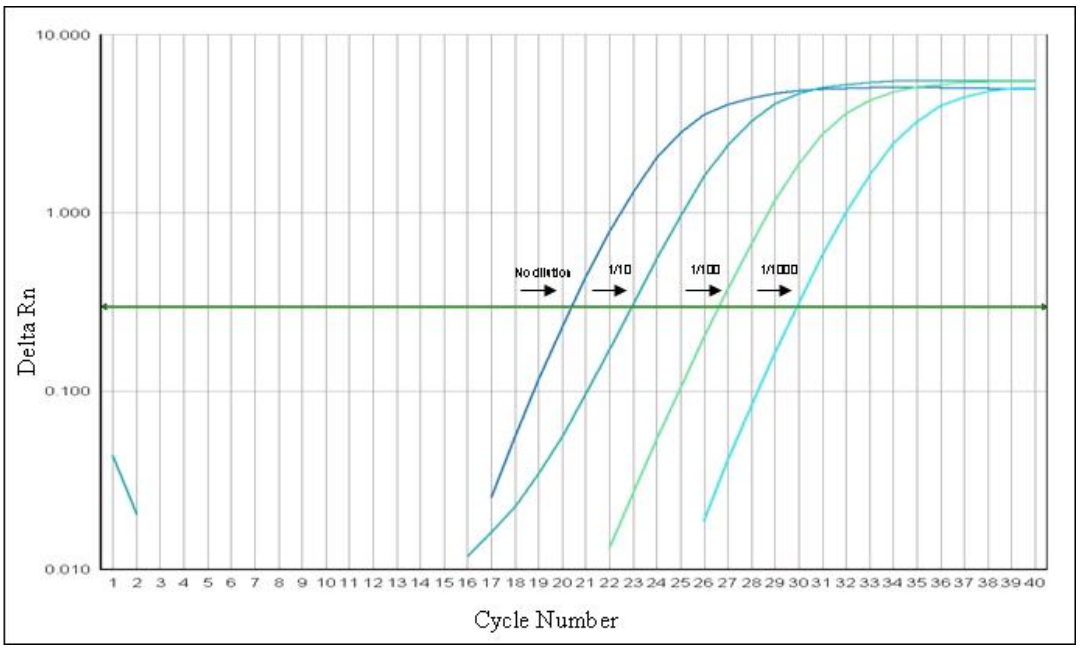


Figure 6.4.4B Amplification plot showing Delta Rn against Cycle Number for representative RT-PCR analysis of PLIP in *O. floccosum*. The cDNA template consisted of serial dilutions constructed from blastospores cultivated for 96 hours.

Dissociation curve analysis (Figures 6.4.4C and 6.4.4D) of sequence verified PLIP – Lg RT-PCR product had a melting temperature of approximately 82.6ºC. The cDNA templates used in the RT-PCR reactions consisted of serial dilutions constructed from mycelium cultivated for 96 hours. Note, four repeat amplifications of multiple 1/10

serial dilutions of mycelial sample produced a melt curve with an additional curve of approximately 74.8ºC and subsequently were not used in standard curve calculations. This did not occur in other samples despite the fact that subsequent dilutions were made from the original sample. Attempts to sequence the 1/10 mycelial amplicons failed as multiple products could not be separated. Electrophoretic analysis of the amplicons produced a single band of approximately the correct size.

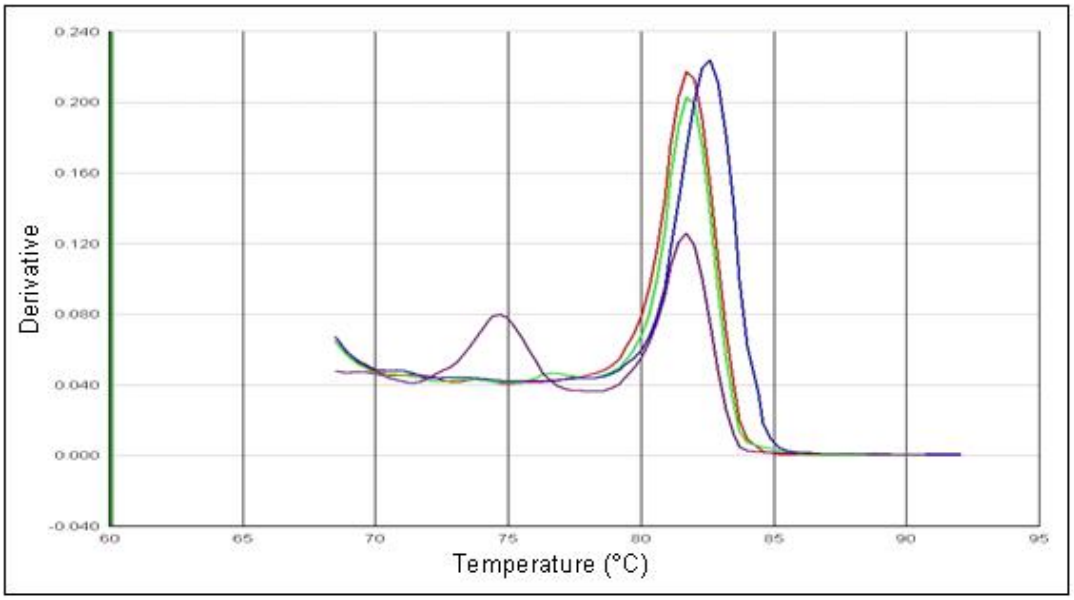


Figure 6.4.4C Dissociation curve showing Derivative against Temperature for representative RT-PCR analysis of PLIP-Lg in *O. floccosum*.

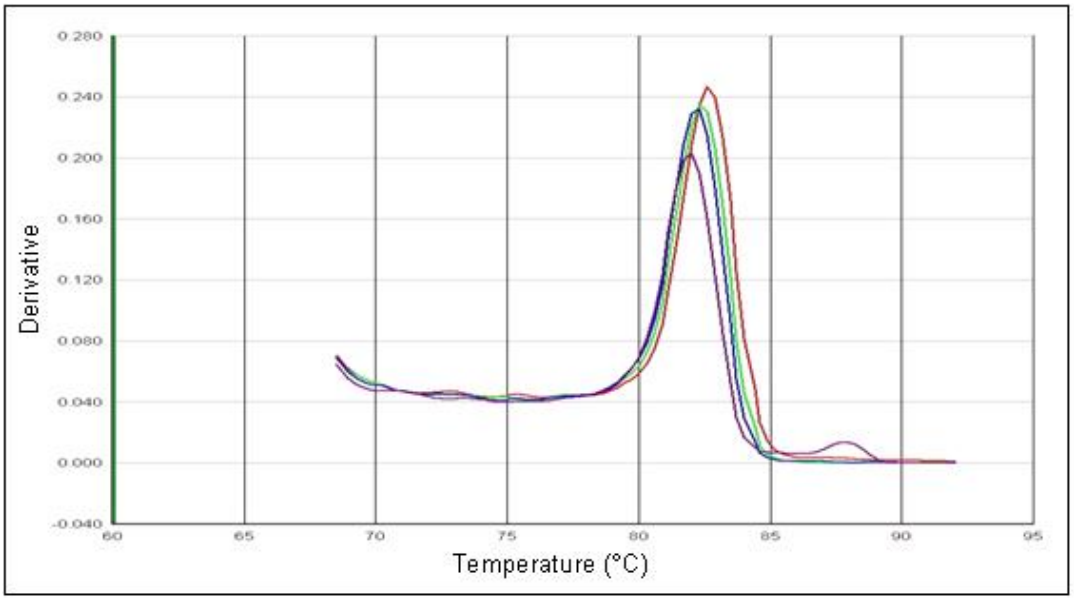


Figure 6.4.3D Dissociation curve showing Derivative against Temperature for representative RT-PCR analysis of PLIP-Lg in *O. floccosum*.

Analysis of R^2 values indicated relatively consistent amplification (Figure 6.4.4E) with values 0.9984 and 0.991 for RT-PCR analysis of 96 hour mycelial and blastospore growth respectively.

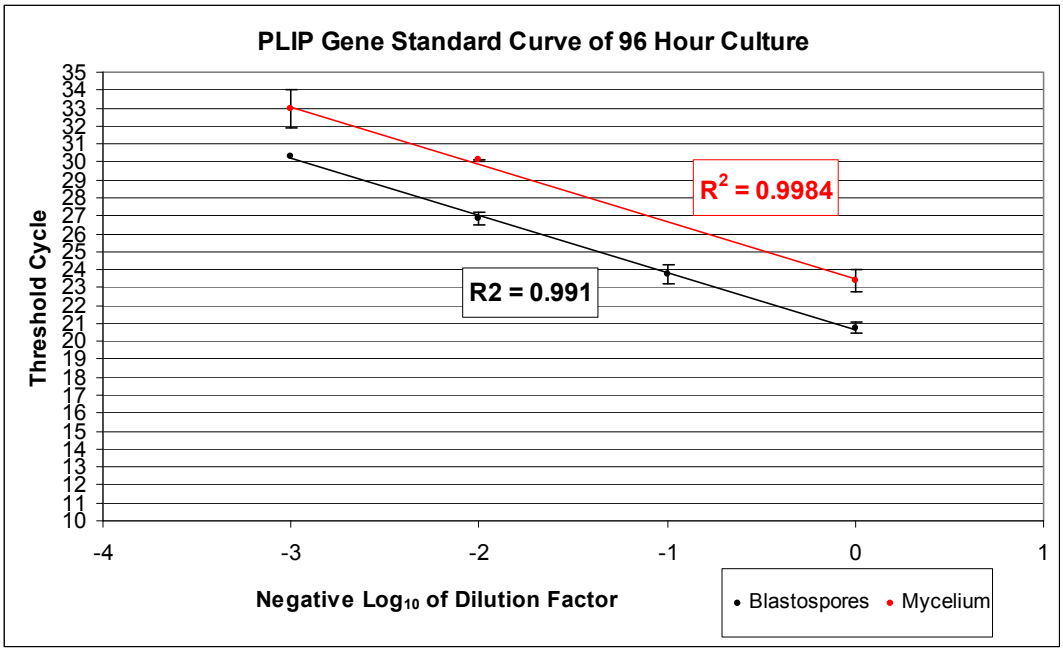


Figure 6.4.4E Standard curve used to calculate the efficiency of PLIP-Lg RT-PCR

The PCR efficiency was calculated using the following equation:

$$\text{PCR efficiency} = 10^{\wedge(1/S)} - 1$$

$$\text{PCR efficiency \%} = (10^{\wedge(1/S)} - 1)100$$

Where S = the slope of all standard curve samples assayed using PLIP-Lg gene specific primers.

$S_{(PLIP)} = -3.31$

PCR efficiency $\%_{(PLIP)} = 101\%$

A PCR efficiency of 101% is considered optimal for RT-PCR analysis, according to *Real-Time PCR Applications Guide- Bulletin 5279* (BIORAD, U.S.A.).

6.4.5 PLIP Expression In Different Somatic States Throughout Growth

The relative levels of DNA expression in mycelial and blastospore somatic cells were compared in 14 cDNA samples constructed over 7 time points. All samples were run in duplicate, and in some instances triplicate or quadruplicate RT-PCR assays. No

exogenous supply of lipid material was added to the medium at any time so any changes in expression were not the result of artificial environmental stimuli.

The average C_T values calculated for between 2-4 PCR replicates at given time points and in both forms of growth are presented in Table 6.4.5. The data indicated that PLIP expression was detected in both blastospores and mycelia at all stages of growth sampled. Data for 24 hour blastospores and 192 hour mycelium was presented in the Table to show that expression was detected, but, as both of these samples did not have sufficient RNA concentrations for cDNA construction, these results were not comparative.

	Blastospore		Mycelia	
Sample Time	Average C_T	Standard deviation	Average C_T	Standard deviation
24	28.03	0.93	19.94	0.01
48	20.29	0.01	24.58	0.20
72	19.59	0.30	24.07	0.01
96	20.74	0.31	23.39	0.58
120	20.08	0.57	19.13	0.05
144	16.01	1.23	18.66	0.28
192	16.94	0.11	17.80	0.06

Table 6.4.5 Average C_T values and standard deviations produced by RT-PCR analysis of PLIP at different time points throughout growth in blastospore and mycelia forms.

The level of PLIP expression typically increased over time in both mycelia and blastospores, although it was quite stable from 48 to 120 h in blastospores. It is interesting to note that despite the lower concentration of RNA template used to construct the cDNA library, expression in mycelia was highest at 192 hours.

The relative expression of PLIP in blastospores compared to mycelial growth was calculated using the following equations:

$$\text{Ratio}_{(\text{blastospores / mycelium})} = E^{C_{T\,(\text{mycelium})} - C_{T\,(\text{blastospores})}}$$

Where $E = 10^{\wedge(1/S)}$

and S = the slope of all standard curve samples assayed using PLIP gene specific primers.

$$E = 10^{\wedge(1/3.31)}$$

$= 2.01$

$$\text{Ratio}_{(\text{blastospores / mycelium})} = 2.01^{\Delta C_T}$$

A graph showing the ratio of PLIP-Lg expression in blastospores compared to mycelium is given in Figure 5.4.5A. In all but one sample time point, 120 hours, the level of PLIP-Lg expression was higher in blastospores than in mycelium. The greatest difference in expression was observed at 72 hours where an approximate 23 fold increase in the level of PLIP-Lg expression in blastospores was observed compared to RNA isolated from mycelium. Interestingly, the difference in the level of expression between the two growth forms was greater in younger cultures where the overall level of expression was less. At 144 hours, when expression was the highest recorded in blastospores, there is only a 6 fold difference in levels of mRNA relative to one another.

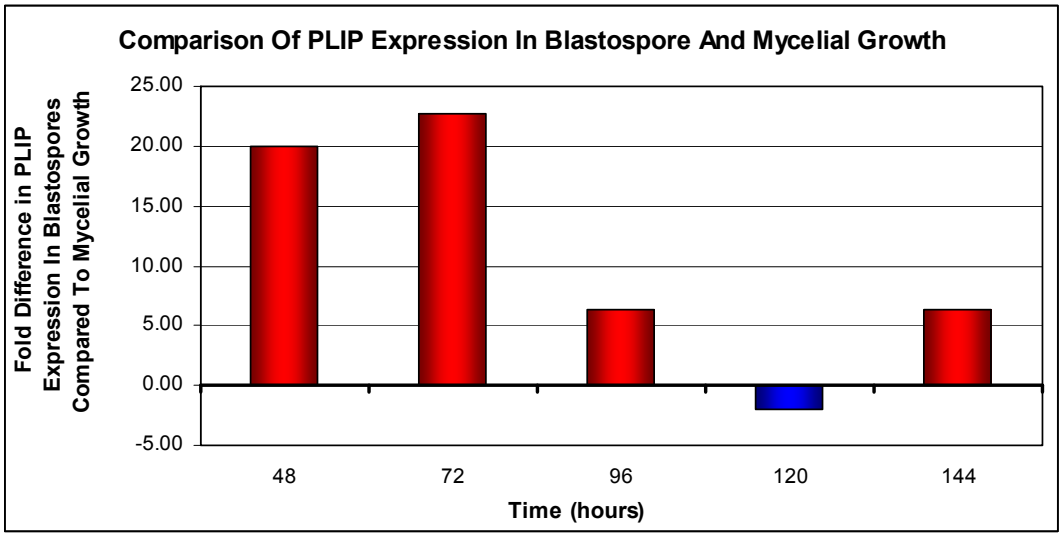


Figure 6.4.5A The ratio of PLIP-Lg expression in blastospores compared to mycelium in *O. floccosum*.

The relative levels of PLIP expression in *O. floccosum* were compared against NADH-UR4, an up-regulated transcript identified by EST analysis reviewed in Chapter 5. Comparisons were based on the assumption that the efficiencies of PLIP and NADH-UR4 RT-PCR, 2.01 and 2.05, respectively, were sufficiently similar and that all other factors with the exception of gene target remained constant. A graph displaying the proportional increase in the level of NADH-UR4 compared to PLIP is presented in Figure 6.4.5B. For calculation purposes, a mean efficiency of 2.03 was

assumed. Given that the same cDNA template was used for all *O. floccosum,* RT-PCR analysis, data for 24 hours and 192 hours were included.

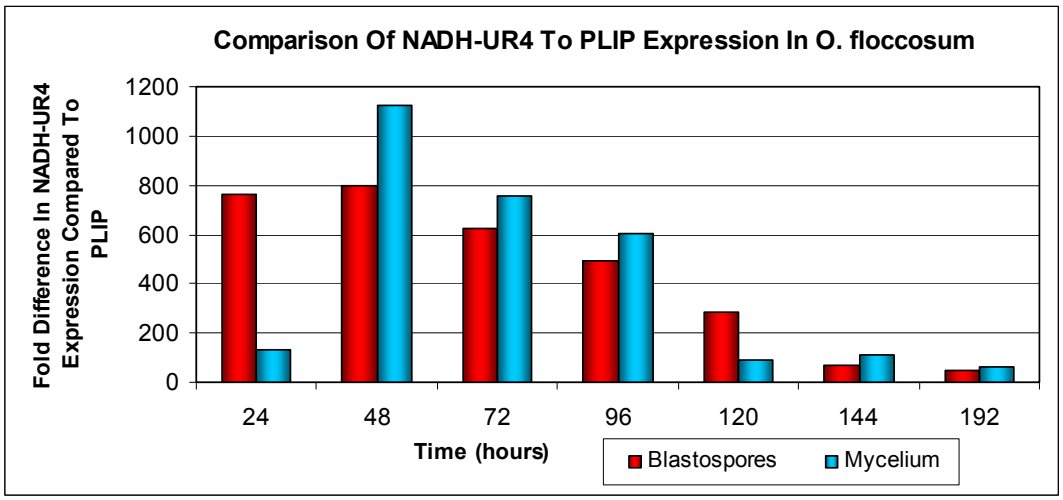


Figure 6.4.5B The proportional level of NADH-UR4 compared to PLIP-Lg expression in *O. floccosum*.

The difference in the levels of expression was greatest in mycelium as opposed to blastospores in all samples with the exception of 24 hours and 120 hours

NADH-UR4 expression was demonstrated in Chapter 5 to be relatively constant in all but 24 hours. Given that NADH-UR4 was by far the most up-regulated transcript in the OF48 dataset, we can imply from the results as shown in Chapter 6 that PLIP expression was relatively high in both mycelia and blastospores in late phase growth.

Given that the greatest level of expression was recorded at 144 hours of growth, when blastospores were essentially stationary, not undergoing cell replication, PLIP-Lg may have a role in membrane homeostasis and remodeling as well as nutrient acquisition from fat deposits within the cell.

7 Conclusions and Future Recommendations

Ophiostomataceae species represent a group of insect-dispersed pyrenomycetes isolated throughout the world. The ascomycete family contains members that are plant pathogens such as *O. novo-ulmi* and *O. clavigerum* as well as many sapstaining fungi that affect the aesthetic value of decaying wood and timber but do not affect its structural integrity (Zimmerman *et al.*, 1995; Schirp *et al.*, 2003a). This PhD thesis research evolved from an original aim to identify biotechnologically and functionally relevant proteins such as lipases and investigate up-regulated transcripts in *Ophiostoma* species with the ultimate goal of identifying a promoter region for use in a recombinant expression vector. Such a vector would therefore be applicable for the rapid production of recombinant protein able to be up-scaled to a commercial volume in a fermenter system.

O. floccosum was identified as the second most abundant *Ophiostoma* species isolated throughout New Zealand. Despite being first described by Mathiesen-Käärik in 1951, little is known about this species. Its morphological characteristics were described by Harrington *et al.* (2001) and subsequently the nucleotide sequence of subtilase, β-tubulin and ribosomal subunit genes have been described as well as the identification, cloning and sequencing of a secreted α-amylase. Apart from the exceptions just listed, very little about the genome or protein production in this fungus has been investigated. This PhD thesis research also included investigation of *O. piliferum* a species which has been developed as a biocontrol agent for worldwide use. Albino strains of both *O. piliferum* and *O. floccosum*, in particular the *O. piliferum* strain 97 marketed as Cartapip™, have been used successfully as bio-control agents to prevent the colonisation of other sapstaining fungi. *O. piliferum* species have been investigated by a number of groups worldwide with results providing insight into growth and morphology, melanin and perithecial developments as well as the production of various hydrolases, specifically secreted proteases and lipases (Zimmerman *et al.*, 1995; McNaughton, 1997; Brush *et al.*, 1999, Harrington *et al.*, 2001; Hoffman and Breuil, 2004).

At the initiation of the PhD thesis research, little was known about the molecular functioning of these organisms. Despite the deposit of 9591 *O. piliferum* EST sequences onto NCBIs dbEST by Tsang *et al*., in 2006, no data has been published on the transcriptome of this organism and little comparative analysis has been made among the different species. Furthering the knowledge and understanding of molecular functioning of this group of fungi may offer many advantages, specifically in terms of identifying target areas for biocontrol, and exploiting its polymorphic growth capacity and potential for *in situ* application for use as a biotechnological tool.

As part of the thesis research, culturing of *O. floccosum* strain J2122 and *O. piliferum* strain 97 were compared using three laboratory methods to observe the range of morphologies produced, as well as cell density, and culture viability in submerged liquid cultures. It was important to characterise the growth of each strain in order to relate molecular functioning to *in vitro* growth. The three conditions of growth produced differences in both the rate of growth measured by blastospore production and the total biomass in both species. The inoculation size, starting pH, media type and concentration and temperature were all kept constant. Therefore, two factors that differed with each of the culturing methods, oxygen dispersion and availability and the level of agitation were key variables affecting the growth of both strains. These investigations have further use in identifying key differences in the types and stages of growth that may aid in the prediction of optimal times for competent cell production, transformation and recombinant growth in future research.

Microscopic analysis of growth of the two strains showed that both species were morphologically similar and shared many time-dependent characteristics of growth. In early growth, blastospores were typically elongated and mycelial fragments had elongated apical ends, likely due to their active growth. By contrast, mid phase blastospores were typically much rounder and more yeast-like in appearance, actively budding and often producing multiple buds. By late phase, blastospore budding was much reduced and blastospores appeared much smaller. During late phase growth mycelial fragments were generally much wider and typically contained buds at the apical ends.

Biomass accumulation for *O. floccosum* was likely affected by exopolysaccharide production as the difference in biomass accumulation between *O. floccosum* and *O. piliferum* was neither reflective of the total blastospore levels achieved, these were relatively proportional between the stirred tank bioreactor and shaken cultures, or the viability levels between the conditions as the level of viable colony forming units was greater in *O. piliferum* fermentation cultures (1.7×10^{11} CFUs compared to 2.36×10^{9} CFUs in *O. floccosum*). Therefore, the differences in biomass were not due to increased filamentous or yeast like growth. The presence of an extracellular hyphal sheath containing proteinacious materials and various polysaccharides was reported in both *O. piceae* and *O. novo-ulmi* (Ericksson *et al.*, 1990; Gharibian *et al.*, 1996; Ouellete *et al.*, 1999; Jeng *et al.*, 2007). Exopolysaccharide production has also been reported for various yeast species (Pavlova *et al.*, 2004) and in both yeast-like and mycelial forms in dimorphic fungi (Reeslev *et al.*, 1997). The role of this slime sheath is still somewhat uncertain. In filamentous fungi, the sheath occurs along the length of the hypha and coats the growing tip (Schirp, 2002). In wood decay and sapstaining fungi it has been suggested that the sheath may protect extracellular enzymes from inactivation and retain these enzymes close to the microorganism, offering a competitive advantage to the fungus *in vivo* (Kalisz *et al.*, 1987; Green III., 1992; Schirp, 2002). During screening of various fungal species for exopolysaccharide production, many phytopathogens were reported to produce exopolysaccharides suggesting that exopolysaccharide production may be a pathogenicity factor (Selbmann, *et al.*, 2003). In *Discula umbrinella*, the proteinaceous sheath was reported to be responsible for the adhesion of conidia to host organisms and glycoproteins were involved in the recognition of the attachment process (Viret *et al.*, 1994). The structure and physical properties of microbial polysaccharides depends on the culture medium composition and growth conditions (Pavlova *et al*., 2004). Exopolysaccharide formation *in vitro* is often accompanied by a change in the pH of the culture medium (Reeslev *et al.*, 1997; Pavlova *et al*., 2004). Since wood is acidic, with a typical pH of 4-5, the extracellular sheath hypothesised for *O. floccosum* may also serve to lower the pH of the culture medium to maintain an environment favourable for enzyme activity (Gharibian *et al.*, 1996; Schirp, 2002).

The production of exopolysaccharides may be of future biotechnological significance. They are used in industry for a wide variety of applications including food additives

(for example thickeners, emulsifiers, leavenings) (Pavlova *et al*., 2004). Further research is required to investigate possible exopolysaccharide production in *O. floccosum*.

In *O. floccosum* late phase growth, fewer mycelia were observed compared to *O. piliferum* cultures and in both late phase cultures blastospores were smaller producing few buds. In *Ophiostoma* species, autolysis was shown to occur in the late stages of *in vitro* growth (Schirp, 2002). Autolysis is characterised by several markers including decreasing viability, declination of dry cell mass, vacuolization, formation of empty hyphae, fragmentation of hyphae, decreasing pellet diameter, production of numerous intracellular and extracellular hydrolases (e.g. proteinase, chitinase, glucanase, RNase, DNase) and the release of ammonia (Emri *et al.*, 2008). Autolysis in *S. cerevisiae* resulted in increased intracellular proteolytic enzymes and subsequent rupturing of the cell (Knorr *et al.*, 1979). This process occurs under carbon starvation; in the case of *O. floccosum* and *O. piliferum* growth investigated in this PhD research, this would likely occur in the late phase of growth. It is thought that autolysis in filamentous fungi is a well-regulated process aiding in survivability. In many filamentous fungi a higher degree of lysis using exogenous lytic enzymes was observed in the cell walls of younger mycelia (Perez-Leblic *et al.*, 1982). It is possible that in *O. floccosum* blastospores are more resistant to autolysis than mycelial fragments. The yeast-like morphological form has been reported to be dominant in other filamentous fungi in carbon-limited liquid cultures (Sami *et al.*, 2001). As blastospores are thought to be a dispersal mechanism in *Ophiostoma* species preferential survival of these spores over active mycelia is likely to be a favourable trait.

Autolysis results in increased secretion of hydrolase enzymes, therefore, it is possible that some of the lytic enzymes produced in later growth could be used in the production of competent cells used for transformation of early stage cells. Further investigation into lytic enzymes produced by *O. floccosum* and *O. piliferum* for use in recombinant expression could potentially be beneficial in the development of these species as recombinant protein expression hosts.

By investigating the transcriptome of an organism, we can better relate genotype to phenotype and further our understanding and/or capability to manipulate the functioning of an organism. One such transcriptomic method includes ESTs. The *O. floccosum* EST dataset (OF48) represented blastospores and inherently some small mycelial fragments in early stage growth in liquid media. Analysis of growth in early phase blastospores versus other phases or morphological forms was two fold in reasoning. From a biotechnological point of view, it was advantageous to understand transcriptomics in particular up-regulated transcripts that could be indicative of a strong promoter at an early stage in a fermentable system. Also, EST datasets produced from other *Ophiostoma* species were generally produced at later stages of growth, often on solid media and typically isolated from or including the mycelial form. Based on the information available regarding the current *Ophiostoma* EST libraries, no one had investigated early stage growth, and therefore these PhD thesis results are original findings in the field.

By assuming that the genome of *O. floccosum* is comparable to the sequenced genomes of other select Pezizomycotina species, the *O. floccosum* EST library produced was predicted to represent between 4.3 and 6.6% of all *O. floccosum* genes. Saccaromycotina species typically have fewer ORFs as was discussed in Section 1.2.2. Based on the estimated representation of the dataset it is likely that the mRNA transcripts sequenced within the library represent moderate to up-regulated transcripts, particularly those UPTs identified with an EST transcript frequency of greater than 2.

Sequence homology does not automatically imply functional similarity. However, it is essentially the only practical existing method that allows large scale annotation of novel genes. Annotation based on sequence homology of the OF48 EST dataset indicated that the majority of transcripts putatively identified were involved with rapid cell proliferation and carbohydrate metabolism as has been observed in many Saccharomycotina (yeast) species. In *S. cerevisiae*, approximately half of all cellular protein was found in fewer than 100 different gene products, which were mostly involved in carbohydrate metabolism or protein synthesis (Futcher *et al*., 1999). If estimates of mRNA transcript abundance to cognate protein level of 1, to 4,000 were

correct and comparative in *Ophiostoma* species, then the high level of transcripts indicated suggest that carbohydrate metabolism and protein synthesis also dominates the majority of proteins produced in *O. floccosum* blastospores during exponential growth.

A number of the UPTs identified in the *O. floccosum* EST dataset involved in metabolism were predicted to form part of the glycolysis and gluconeogenesis pathways. In many fungi, glucose is converted to glucose-6-phosphate or fructose-6-phosphate to pyruvate through the glycolytic pathway. During respiration, pyruvate is usually oxidized to CO_2 through the tricarboxylic acid (TCA) cycle (Arraes *et al*., 2005). A summary of the glycolysis and gluconeogenesis pathways is presented in Figure 7.1. Those enzymes in the pathways with homologous transcripts in the *O. floccosum* EST dataset are boxed in red. Glycolysis and the TCA cycle have a dual role in many fungi; they produce energy in the form of ATP, NADH or NADPH and provide building blocks to synthesise other biomolecules (Flores *et al.*, 2000; Arraes *et al.*, 2005). Of the putative genes identified predicted to be involved in glycolysis and gluconeogenesis, many were homologous to transcripts in other *Ophiostoma* datasets described in Table 4.3.4.

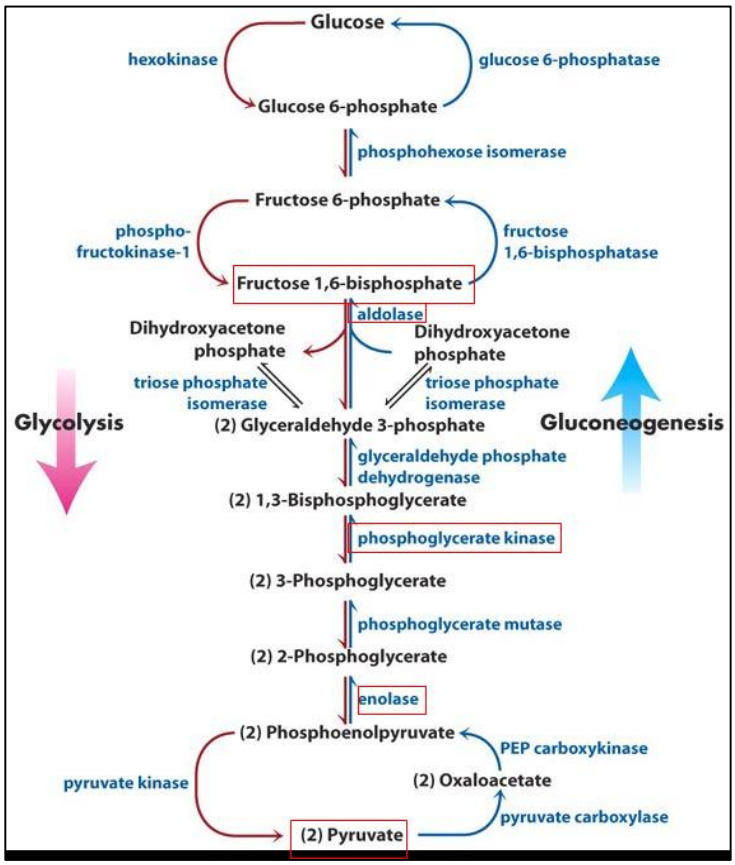


Figure 7.1 A summary of the glycolysis and gluconeogenesis. Those enzymes with homologous transcripts in the *O. floccosum* EST dataset are boxed in red. The figure was adapted from Dr Thomas O. Baldwin, http://www.biochem.arizona.edu/classes/bioc462/462b/graphics/Glycolysis.

Many fungal species utilise both aerobic (respiration) and anaerobic (fermentation) pathways to obtain energy from glucose in the form of ATP. *S. cerevisiae* preferentially ferments glucose-producing ethanol and CO_2 by anaerobic metabolism. By contrast, in many filamentous Pezizomycotina species, respiration is preferentially used (Chambergo *et al.*, 2002; Kenealy and Dietrich, 2004; Kayali *et al.*, 2005). In the dimorphic fungus *Mucor racemosus*, the filamentous form is capable of respiration whereas the yeast-like form exploits fermentation (Chambergo *et al.*, 2002). The presence of transcripts in the *O. floccosum* EST dataset putatively identified as being involved in the TCA cycle and the up-regulation of transcript encoding mitochondrial genes was interpreted as being indicative that respiration is preferred in *O. floccosum* in the blastospore form of growth, and likely in the mycelial form, also. Under the laboratory conditions for growth investigated, increasing the oxygen (O_2) concentration of the liquid medium accelerated the level of blastospore and total biomass production in *O. floccosum* and *O. piliferum*. However, this does not reflect

the situation *in vivo* as increased O_2 is unlikely to ever occur under the natural progression of colonisation and growth normally performed by these fungi. Rather, there is evidence to suggest that many wood degrading and inhabiting fungi, including *Ophiostoma* species are subject to O_2 limitation during growth on wood (Thacker and Good, 1952; Kenealy and Dietrich, 2004). Also, O_2 limitation is likely to occur in small pockets *in vitro*, for instance O_2 may be limited in the centre of filamentous 'pellets' or balls produced during agitated growth. O_2 limitation in wood inhabiting fungi has been little investigated in the literature. It was proposed that for *Phanerochaete chrysosporium,* as with many other filamentous fungi including *Ophiostoma clavigerum, O. mintium, Ceratocystis resinifera*, and *C. rufipenni* investigated under conditions of O_2 limitation, that the fungi are primarily aerobic, but that they can sustain themselves under limited oxygen conditions (Sing *et al.*, 1992; Skory *et al.*, 1997; Solheim and Krokene 1998a, 1998b; Kenealy and Dietrich, 2004; Morin *et al.*, 2007). In the research conducted on *P. chrysosoporium* no increase in biomass was observed under O_2 limited conditions (Kenealy and Dietrich, 2004). This would support the reduced levels of growth in *O. floccosum* and *O. piliferum* in terms of cell density observed in statically grown liquid cultures, discussed as condition 3 in Chapter 3. Also, during static growth in liquid cultures, much of the filamentous growth able to be observed visually unaided was near the top of the culture medium further indicating that respiration is preferred by these species. The presence of putative alcohol dehydrogenase genes within *O. floccosum* OF48 EST dataset, and other *Ophiostoma* species including *O. piliferum,* of which two proportionally up-regulated transcripts were identified, suggests that anaerobic growth is possible in these organisms. Alcohol dehydrogenase is a key enzyme in fermentative growth converting pyruvate to ethanol. In *P. chrysosporium*, alcohol dehydrogenase activity was detected in ethanol producing cultures subjected to O_2 limitation but not in cultures where O_2 was not specifically limited (Kenealy and Dietrich, 2004). Based on the evidence provided from this PhD research and findings in other wood inhabiting fungi (Sing *et al.*, 1992; Skory *et al.*, 1997; Kenealy and Dietrich, 2004), it is likely that *O. floccosum* has the ability to utilise carbon under both aerobic and anaerobic conditions. In nature, fermentative metabolism in *Ophiostoma* species may allow the maintenance of active cells which could exploit new sources of O_2 as they become available.

The PhD thesis research showed that few genes were identified involved in biomass degradation despite the previously reported presence of hydrolase enzymes in *O. floccosum* during exponential growth in the supernatant of mixed mycelial and blastospore cultures grown in liquid medium (Schirp, 2002). This may be because the transcription of hydrolase genes was much lower compared to genes involved in cell proliferation and glucose metabolism and, therefore, not detected in the limited number of EST sequences produced from the OF48 cDNA library. Another explanation may be that hydrolase expression is somewhat limited to the mycelial rather than the yeast-like form of *O. floccosum*. Genome analysis of Saccharomycotina has shown that they produce fewer genes involved in biomass degradation such as glycoside hydrolases compared to Pezizomycotina (as was discussed in Section 1.2.3.4). However, the secretion of many hydrolases, in particular proteinases, is a key factor in virulence in fungal pathogens (Calderone, 2002). The yeast-like phase is thought to be a means of asexual dispersion in *Ophiostoma* species and the ability to secrete hydrolases in order to assimilate nutrients for colonization and growth would therefore be advantageous in yeast-like cells. Alternatively, many genes identified in plant cell wall degradation in filamentous fungi were found to be repressed by the presence of easily metabolisable carbon sources such as glucose and induced only in the presence of substrate polymers (Arvo *et al.*, 2005). Hence, the readily available carbon in the form of glucose in the nutrient rich media used may have prevented the transcription of many hydrolases.

Of the 598 *O. floccosum* OF48 UPTs annotated using BLASTx and TargetIdentifier 166 UPTs, approximately 72%, were significantly homologous to known or predicted protein sequences in the public databases. Of the 432 annotated UPTs, 110 aligned with the greatest homology to bacterial sequences. Alignment of UPTs to bacterial sequences was not observed to the same extent in the other *Ophiostoma* datasets. The reasons for this may be one or a combination of several factors. Firstly, and in the opinion of the author most likely, the lesser number of sequences with fungal homology was likely due to a lack of fungal sequences available in the databases. To date, the complete genome of a sapstaining fungus has not been published and functional analysis using mutation and gene knockout methods of *Ophiostoma* genes is

limited. [Colette Breuil's group is currently working on the full genomic sequencing of *G. clavigerum* but this is not published yet (personal communication). While genomic research on *Ophiostoma* species is way behind compared to research on "model" species, there is nonetheless work going on including two recent papers by Bouvet describing DNA transposons in *O. ulmi* and *O. novo-ulmi* (Bouvet *et al*., 2007, 2008). Secondly, the homology of EST sequences to bacterial genes may indicate contamination of the OF48 cDNA library with bacterial genomic DNA. A number of possible opportunities existed for bacterial contamination during the production and cloning of a cDNA library, however, stringent methods were used to ensure that bacterial contamination was unlikely. All cultures were screened microscopically at both 40X and 100X magnification and plated on nutrient rich general media to identify any possible bacterial contaminants. To avoid the contamination of recombinant phage or *E. coli* host cells, isolated plaques were amplified using PCR methods before being sent for sequencing. Additionally the cDNA library was screened using universal bacterially specific primers to identify the presence of contaminating bacterial DNA. It was therefore unlikely that contamination of the cDNA library with bacterial DNA occurred.

The assessment of genetic relatedness between organisms using the EST data is limited as EST sequences represent a 'snapshot' of cellular functioning at a particular time in response to exogenous and endogenous stimuli. Even with all variables involved in growth and cDNA construction kept constant, there are many factors that would limit the comparativeness of EST data. For example, the completeness of the transcriptomic profile represented by an EST dataset is related to its size. It is unlikely that all protein being transcribed was represented within each dataset. Also, EST sequences are typically short and up to 7 mRNA's differing by their 3' ends may be produced from a single ORF due to variable transcript termination and polyadenylation (Fukumura, 2003). As was reported in Section 4.3.4, approximately 13% of all *Ophiostoma* EST sequences analysed in this study had homology to at least one other UPT identified from another *Ophiostoma* species. This indicated that despite variation in the known methods of culturing and cDNA construction, comparisons of EST sequence data between *Ophiostoma* species was an effective method of identifying homologous genes between the species.

Many of the putative genes identified in common between *O. floccosum* and the other *Ophiostoma* datasets were involved in protein production and core metabolism. For example, a number of putative genes involved in transcription and translation such as ribosomal subunits, initiation and elongation factors as well as transcripts encoding histones were common to many of the datasets. As was mentioned above, a number of transcripts putatively encoding enzymes involved in glycolysis and gluconeogenesis were identified in common between the datasets. Given the dominance of transcript identified within the *O. floccosum* EST dataset that were predicted to have a cellular role in these two categories, it is not surprising that the findings were biased toward these two functional groups.

Of particular interest in the analysis of the *Ophiostoma* datasets was the identification of genes putatively encoding for peptidases (also termed proteases, proteinases and proteolytic enzymes). While some peptidases serve a useful application in industry, peptidase production can severely inhibit the production of recombinant protein in fungal protein expression systems. Wood sapstaining fungi are known to produce multiple proteases that break down wood protein. Wu *et al.* (2006) identified by N terminal sequencing, two subtilisin-like serine peptidases reported to be strongly secreted in *O. floccosum* cultures. Subtilases are the dominant extracellular proteases produced by *O. floccosum* (Hoffman and Breuil, 2004). To date, no other types of peptidase have been reported for this species. From the homology based putative identification of proteases in *Ophiostoma* species in this PhD thesis research, it was apparent that all *Ophiostoma* examined contained homologues to peptidases in the MEROPS database with the exception of *O. piceae.* It is likely that no putative peptidases were identified in the *O. piceae* dataset because of its small size; with only 34 EST sequences it is obvious that the majority of proteins produced by this species were not likely to be represented. Also, this dataset was produced from two suppressive subtracted cDNA libraries designed to identify genes transcribed in either mycelial or blastospore form only. It is possible that any proteases produced by *O. piceae* are expressed in both forms of growth. This is also supported by the fact that no proteases were identified that were present in only the HN yeast dataset and the *O. floccosum* yeast dataset.

Proportionally, the number and frequency of peptidases identified in the EST datasets was reasonably high. With little published in the literature on total protease production, both intracellular and secreted, it is difficult to determine if this level is reflective of other fungal species examined or if it is a feature of these sapstaining fungi. Caution should be taken when interpreting the protease relative frequencies, as the transcription of protease genes is likely to be reflective of the culture conditions used to construct the cDNA used for EST analysis. As was reported for *Ophiostoma* species previously in the literature, a high number of putative serine peptidase families were identified within each of the data sets (Breuil and Huang, 1994; Hoffman and Breuil, 2004; Wu *et al.*, 2006). Proportionally high numbers of different metallo protease families were also identified. Only one UPT within any of the *Ophiostoma* datasets was identified as being homologous to a glutamic protease. Glutamic proteases are thought to only be found in fungi. Homologues have been identified in the genomes of many ascomycete species with the exception of the class Saccharomycetales (Sims *et al.*, 2004). The peptidase homologues identified in *O. floccosum* and their likely function was reviewed in Section 4.3.4.1. The prediction of sub-cellular localisation and secretory signals was not conducted as it was likely that many of the UPTs identified contained only partial ORFs. This PhD thesis research constitutes the first reported findings of putative peptidases in the aspartic, cysteine, glutamic and threonine peptidase families in *Ophiostoma* species.

Codon usage in predicted ORFs within UPTs of *Ophiostoma* species was investigated with the exception of *O. piceae* (Section 4.3.4.2). This research is the first reported look into codon usage in *Ophiostoma floccosum*, *O. piliferum, O. clavigerum* and *O. novo-ulmi*. As was discussed in Section 1.2.5.4, it has been established that in a variety of organisms, highly expressed genes exhibit preference for some of the codons encoding the same amino acids (Grantham *et al.*, 1980; Wada *et al.*, 1990; Jansen *et al.*, 2003; Carbone *et al.*, 2004). This is referred to as codon bias. The 'biased' codons are codons recognised by the most abundant isoacceptor tRNA species (Bennetzen and Hall, 1982; Ikemura, 1982; Berg and Kurland, 1997). A significant amount of research into fungal codon usage, and bacterial codon usage has resulted from the need to optimise and subsequently increase heterologous protein expression as the existence of

slightly different codes in different organisms is a significant barrier to heterologous gene expression. For prokaryotes codon usage has been identified as the single most important factor in gene expression (Lithwick and Marghalit, 2003; Gustafsson *et al.*, 2004).The codon usage of putative genes predicted to be up-regulated based on high transcript frequency within each of the EST datasets were compared to those identified as singlet sequences. The codon usage in up-regulated transcripts identified in all of the *Ophiostoma* species investigated was consistent with the exception of *O. floccosum*. As discussed in Section 4.2.4.2, the large difference between the HF dataset in *O. floccosum* in terms of preferred codon usage, GC content and positioning, and the HF datasets in the other *Ophiostoma* species, was likely due to the combined effect of a small number (7) of up-regulated *O. floccosum* transcripts contained in the dataset, and the likelihood that many of these putative genes were mitochondrially encoded. In contrast, the codon usage in the *O. floccosum* HFC dataset, containing all likely medium and up-regulated fragments in the EST library, was consistent in terms of codon usage with the other *Ophiostoma* species investigated. There seemed to be little difference between the codon usage of low frequency and high frequency datasets in *Ophiostoma* species. This may be because the predicted genes contained within the low frequency datasets did not reflect low frequency transcript or that there was little difference between them and that Codon bias is not a significant factor in the regulation of protein expression in *Ophiostoma* species.

Comparisons of the preferred codon usage between different organisms (provided in Table 4.3.4.2B) indicated that codon usage in up-regulated transcripts was more similar between Pezizomycotina species and humans than Saccharomycotina and humans. However, in order to express recombinant proteins of human origin in *Ophiostoma* species, it is likely that the codon usage for valine, serine and arginine would have to be optimised. It is important to note that codon usage was determined from predicted ORFs within UPTs. Therefore, false predictions of ORFs or the prediction of partial ORFs that are not in frame will skew the results. Given the consistency of codon usage observed between the *Ophiostoma* datasets, with the exception of the *O. floccosum* OF48-HF dataset, it is unlikely that any such false predictions have significantly impacted on the results.

To exploit the *Ophiostoma* genus as an expression host an array of gene promoters must be identified for the different forms of growth. A central premise to this thesis is that up-regulated transcripts are fairly indicative of transcriptional control by a strong promoter. The most up-regulated transcripts from both the *O. floccosum* and *O. piliferum* datasets were investigated using genome walking methods to further elucidate the gene sequence and RT-PCR was used to compare expression throughout growth. Currently, little is understood about the differences in total protein expression and production between mycelial and yeast-like morphological forms of growth in *Ophiostoma* species. Ideally, a strong promoter that is up-regulated in both mycelial and blastospore forms of growth would be advantageous in a recombinant expression system.

In total 135 *O. floccosum* EST fragments were identified as being homologous to at least 4 different subunits of NADH-ubiquinone oxidoreductase Complex I representing 11.2% of the total EST library. Of these 135 EST fragments, 59, representing 4% of the total EST library, aligned to form a single consensus sequence which had the greatest homology to subunit 4 of the NADH-ubiquinone oxidoreductase complex. Nicotinamide Adenine Dinucleotide (NAD) and its reduced form (NADH) are electron (e^-) plus proton (H^+) carriers that transport $2e^-$ and $2H^+$ at a time (Madigan and Martinko, 2006). NADH-ubiquinone oxidoreductase oxidizes NADH, transferring two electrons to ubiquinone with concurrent translocation of four protons across the mitochondrial membrane (Clason *et al.*, 2007). In many fungi, as well as plants and prokaryotes, electron transfer from NADH to the ubiquinone pool of the respiratory chain is carried out by 2 kinds of enzymes: 1) Complex I or 2) alternative NADH-ubiquinone oxidoreductases (Kerscher *et al.*, 1999; Joseph-Horn *et al.*, 2001; Kerscher *et al.*, 2001).Complex 1 is a large, multi subunit, mitochondrial enzyme. In *Yarrowia lipolytica* it is composed of at least 38 different subunits (Abdrakhmanova *et al.*, 2004) and at least 35 unique subunits in *Neurospora crassa* and *Aspergillus niger* (Videira, 1998; Joseph-Horn *et al.*, 2001) with a molecular mass of around 1 MDa in some eukaryotes (Kerscher *et al.*, 2001; Friedrich and Böttcher, 2004). The simplest form of proton-pumping NADH-ubiquinone oxidoreductase known exist in bacteria, consisting of 14 subunits with a molecular mass of around 535 kDa. Seven of these subunits are hydrophobic proteins predicted to fold into 54 α-

helices across the membrane (Friedrich and Scheide, 2000). These seven subunits have homologues to mitochondrially encoded subunits in eukaryotes. The remaining subunits in eukaryotes are thought to be encoded by nuclear genes which are then synthesised in the cytoplasm and imported into the mitochondria (Videira, 1998; Friedrich and Böttcher, 2004). Complex I oxidises NADH produced mainly by the pyruvate dehydrogenase complex in the citric acid cycle and in β-oxidation. Therefore, its NADH binding site faces the matrix side of the mitochondrial inner membrane (Kerscher *et al.*, 2001). As in other eukaryotes, fungal Complex I forms an L-shaped structure with the subunits associated with proton translocation of ubiquinone situated within the hydrophobic membrane region. Complex I is highly conserved in terms of composition, structure and function in most eukaryotes with the exception of some yeasts. In *Saccharomyces cerevisiae, Saccharomyces carlsbergii* and *Kluyveromyces lactis*, Complex I is absent (Büschges *et al.*, 1994; Joseph-Horne *et al.*, 2001). In *S. cerevisiae* this is likely due to the organisms' preference for ethanolic fermentation, rather than respiration as its preferred mode of glucose utilisation (Kerscher *et al.*, 2001). Despite the importance of Complex I and the fact that it is the largest complex in the respiratory chain, it is perhaps the least understood. Currently, the exact positioning and functioning of the various subunits is unclear likely due to the fact that targeted manipulation of the mitochondrial genome is difficult if not impossible (Videira and Duarte, 2001; Clason *et al.*, 2007). In *N. crassa* it is thought that both subunits 4 and 5, identified as being up-regulated in *O. floccosum*, are located in the peripheral or 'ankle' part of the complex structure. The 'ankle' is thought to protrude from the membrane so it is on the matrix side of the mitochondrial membrane and contains the binding site for NAD(H), and the input electron transfer chain (Videira and Duarte, 2001). It was suggested that subunit 4 may function in quinone interaction (Videira and Duarte, 2001). However, NADH-UR4 identified in *O. floccosum* is thought to be mitochondrially encoded subunit and is therefore likely to be a hydrophobic protein predicted to fold into α-helices across the membrane (Friedrich and Scheide, 2000). Ultimately the exact function of subunits 4 and 5 remains unclear and the reasons as to why subunits 4 and 5 should be significantly more up-regulated in OF48 are not apparent. It has been reported in the literature that transcription of genes encoding alternative NADH dehydrogenases may be regulated throughout growth in some fungi. However, there is little reported on the expression of genes

encoding subunits of complex I. Based on the RT-PCR analysis presented in Section 5.3.6.5 it would appear that the subunit 4 NADH-ubiquinone oxidoreductase homolog identified in *O. floccosum* is constitutively expressed throughout growth in both the mycelial and blastospore forms.

Within the *O. piliferum* OP EST dataset, 178 EST fragments representing 1.9% of the OP dataset aligned to form the UPT identified as OP0411. Genome walking methods were used to elucidate a 4108 bp nucleotide sequence fragment encoding a 324 aa predicted protein homologous to a heat shock protein in *N. crassa*. Heat shock proteins are thought to act as chaperones to aid in the folding of other protein to ensure an active conformation. They are often produced in response to stress stimuli, including heat, oxidizing conditions and exposure to toxic compounds. The suggested function of this predicted protein would imply that expression was either repressed or activated in response to certain stimuli. Based on the RT-PCR results, it would indicate that HSP was expressed throughout growth at a relatively consistent level in both blastospore and mycelial forms in *O. piliferum*. The number of mRNA transcripts produced were typically greater in mycelia compared to blastospores and a slight increase in mRNA level was seen in mid to late phase growth in mycelia with up to 17 fold higher level in mycelia compared to blastospores at 144 hours of growth.

Both *O. piceae* and *O. piliferum* were reported to produce extracellular lipases that hydrolyse triglycerides in wood extractives into free fatty acids resin acids (or whatever R group is attached through oxygen to the glycerol backbone) and glycerol during colonisation (Abraham *et al*., 1998; Gao and Breuil, 1998). Not only do lipases serve an important role in wood colonisation, they also have wide ranging roles and uses in industry, including the direct application of *O. piliferum* albino strain Cartapip ™ 97 to reduce pitch/wood extractives in paper manufacturing (Blanchette *et al.*, 1992; Farrell *et al.*, 1992; Wendler *et al.*, 1992). Using PCR methods of bioprospecting and genome walking methods to elucidate the nucleotide sequence upstream and downstream of amplicons, a 96 kd protein was identified with homology to a phosphatidic acid-preferring phospholipase A1 gene referred to as PLIP-Lg. The predicted protein contained a conserved motif encoding a catalytic serine thought to form part of a conserved nucleophilic elbow in lipases. In addition to this consensus

sequence, three conserved domains, a DDHD domain, WWE domain and abhydrolase domain, were identified by homology in the predicted phospholipase gene. The presence of all three domains has been reported in a number of lipase and phospholipase proteins (Kato *et al.*, 2002; Arrese *et al.*, 2006; Seo *et al*., 2008) further supporting its likely function as a phospholipase protein. As was described in Section 1.2.1.2.3, phospholipases are involved in a diverse number of processes including membrane homeostasis and remodeling, nutrient acquisition, generation of bioactive molecules, as well as pathogenesity and virulence (Köhler *et al.*, 2006). Using prediction based software it was hypothesised that PLIP-Lg was likely to encode an intracellular protein as no secretory signal was identified.

RT-PCR indicated that the predicted phospholipase gene PLIP-Lg was expressed throughout growth in both blastospore and mycelial forms and that the level of expression typically increased with the age of the culture. Codon usage of the predicted ORF was consistent with it being an up-regulated protein. Given the low abundance of PLIP in early expression relative to NADH-UR4, an up-regulated transcript identified in the *O. floccosum* EST dataset, it was not surprising that this gene was not detected within the EST dataset using homology methods. Expression of PLIP-Lg was typically greater in blastospores compared to mycelia. Given that the greatest level of expression was recorded at 144 hours of growth when blastospores were not as actively reproducing, PLIP-Lg may have a role in membrane homeostasis and remodeling as well as nutrient acquisition from fat deposits within the cells. These processes may be part of basic regulatory functioning or they could possibly be involved with autolysis or senescence. Senescence is an inherent degenerative program in multicellular organisms that is manifested by a progressive decline in cellular energy production culminating with the death of a part of the whole organism. Senescence has been observed in a number of filamentous fungi. A number of UPTs within the *O. floccosum* dataset identified as being most homologous to senescence related proteins. Both of the processes, autolysis due to carbon starvation and senescence, may explain the smaller size of late phase blastospores compared to early and mid phase growth.

Attempts to clone the predicted phospholipase gene in *E. coli* were not successful. A number of factors may have caused or contributed to the lack of expressed protein, many of which were discussed in Section 6.3.4. It is possible that the predicted phospholipase was toxic to the cell and selective pressures within *E. coli* may have prevented production of the protein. Also, differences between bacterial and fungal protein production such as codon usage, required post translational modifications for active conformation and the absence of required chaperones among other factors may have resulted in unstable protein that could not be observed (Arrese, 2006). It is likely that future research into PLIP-Lg expression would be more successful in a eukaryotic system such as the yeast systems *P. pastoris* and *S. cerevisiae*.

Interestingly, the three putative genes (NADH-UR4, HSP and PLIP-Lg) investigated in this PhD thesis research did not contain any predicted introns based on nucleotide sequence. A number of genes, both nuclear and mitochondrially encoded, have been identified in *Ophiostoma* species containing introns (Gibb and Hausner, 2005; Paoletti *et al.*, 2005; Wu *et al.*, 2006). Either the results are accurate and these genes do not have introns or possibly the detection software used in the PhD thesis research, despite being trained on eukaryotic genes, was not suitable for *O. floccosum* and *O. piliferum*. Within the *O. floccosum* PLIP-Lg ORF, PCR amplicons produced from cDNA were the same size and contained the same nucleotide sequence as PCR amplicons produced from genomic DNA indicating that no introns were present in this putative gene.

Many proteins investigated in filamentous fungi are regulated mainly at the transcriptional level (Aro *et al*., 2005). Cis-acting regulatory elements are short conserved motifs typically 5-20 nucleotides in length usually found in the vicinity of a promoter. Transcription factors bind to the DNA of specific cis-acting regulatory elements and orchestrate the initiation or transcription. In many cases, the role of activators is to recruit transcription factors to the promoter site rather than directly recruit the polymerase itself. Within the *O. floccosum* and *O. piliferum* promoter regions investigated *in silico*, a number of potential transcription factor binding sites were identified. These included both repressors and activators identified in other fungal species. Transcriptional repressors bind to the specific sites on DNA and prevent transcription of nearby genes whereas activator proteins recruit basal

transcriptional machinery to the promoter. Many eukaryotic genes are reported to be controlled by a multiplicity of transcription factors. Within *Ophiostoma* species there have been few investigations into transcriptional regulation and the promoter regions of genes identified (Bowden *et al.*, 1994; Pereira *et al.*, 2000; Wu *et al*., 2006). Perhaps most significantly with regards to this PhD research was the elucidation of the 5' promoter region of an α-amylase gene in *O. floccosum* (Wu *et al.*, 2006). As was described in Section 5.1, within the α-amylase promoter region, a few key transcriptional factors were reported including a possible TATA box, a conserved CCAAT box and five putative CAAT motifs. The identification of putative TATA boxes was a consistent finding in all three of the putative genes investigated. The conserved TATA motif is often located approximately 30 bp upstream from the site where transcription begins and in many eukaryotes it tends to be surrounded by GC rich sequences (Wolner and Gralla, 2000). However, within the *Ophiostoma* promoter regions investigated in this study, and including the α-amylase investigated by Wu *et al*. (2006), the presence of GC rich sequences was not determined. Rather in the *O. piliferum* predicted HSP gene, the conserved TATA motif was flanked in the 3' region by a long string of adenines. Replicate sequencing of the nucleotide fragment confirmed that this was not the result of sequence error. The TATA binding protein recognises and binds to the TATA box. It is a general transcription factor required for initiation of transcription in eukaryotes (Pongsunk, *et al.*, 2005). TATA binding proteins have been cloned and characterised in a number of fungi. From the literature reviewed these small proteins (approximately 30 kDA) would appear to be relatively conserved at the amino acid level. During transcription initiation the TATA binding protein acts as a common subunit that interacts with other specific transcription factors resulting in the formation of a transcription pre-initiation complex by all three RNA polymerases (Hernandez, 1993; Pongsunk *et al.*, 2005). Data presented by Pongsunk *et al.*, (2005) suggested that the TATA binding protein had a role in dimorphic control within *Penicilium marneffei* with it being required for filamentous growth but less relevant to the growth of the pathogenic yeast form. Indeed, several studies that identified transcription factors involved in morphological control and switching in *P. marneffei* were reported to be RNA polymerase II-dependent (Borneman *et al.*, 2002; Pongsunk *et al.*, 2005). Further investigation into the role of the TATA binding protein in *Ophiostoma* species may therefore provide greater insight into the control of

dimorphism in these organisms. Many genes have been identified in fungi that do not have a conserved TATA motif, mostly those encoding general metabolic proteins.
The 5' promoter regions of both the *O. floccosum* putative PLIP-Lg gene and *O. piliferum* HSP gene contained conserved CCAAT motifs as was seen in the α-amylase investigated by Wu *et al*. (2006). CCAAT sequences are found in the 5' promoter regions of approximately 30% of eukaryotic genes and have been identified in the promoter regions of genes encoding plant cell wall degrading enzymes and transcriptional factors modulating expression (Aro *et al*. 2005). The protein contains two subunits, both of which are required for DNA binding. The B subunit is thought to be related to the yeast HAP proteins (Aro *et al*. 2005).

Also, within the 5' promoter regions of both the *O. floccosum* putative PLIP-Lg gene and *O. piliferum* HSP gene, were conserved motifs representing putative NIT2 activators. In both *A. nidulans* and *N. crassa* NIT2 proteins encode putative zinc finger DNA-binding domains and function in the activation of nitrogen-regulated genes. It is possible that the lack of observed homologous transcriptional factor motifs in the 5' promoter region of the *O. floccosum* NADH-UR4 gene is due to a lack of nucleotide sequence at the 5' end compared to both *O. floccosum* putative PLIP gene and *O. piliferum* HSP genes. It is likely that further elucidation of nucleotide sequence upstream of the 5' end would identify other putative transcriptional factors.

The identification of transcriptional elements from conserved motifs is purely speculative. The functional regions within upstream promoters need to be determined experimentally using knockout and mutational cloning methods. This requires an efficient method of transformation and expression in the subject organism, or organisms sufficiently related, and was beyond the scope of this PhD thesis research. Cloning and mutational analysis to assess gene function has been reported in various *Ophiostoma* species (Sorbo *et al.*, 2000; Tanguay and Breuil, 2003; Hoffman and Breuil, 2004). *Agrobacterium tumefaciens* mediated transformation into both *O. piliferum* and *O. piceae* has been reported (Tanguay and Breuil, 2003; Hoffman and Breuil, 2004). The function of a conserved Cerato-ulmin gene from *O. ulmi* and *O. novo-ulmi* was determined by cloning and expressing it in *O. piceae* (Del Sorbo *et al.*, 2000). To date there has been no report of transformation and expression in *O.*

floccosum. Future research into cloning and transformation of both homologous and recombinant proteins in *O. floccosum* is essential for its use as a possible recombinant expression host.

In summary, in this PhD thesis research the following hypotheses were investigated and proven correct.

- Analysis of ESTs constructed from blastospore dominant cultures at log phase in *O. floccosum* revealed up-regulation of specific genes.
- The levels of up-regulated mRNA transcripts in *O. floccosum* and *O. piliferum* were different between blastospore and mycelial morphological forms of growth.
- The levels of unique mRNA transcripts changed during different stages of growth

Some of the findings of this PhD thesis research were presented at the 'Protein Expression Europe' conference in Prague, Czech Republic and the extended abstract written for this conference is given in Appendix 5.

Concerning future research that this PhD thesis research suggests, the following objectives were identified as possible milestones, both fundamental and applied (the order does not represent a ranking in terms of importance):

- The analysis of all *Ophiostoma* EST datasets to identify possible alternative metabolic pathways such as the glyoxylate cycle and identify the enzymes used in fatty acid as well as purine and pyrimidine metabolism.
- Compare *Ophiostoma* sequences to those in the yeast (*S. cerevisiae*) protein database of the Munich Information Centre For Protein Sequences (MIPS) to understand the relatedness of these ascomycetes.
- The cloning of the putative Phospholipase A1 gene identified in *O. floccosum* in a eukaryotic vector system to establish its functional role and phenotypic significance within the organism
- The determination of functionally significant transcription factors in the 5' promoter regions of NADH-UR4, HSP and PLIP-Lg using mutational cloning and reporter vector systems.

- The identification of other possible genes of interest homologous to multiple *Ophiostoma* species that may be controlled by a strong promoter for use in a recombinant expression vector.
- The investigation of exopolysaccharide production in *O. floccosum* in relation to various growth conditions such as media, temperature, pH, the level of oxygen and the degree of agitation.

References

Abdrakhmanova, A., Zickermann, V., Bostina, M., Radermacher, M., Schägger, H., Kerscher, S., & Brandt, U. (2004). Subunit compostition of mitochondrial complex I from the yeast *Yarrowia lipolytica*. *Biochimica et Biophysica Acta, 1658*, 148-156.

Abraham, L., & Breuil, C. (1996). Isolation and characterization of a subtilisin-like serine proteinase secreted by the sap-staining fungus *Ophiostma piceae*. *Enzyme and Microbial Technology, 18*(2), 133-140.

Abraham, L., Hoffman, B., Gao, Y., & Breuil, C. (1998). Action of Ophiostoma piceae proteinase and lipase on wood nutrients. *Canadian Journal of Microbiology, 44*, 698-701.

Abraham, L. D., Chow, D. T., & Breuil, C. (1995). Characterization of the cleavage specificity of a subtilisin-like serine proteinase from Ophiostoma piceae by liquid chromatography/mass spectrometry and tandem MS. *FEBS Letters, 374*(2), 208-210.

Alexopoulos, C. J., Mims, C. W., & Blackwell, M. (1996). *Introductory Mycology* (4th ed.): John Willey and Sons, Inc.

Amador, A., Papaceit, M., & Juan, E. (2001). Evolutionary change in the structure of the regulatory region that drives tissue and temporally regulated expression of alcohol dehydrogenase gene in *Drosophila funebris. Insect Molecular Biology, 10*(3), 237-247.

Anand, M., Chakraburtty, K., Marton, M. J., Hinnebusch, A. G., & Kinzy, T. G. (2003). Functional interaction between yeast translation eukaryotic elongation factor (eEF) 1A and eEF3. *Journal of Biological Chemistry, 278*(9), 6985-6991.

Andersen, C. B. F., Becker, T., Blau, M., Anand, M., Halic, M., Balar, B., Mielke, T., Boesen, T., Pedersen, J. S., Spahn, C. M. T., Kinzy, T. G., Andersen, G. R., & Beckman, R. (2005). Structure of eEF3 and the mechanism of transger RNA relase from the E-site. *Nature, 443*, 663-668.

Aravind, L. (2001). The WWE domain: a common interaction module in protein ubiquitination and ADP ribosylation. *TRENDS in Biochemical Sciences, 26*(273-275).

Archer, D. B., & Peberdy, J. F. (1997). The molecular biology of secreted enzyme production by fungi. *Critical Reviews in Biotechnology, 17*(4), 273-306.

Arraes, F. B. M., Benoliel, B., Burtet, R. T., Costa, P. L. N., Galdino, A. S., Lima, L. H. A., Marinho-Silva, C., Oliveira-Pereira, L., Pfrimer, P., Procopio-Silva, L., Reis, V. C., & Felipe, M. A. (2005). General metabolism of the dimorphic and pathogenic fugus *Paracocccidioides brasiliensis*. *Genetics and Molecular Research, 2*, 290-308.

Arrese, E. L., Patel, R. T., & Soulages, J. L. (2006). The main triglyceride-lipase from the insect fat body is an active phospholipase A1: identification and characterization. *Journal of Lipid Research, 47*, 2656-2667.

Arvas, M. (2007). *Comparative and functional genome analysis of fungi for development of the protein production host Trichoderma reesei.* VTT Publications.

Bakshi, B. K. (1951). Studies on four species of *Cerotocystis*, with a discussion on fungi causing

sap-stain in Britain. *Mycological Papers, 35*, 1-16.

Ballard, R. G., Walsh, M. A., Cole, W. E. (1982). Blue-stain fungi in xylem of lodgepole pine: a light-microscope study on extent of hyphal distribution. *Canadian Journal of Botany 60*, 2334-2341.

Barlow, J. J., Mathias, A. P., Williamson, R., & Gammack, D. B. (1963). A simple method for the quantitative isolation of undegraded high molecular weight Ribonucleic Acid. *Biochem. Biophys. Res. Commun., 13*, 61-66.

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology, 10*(5), 411-421.

Beckman, C. H. (1956). Production of pectinase, cellulase, and growth-promoting substance by *Ceratostomella ulmi*. *Phytopathology 46*, 605-609.

Becht, P., Vollmeister, E., & Feldbrugge, M. (2005). Role for RNA-ginding protein implicated in pathogenic development of *Ustilago maydis*. *Eukaryotic Cell, 4*(1), 121-133.

Bell, P. J. L., Sunna, A., Gibbs, M. D., Curach, N. C., Nevelainen, H., & Bergquist, P. L. (2002). Prospecting for novel lipase genes using PCR. *Microbiology. , 148*, 2284-2291.

Bennetzen, J. L., & Hall, B. D. (1982). Codon selection in yeast. *J. Biol. Chem, 257*, 3026-3031.

Berg, O. G., & Kurland, C. G. (1997). Growth rate-optimised tRNA abundance and codon usage. *Journal of Molecular Biology, 270*, 544-550.

Berk, A. J. (1999). Activation of RNA polymerase II transcription. *Current Opinion in Cell Biology, 11*, 330-335.

Bezhani, S., Sherameti, I., Pfannschmidt, P., & Oelmuller, R. (2001). A repressor with similarities to prokaryotic and eukaryotic DNA hilicases control the assembly of the CAAT box binding complex at a photosynthesis gene promoter. *Journal of Biological Chemistry, 26*, 23785-23789.

Bhadauria, V., Zhao, W. S., Wang, L. X., Zhang, Y., Liu, J. H., Yang, J., Kong, L. A., & Peng, Y. L. (2007). Advances in fungal proteomics. *Microbiological Research, 162*(3), 193-200.

Binz, T., & Canevascini, G. (1996). Xylanases from the Dutch elm disease pathogens *Ophiostma ulmi* and *Ophiostoma novo-ulmi*. *Physiological and Molecular Plant Pathology, 49*, 159-175.

Binz, T., Gremaud, C., & Canevascini, G. (1997). Production and purification of and extracellular β-galactosidase from the Dutch elm disease fungus *Ophiosoma novo-ulmi. Canadian Journal of Microbiology., 43*, 1011-1016.

Birren, B. (2003). *A white paper for fungal comparative genomics*. Cambridge, MA Whitehead Institute / MIT Center for Genome Research.

Blanchette, R. A., Farrell, R. L., Burnes, T. A., Wendler, P. A., Zimmerman, W., Brush, T. S., & Snyder, R. A. (1992). Biological contro of pitch in pulp and paper production by *Ophiostoma piliferum*. *Tappi Journal, 75*, 102-106.

Borneman, A. R., Hynes, M. J., & Andrianopoulos, A. (2001). The basic helix-loop-helix protein with similarity to the fungal morphological regulators, Phd1p, Efg1p and StuA, control conidiation but not dimorphic growth in *Penicillium marneffei*. *Molecular Microbiology, 44*, 621-631.

Bouvet, G. F., Jacobi, V., Plourde, K. V., & Bernier, L. (2008). Stress-induced mobility of *OPHIO1* and *OPHIO2* DNA transposons of the Dutch elm disease fungi. *Fungal Genetics and Biology, 45*, 565-578.

Bowden, C. G., Hintz, W. E., Jeng, R., Hubbes, M., & Horgen, P. A. (1994). Isolation and characterization of the cerato-ulmin toxin gene of the Dutch elm disease pathogen, *Ophiostoma ulmi*. *Current Genetics, 25*, 323-329.

Bowman, S. M., & Free, S. J. (2006). The structure and synthesis of the fungal cell wall. *Bioessays, 28*, 799-808.

Bradner, R. J., Bergquist, P. L., J., T. o. V. S., & H., N. (2003). The application of PCR for the isolation of a lipase gene from the genomic DNA of an Antarctic microfungus. *Current Genetics, 44*, 224-230.

Braisier, C. M. (1991). *Ophiostoma novo-ulmi* sp. nov., causative agent of Dutch elm disease pandemics. *Mycopathologia, 115*, 151-161.

Braisier, C. M., & Kirk, S. A. (1993). Sibling species within *Ophiostoma piceae*. *Mycological Research, 97*, 811-816.

Braisier, C. M., & Mehrotra, M. D. (1995). *Ophiostoma himl-ulmi* sp. nov, a new species of Dutch elm disease fungus endemic to the Himalayas. *Mycological Research, 99*, 205-215.

Brush, T. S., Chapman, R., Kurzman, R., & Williams, D. P. (1999). Purification and characterization of extracellular Lipases From *Ophiostoma piliferum*. *Bioorganic and Medicinal Chemistry, 7*, 2131-2138.

Brush, T. S., Farrell, R. L., & Ho, C. (1994). *Tappi Journal, 77*, 155.

Buchan, J. R., Aucott, L. S., & Stansfield, I. (2006). tRNA properties help shape codon pair preference in open reading frames. *Nucleic Acids Research, 34*(3), 1015-1027.

Büschges, R., Bahrenberg, G., Zimmermann, G., & Wolf, K. (1994). NADH:unbiquinone oxidoreductase in obligate aerobic yeasts. *Yeast, 10*, 475-479.

Bustin, S. A. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology, 29*, 23-39.

Bustin, S. A., & Nolan, T. (2004). Pitfalls of Quantitative Real-Time Reverse-Transcription Polymerase Chain Reaction. *Journal of Biomolecular Techniques, 15*(3), 155-166.

Calderone, R. A. (2002). *Candida and Candidiasis*: ASM Press.

Calero-Rueda, O., Plou, F. J., Ballesteros, A., Martinez, A. T., & Martinez, M. J. (2002). Production, isolation and characterization of a sterol esterase from *Ophiostoma piceae*. *Biochimica et biophysica acta. Proteins and proteomics, 1599*, 1-2, 28-35.

Carbone, A., Zinovyev, A., & Képès, F. (2003). Codon adaptation index as a measure of dominating codon bias. *Bioinformatics, 19*(16), 2005-2015.

Carlsen, M., Nielsen, J., Villadsen, J. (1996) Growth and α-amylase production by *Aspergillus oryzae* during continuous cultivations. *Journal of Biotechnology 45*(1), 81-93.

Castillo, L., Martines, A. I., Garcera, A., Elorza, M. V., Valentin, E., & Sentandreu, R. (2003). Functional analysis of the cysteine residues and the repetitive sequence of *Saccharomyces cerevisiae* Pir4/Cis3: The repetitive sequence is needed for binding to the cell wall beta-1,3-glucan *Yeast, 20*, 1291-1306.

Chambergo, F. S., Bonaccorsi, E. D., Ferreira, A. J. S., Ramos, A. S. P., Ferreira, J. R., Abrahão-Neto§, J., Farah, J. P. S., & El-Dorry, H. (2002). Elucidation of the metabolic fate of glucose in the filamentous fungi *Trichoderma reesei* using expressed sequence tag (EST) analysis and cDNA microarrays. *Journal of Biological Chemistry, 277*(16), 13983-13988.

Chou, Q., Russell, M., Birch, D. E., Raymond, J., & Bloch, W. (1992). Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Research, 20*, 1717-1723.

Clason, T., Zickermann, V., Ruiz, T., Brandt, U., & Radermacher, M. (2007). Direct localization of the 51 and 24 kDa subunits of mitochondrial complex I by three-dimensional difference imaging. *Journal of Structural Biology, 159*, 433-442.

Coghlan, A., & Wolfe, K. H. (2000). Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. *Yeast, 16*, 1131-1145.

Cornell, M. J., Alam, I., Soanes, D. M., Wong, H. M., Hedeler, C., Paton, N. W., Rattra, M., Hubbard, S. J., Talbot, N. J., & Oliver, S. G. (2008). Comparative genome analysis across a kingdom of eukaryotic organisms: Specialization and diversification in the Fungi. *Genome Research*, 1809-1821.

Cranswick, A. M., Rook, D. A., Zabkiewicz, J. A. (1987). Seasonal changes in carbohydrate concentration and composition of different tissue types of *Pinus radiate* trees. *New Zealand Journal of Forestry Science 17*, 229-245.

Cuomo, C. A., Güildener, U., Xu, J., Trail, F., Turgeon, B. G., Pietro, A. D., Walton, J. D., Ma, L., Baker, S., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y., DeCaprio, D., Gale, L. R., Gnerre, S., Goswami, R. S., Hammond-Kosack, K., Harris, L. J., Hilburn, K., Kennell, J. C., Kroken, S., Magnuson, J. K., Mannhaupt, G., Mauceli, E., Mewes, H. W., Mitterbauer, R., Mauelbauer, G., Münsterkötter, M., Nelson, D., O'Donnell, K., Ouellet, T., Qi, W., Quesneville, H., Isabel, M., Roncero, G., Seong, K., Tetko, I. V., Urban, M., Waalwijk, C., Ward, T. J., Yao, J., Birren, B. W., & Kistler, H. C. (2007). The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization. *Science, 317*, 1400-1402.

Curan, J. F., & Yarus, M. (1989). Rates of aminoacytl-tRNA selection at 29 sense codons *in vivo*. *Journal of Molecular Biology, 209*, 65-77.

Cygler, M., Schrag, J. D., Sussman, J. L., Harel, M., Silman, I., Gentry, M. K., & Doctor, B. P. (1993). Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. *Protein Science, 2*, 366-382.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., Davis, R. W., & Steinmetz, L. M. (2006). A high-resolution map of transcription in the yeast genome. *PNAS, 103*(14), 5320-5325.

Davidson, R. W. (1958). Additional species of *Ophiostomataceae* from Colorado. *Mycologia, 50*, 661-670.

de Hoog, G. S. (1993). *Sporothrix-like anamorphs of Ophiostoma species and other fungi.* : APS Press, St. Paul. Minnesota.

de Hoog, G. S., & Scheffer, R. J. (1984). *Ceratocystis* versus *Ophiostoma*: a reappraisal. *Mycologia, 76*, 292-299.

Dekkers, K. L., You, B. J., Gowda, V. S., Liao, H. L., Lee, M. H., Ueng, P. P., & Chung, K. R. (2006). The *Cercospora nicotianae* gene encoding dual *O*-methyltransgerase and FAD-dependent monooxygenase domains mediated cercosporin toxin biosynthesis. *Fungal Genetics and Biology, 44*(5), 444-454.

Del Sorbo, G., Scala, F., Parrella, G., Lorito, M., Comparini, C., Ruocco, M., & Scala, A. (2000). Functional expression of the gene cu encoding the phytotoxic hydrophobin cerato-ulmin, enables *Ophiostoma quercus*, a nonpathogen on Elm to cause symptoms of Dutch Elm disease. *The American Phytopathological Society 13*(1), 43-53.

Derewenda, U., Brzozowski, A. M., Laswon, D. M., & Derewenda, Z. S. (1992). Catalysis at the interface: the anatomy of a conformational change in a triglyceride lipase. *Biochemistry, 11*, 1532-1541.

Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, K., N., G., Sverdlov, E. D., & Seibert, P. D. (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *PNAS, 93*, 6025-6030.

Dietrich, F. S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S., Wing, R. A., Flavier, A., Gaffney, T. D., & Philippsen, P. (2004). The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science, 304*(5668), 304-307.

DiGuistini, S., Ralph, S. G., Lim, Y. W., Holt, R., Jones, S., Bohlmann, J., & Breuil, C. (2007). Generation and annotation of lodgepole pine and oleoresin-induced expressed sequences from the blue-stain fungus *Ophiostoma clavigerum*, a Mountain Pine Beetle-associated pathogen. *FEMS Microbiological Letters, 267*(2), 151-158.

Dogra, N., & Breuil, C. (2004). Suppresive subtractive hybridization and differential screening identified gene differentially expressed in yeast and mycelial form of *Ophiostoma piceae*. *FEMS Microbiological Letters, 238*, 175-181.

Drabløs, F., & Petersen, S. B. (1997). Identification of conserved residues in family of esterase and lipase sequences. *Methods in Enzymology, 284*, 28-61.

Duarte, M., Peters, M., Schulte, U., & Videira, A. (2003). The internal alternative NADH dehydrogenase of *Neurospora crassa* mitochondria. *Biochemical Journal, 371*, 1005-1011.

Duret, L., & D., M. (1999). Expression pattern and, surprisingly, gene lengh shape codon usage

in Caenorhabditis, Drosphila, and Arabidopsis. *Proc Natl Acad Sci USA, 96*, 4482-4487.

Earnshaw, W. C., Martins, L. M., & Kaufmann, S. H. (1999). Mammalian caspases: Structure, activation, substrates, and functions during apoptosis. *Annual Review of Biochemistry, 68*, 384-424.

Ebbole, D. J., Jin, Y., Thon, M., Pan, H., Bhattarai, E., Thomas, T., & Dean, R. (2004). Gene discovery and gene expression in the rice blast fungus, *Magnaporthe grisea*: Analysis of Expressed Sequence Tags. *Molecular Plant-Microbe Interactions, 17*(12), 1337-1347.

Eckart, M. R., Bussineau, C. M. (1996). Quality and authenticity of heterologous protein synthesized in yeast. *Current Opinion in Biotechnology 7*(5), 525-530

Emri, T., Molnar, Z., & Szilagyi, M. (2008). Regulation of autolysis in *Aspergillus nidulans*. *Applied Biochemistry and Biotechnology.*

Erikkson, K. E., Blanchette, R. A., & Ander, P. (1990). Microbial and enzymatic degradation of wood and wood components. *Springer-Verlag berlin Heidelberg* 407.

Eyre-Walker, A. (1996). Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? *Mol Biol Evol, 13*, 864-872.

Farrell, R. L., Blanchette, R. A., Brush, T. S., Gysin, B., Hadar, Y., Perollaz, J. J., & Wendler, P. A. (1992, May 27-30). *"Cartapip" A biopulping priduct for control of pitch and resin acid problems in pulp mills.* Paper presented at the 5th International Conference on Biotechnology in the Pulp and Paper Industry, Kyoto, Japan.

Farrell, R. L., Blanchette, R. A., Brush, T. S., Hadar, Y., Iverson, S., Krisa, K., Wendler, P. A., & Zimmerman, W. (1993). Cartapip TM: a biopulping product for control of pitch and resin acid problems in pulp mills. *Journal of Biotechnology, 30*, 115-122.

Farrell, R. L., Hadar, E., Kay, S. J., Blanchette, R. A., & Harrington, T. C. (1997). *Survey of sapstain organism in New Zealand and albino anti-sapstain fungi.* Paper presented at the Biology and Prevention of Sapstain, Whistler, Canada.

Fengel, D., & Wegener, G. (1989). *Wodd: Chemistry, Ultrastructure, Reactions*. New York.

Fessele, S., Maier, H., Zischek, C., Nelson, P. J., & Werner, T. (2002). Regulatory context is a crucial part of gene function. *Trends in Genetics, 18*, 60-63.

Finn, R. D., Mistry, J., Schuster, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L. L., & Bateman, A. (2006). Pfam: clans, web tools and services. *Nucleic Acids Research, Database Issue 34*, D247-D251.

Flores, C. L., Rodriguez, C., Petit, T., & Gancedo, C. (2000). Carbohydrate and energy-yielding metabolism in non-conventional yeasts. *FEMS Microbiological Review 24*, 507-529.

Fojan, P., Jonson, P. H., Petersen, M. T. N., & Peterson, S. B. (2000). What distinguishes an esterase from a lipase: A novel structural approach. *Biochimie, 82*, 1022-1041.

Friedrich, T., & Böttcher, B. (2004). The gross structure of the respiratory complex I: a Lego System. *Biochimica et Biophysica Acta, 1608*, 1-9.

Friedrich, T., & Scheide, D. (2000). The respiratory complex I of Bacteria, Archaea, and Eucarya and its module common with membrane-bound multisubunit hydrogenases

FEBS Letters, 479, 1-5.

Fukumura, R., Takahashi, H., Saito, T., Tsutsumi, Y., Fujimori, A., Sato, S., Tatsumi, K., Araki, R., & Abe, M. (2003). A sensitive transcriptome analysis method that can detect unkown transcripts. *Nucleic Acids Research, 31*(16), 94.

Futcher, B., Latter, G. I., Monardo, P., Mclaughlin, C. S., & Garrels, J. I. (1999). A sampling of the Yeast Proteome. *Molecular and Cellular Biology, 19*(11), 7357-7368.

Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R., Batzoglou, S., Lee, S. I., Basturkment, M., Spevak, C. C., Clutterbuck, J., Kapitonov, J., Jurka, J., Scazzoxxhio, C., Frarman, M., Butler, J., Purcell, S., Harris, S., Braus, G. H., Draht, O., Busch, S., DÉnfert, C., Bouchier, C., Goldman, G. H., Bell-Pedersen, D., Griffiths-Jones, S., Doonan, J. H., Yu, J., Vienken, K., Pain, A., Freitag, M., Selker, E. U., Archer, D. B., Penalva, M. A., Oakley, B. R., Momany, M., Tanaka, T., Kumagai, T., Asai, K., Machida, M., Nierman, W. C., Denning, D. W., Caddick, M., Hynes, M., Paoletti, M., Fischer, R., Miller, B., Dyer, P., Sachs, M. S., Osmani, S. A., & Birren, B. W. (2005). Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. orzyae*. *Nature, 438*(7071), 1105-1115.

Gao, Y., & Breuil, C. (1995). Extracellular lipase production by a sapwood-staining fungus, *Ophiostoma piceae*. *World Journal of Microbiology and Biotechnology, 11*, 638-642.

Gao, Y., & Breuil, C. (1998). Properties and substrate specificities of an extracellular lipase purified from *Ophiostoma piceae*. *World Journal of Microbiology and Biotechnology. , 14*, 421-429.

Gasch, A. P., & Werner-Washburne, M. (2002). The genomics of yeast responses to environmental stress and starvation. *Funct Integr Genomics, 2*, 181-192.

Ghaemmaghami, S., Huh, W., Bower, K., Howson, R. W., Bell, A., Dephoure, N., O'Shea, E. K., & Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature, 425*, 737-741.

Gharibian, S., Hoffer, C., Abraham, L. D., & Breuil, C. (1996). Localizing an *Ophiostoma piceae* proteinase in sapwood of four tree species using polyclonal antibodies *New Phytologist 133*, 673-679.

Ginzinger, D. (2002). Gene quantification using real-time quantitative PCR:an emerging technology hits the mainstream. *Experimetnal Hematology, 30*, 503-512.

Gold, S. E., Duncan, G., Barrett, K., & Kronstad, J. (1994). cAMP regulated morphogneesis in the fungal pathogen *Ustilago maydis*. *Genes Dev, 8*, 2805-2816.

Gordon, C. L., Archer, D. B., Jeenes, D. J., Doonan, J. H., Wells, B., Trinci, A. P. J., & Robson, G. D. (2000). A glucoamylase:GFP fusion to study protein secretion by individual hyphae of *Aspergillus niger*. *Journal of microbiological methods, 42*(1), 39-48.

Green, F. I., Clausen, C. A., Larsen, M. J., & Highley, T. L. (1992). Immuno-scanning electron microscopic localization of extracellular wood-degrading enzymes within the fibrillar sheath of the brown-rot fungus *Postia placenta*. *Canadian Journal of Microbiology 38*, 898-904.

Gubb, E. A., & Hausner, G. (2005). Optional mitochondrial introns and evidence for a homing-

endonuclease gene in the mtDNA *rnl* gene in *Ophiostoma ulmi s. lat. Mycological Research, 109*(10), 1112-1126.

Gustafsson, C., Govindarajan, S., & Minshull, J. (2004). Codon bias and heterologous protein expression. *TRENDS in Biotechnology, 22*(7).

Halmschlager, E., Messner, R., Kowalski, T., & Prillinger, H. (1994). Differentiation of *Ophiostoma piceae* and *Ophiostoma quercus* by morphology and RAPD analysis. *Systematic and Applied Microbiology, 17*, 554-562.

Halsted, B. D. (1890). Some fungus diseases of sweet potato. *New Jersey Agricultural College Experiment Station Bulletin, 76*, 3-32.

Harrington, T. C. (1981). Cycloheximide sensitivity as a taxonomic character in *Ceratocystis*. *Mycologia, 72*, 1123-1129.

Harrington, T. C. (1987). New combinations in *Ophiostoma* of *Ceratocustis* species with *Leptographium* anaporphs. *Mycotaxon, 28*, 39-43.

Harrington, T. C., McNew, D., Steimel, J., Hofstra, D., & Farrell, R. L. (2001). Phylogeny and taxonomy of the *Ophiostoma piceae* complex and the Dutch elm disease in fungi. *Mycologia, 93*(1), 111-136.

Haryati, T. (2001). *Optimisation of biocontrol of Pinus radiata; extractive degradation using albino Ophiostma sp. fungi* The University of Waikato, Hamilton.

Hausner, G., & Reid, J. (2003). Notes on *Ceratocystis brunnea* and some other *Ophiostoma* species based on partial ribosomal DNA sequence analysis. *Canadian Journal of Botany, 81*, 864-876.

Hausner, G., Reid, J., & Klassen, G. R. (1993a). On the subdivision of *Ceratocystis s.l.*, based on partial ribosomal DNA sequences. *Canadian Journal of Botany., 71*, 52-63.

Hausner, G., Reid, J., & Klassen, G. R. (1993b). *Ceratocystiopsis*: a reappraisal based on molecular criteria. *Mycological Research, 97*, 625-633.

Hawksworth, D. L. (1997). The fascination of fungi: exploring fungal diversity. *Mycologist, 11*, 18-22.

He, H., von der Haar, T., Singh, C. R., Ii, M., Li, B., Hinnebusch, A. G., McCarthy, J. E. G., & Asano, K. (2003). The yeast eukaryotic initiation factor 4G (eIF4G) HEAT domain interacts with eIF1 and eIF5 and is involved in stringent AUG selection. *Molecular and Cellular Biology, 23*(15), 5431-5445.

Held, B. W., Thwaites, J. M., Farrell, R. L., & Blanchette, R. A. (2003). Albino Strains of Ophiostoma Species for Biological Control of Sapstaining Fungi. *Holzforschung,, 57*, 237-242.

Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor? . *Genes Dev., 7*, 1291-1308.

Hoffman, B., & Breuil, C. (2004). Analysis of the distribution and regulation of three representative subtilase genes in sapstaining fungi. *Fungal Genetics and Biology, 41*, 168-175.

Hofmann, K., Bucher, P., Falquet, L., & Bairoch, A. (1999). The PROSITE database, its status in 1999. *Nucleic Acids Research, 27*, 215-219.

Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., M.R., G., Golub, T. R., Lander, E. S., & Young, R. A. (1998). Dissecting the regulatory circuitry of a eukaryote genome. *Cell, 95*, 717-728.

Hubbes, M. (1989). Pathogenic virulence and host resistance in Dutch elm disease. *Naturalist Canada, 115*, 157-161.

Hudson, H. J. (1986). *Fungal Biology*. Baltimore, USA.

Hunt, J. (1956). Taxonomy of the genus *Ceratocystis*. *Lloydia, 19*, 1-58.

Idnurm, A., & Howlett, B. J. (2001). Pathogenicity islands of phytopathogenic fungi. *Molecular Plant Physiology, 2*, 241-255.

Ikemura, T. (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Difference in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. *Journal of Molecular Biology, 158*, 573-597.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution, 2*, 13-34.

Iseli, C., Jongeneel, C. V., & Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences *Proc Int Conf Intell Syst Mol Biol.*, 138-148.

Jaeger, K. E., Dijkstra, B. W., & Reetz, M. T. (1999). Bacterial biocatalysts: Molecular biology, three dimensional structures, and biotechnological applications of lipases. *Annual Review of Microbiology, 53*, 315-351.

Jaeger, K. E., Ransac, S., Dijkstra, B. W., & Reetz, M. T. (1994). Bacterial Lipases. *FEMS Microbiological Review, 15*, 29-63.

Jaillon, O., Bouhouche, K., Gout, J. F., Aury, J. M., Noel, B., Saudemont, B., Nowacki, M., Serrano, V., Porcel, B. M., Segurens, B., Le Mouel, A., Lepere, G., Schachter, V., Betermier, M., Cohen, J., Wincker, P., Sperling, L., Duret, L., & Meyer, E. (2008). Translational control of intron splicing in eukaryotes. *Nature, 451*(7176), 359-362.

Jansen, R., Bussemaker, H. J., & Gerstein, M. (2003). Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Research 31*(8), 2242-2251.

Jansen, R., & Gerstein, M. (2000). Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins. *Nucleic Acids Research, 28*, 1481-1488.

Janssen, P. H., Monk, C. R., & Morgan, H. W. (1994). A thermophilic, lipolytic *Bacillus* sp., and continuous assay of it *p*-nitrophenyl palmitate esterase activity. *FEMB Microbiology Letter, 120*, 195-200.

Jeng, R., Huang, C., Sain, M., Hubbes, M., Rodriguez, A., & Savile, B. (2007). Starch-like exopolysaccharide produced by the filamentous fungi *Ophiostoma ulmi* and *O. novo-ulmi*. *Forest Pathology, 37*, 80-95.

Jensen, L. J., Gupta, R., Blom, N., Devos, D., Tamames, J., Kesmir, C., Nielsen, H., Staerfeldt, H. H., Rapacki, K., Workman, C., Andersen, C. A. F., Knudsen, S., Krogh, A., Valencia, A., & Brunak, S. (2002). Ab initio prediction of human orphan protein function from post-translational modifications and localization features. *Journal of Molecular Biology, 319*, 1257-1265.

Jensen, L. J., Staerfeldt, H. H., & Brunak, S. (2003). Prediction of human protein function according to Gene Ontology categories. *Bioinformatics, 19*, 635-642.

Johnston, J. R. (1994). *Molecular Genetics of Yeast, a Practical Approach*: IRL Press, Oxford.

Jonegeneel, C. V. (2000). Searching the expressed sequence tag (EST) databases: Panning for genes. *Briefings in Bioinformatics 1*(1), 76-92.

Joseph-Horne, T., Hollomon, D. W., & Wood, P. M. (2001). Fungal respiration: a fusion of standard and alternative components. *Biochimica et Biophysica Acta, 1504*, 179-195.

Jukes, T. H., & Osawa, S. (1990). The gentic code in mitochondria and chloroplasts. *Experientia, 46*, 1117-1126.

Kalisz, H. M., Wood, D. A., & Moore, D. (1987). Production, regulation and release of extracellular proteinase activity in Basidiomycete fungi. *Transactions of the British Mycological Society 88*, 221-227.

Kamatani, T., & Yamamoto, T. (2007). Comparison of codon usage and tRNAs in mitochondrial genomes of *Candida* species. *Biosystems, 90*, 362-370.

Kapranov, P., Willingham, A. T., & Gingerase, T. R. (2007). Genome-wide transcription and the implication for genomic organization. *Nature Reviews Genetics, 8*, 413-423.

Karlin, S., Mrazek, J., & Campbell, A. M. (1998). Codon usages in different gene classes of the Eschericia coli genome. *Molecular Microbiology, 29*, 1341-1355.

Kato, T., Morita, M. T., & Fukaki, H. (2002). SGR2, a phospholipase-like protein and ZIG/SGR4, a SNARE, are involved in the shoot gravitropism of Arabidopsis. *Plant Cell., 14*, 33-46.

Kayali, H. A., Tarhan, L., & Soran, H. (2005). Variations of alcohol dehydrogenase activity and fermentative pyruvate, ethanol production of *F. equiseti* and *F. acuminatum* depend on the yeast extract and urea concentrations. *Enzyme and Microbial Technology 36*, 706-711.

Kenealy, W. R., & Dietrich, D. M. (2004). Growth and fermentation responses of *Phanerochaete chrysosporium* to O2 limitation. *Enzyme and Microbial Technology 34*, 490-498.

Kerscher, S. (2000). Diversity and origin of alternative NADH:ubiquinone oxidoreductases. *Biochimica et Biophysica Acta, 1459*, 274-283.

Kerscher, S., Okun, J. G., & Brandt, U. (1999). A single ecternal enzyme confers alternative NADH:ubiquinone oxidoreductase activity in *Yarrowia lipolytica*. *Journal of Cell Science, 112*, 2347-2354.

Kerscher, S. J., Eschemann, A., Okun, P. M., & Brandt, U. (2001). External alternative NADH:ubiquinone oxidoreductase redirected to the internal face of the mitochndrial inner membrane rescues complex I deficiency in *Yarrowia lipolytica*. *Journal of Cell Science, 114*(21), 3915-3921.

Kim, J. J., Allen, E. A., Humble, L. M., & Breuil, C. (2005). Ophiostomatoid and basidiomycetous fungi associated with green, red, and grey lodgepole pines after mountian pine beetles (*Dendroctonus pnderosae*) infestation. *Canadian Journal of Forest Research, 35*, 274-284.

Kim, S. H., Uzunovic, A., & Breuil, C. (1999). Rapid detection of *Ophiostoma piceae* and *O. quercus* in stained wood by PCR. *Applied and Environmental Microbiology, 65*, 287-290.

Kirk, P. M., Cannon, P. F., David, J. C., & Stalpers, J. A. (2001). *Ainsworth and Bisby's Dictionary of the Fungi* (9th Edition ed.).

Knorr, D., Shetty, K. J., Hood, L. F., & Kinsella, J. E. (1979). An enzymatic method FOB yeast autolysis. *Journal of Food Science 44*(5), 1362-1365.

Köhler, G. A., Brenot, A., Hass-Stapleton, E., Agabian, N., Deva, R., & Nigam, S. (2006). Phospholipase A2 and phosphlipase B activities in fungi. *Biochemica et Biophysica Acta, 1761*(11), 1391-1399.

Köhler, G. A., Brenot, A., Hass-Stapleton, E., Agabian, N., Deva, R., & Nigam, S. (2006). Phospholipase A2 and Phospholipase B activities in fungi. *Biochimica et Biophysica Acta 1761*, 1391-1399.

Kost, T. A. (1999). Expression vectors and delivery systems: Tools for determining gene function and gene therapy Editorial overview. *Current Opinion in Biotechnology 10*(5), 409-410.

Kupfer, D. M., Reece, C. A., Clifton, S. W., Roe, B. A., & Prade, R. A. (1997). Multicellular ascomycetous fungal genomes contain more than 8000 genes. *Fungal Genetics and Biology, 21*, 364-372.

Lafay, B., Lloyd, A. T., McLean, M. J., Devine, K. M., Sharp, P. M., & Wolfe, K. H. (1999). Proteome composition and codon usage in spirochaetes: species-specific and DNA strand-specific mutational biases. *Nucleic Acids Research 27*, 1642-1649.

Lang, C., & Looman, A. C. (1995). Efficient expression and secretion of *Aspergillus niger* polygalacturonase in *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology, 44*, 147-156.

Lekanne Deprez, R. H., Fijnvandraat, A. C., Ruijter, J. M., & Moorman, A. F. M. (2002). Sensitivity and accuracy of quantitative real-time polymerase chain reaction using SYBR green I depends on cDNA synthesis conditions. *Analytical Biochemistry, 307*(1), 63-69.

Lithwick, G., & Margalit, H. (2003). Hierachy of sequence-dependent features associated with

prokaryotic translation. *Genome Research, 13*, 2665-2673.

Liu, W., & Saint, D. A. (2002). A new quantitative method of Real Time Reverse Trascription Polymerase Chain Reaction assay based on simulation of polymerase chain reaction kinetics. *Analytical Biochemistry, 302*, 52-59.

Lloyd, A. T., & Sharp, P. M. (1991). Codon usage in *Aspergillus nidulans*. *Molecular and General Genetics, 230*, 284-294.

Lloyd, A. T., & Sharp, P. M. (1992). Evolution of codon usage patterns: the extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*. *Nucleic Acids Research, 20*(20), 5289-5295.

Lo, H. J., Kohler, J. R., DiDomenico, B., Loebenberg, D., Cacciapuoti, A., & Fink, G. R. (1997). Nonfilamentous *C. albicans* mutants are avirulent. *Cell, 90*, 939-949.

Lockett, T. J., & Ashburner, M. (1989). Temporal and spatial utilization of the alcohol dehydrogenase gene promoters during the development of *Drosophila melanogaster*. *Develompmental Biology, 134*, 430-437.

Loria, P. M., Duke, A., Rand, J. B., & Hobert, O. (2003). Two neuronal, nuclear-localized RNA binding proteins involved in synaptic transmission. *Current Biology, 13*, 1317-1323.

Luttik, M. A. H., Overkamp, K. M., P., K., de Vries, S., van Dijken, P., & Pronk, J. T. (1998). The *Saccharomyces cerevisiae NDE1* and *NDE2* genes encode separate mitochondrial NADH dehydrogenases catalyzing the oxidation of cytosolic NADH. *Journal of Biological Chemistry, 273*, 24529-24534.

Lynch, M. (2006). The origins of eukaryotic gene structure. *Molecular Biology and Evolution, 23*(2), 450-468.

Madhani, H. D., & Fink, G. R. (1998). The control of filamentous differentiation and virulence in fungi. *Trends in Cell Biology, 8*, 348-353.

Madigan, M. T., & Martinko, J. M. (2006). *Brock. Biology of Microorganisms* (11 ed.): Pearson Prentice Hall, Pearson Education, Inc.

Mannhaupt, G., Montrone, C., Haase, D., Mewes, H. W., Aign, V., Hoheisel, J. D., Fartmann, B., Nyakatura, G., Kempken, F., Maier, J., & Schulte, U. (2003). What's in the genome of a filamentous fungus? Analysis of the *Neurospora* genome sequence. *Nucleic Acids Research, 31*(7), 1944-1954.

Martinez, D., Larrondo, L. F., Putnum, N., Gelpke, M. D. S., Huang, K., Chapman, J., Helfenbein, K. G., Ramaiya, P., C., D. J., Larimer, F., Coutinho, P. M., Henrissat, B., Berka, R., Cullen, D., & Rokhsar, D. (2004). Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature Biotechnology*.

Mathews, B., Sonenberg, N., & Hershey, J. W. B. (1996). *Origins and targets of translational control*. New York: Cold Spring Harbor Laboratory Press.

Mathiesen-Käärik, A. (1960). Studies on the ecology, taxonomy and physiology of Swedish insect-associated blue stain fungi. *Oikos, 11*, 1-25.

McKenzie, E. H. C. e. (2004). *Introduction to Fungi of New Zeland* (1 ed.): Fungal Diversity

Press, Hong Kong.

McNaughton, C. L. (1997). *The purification and characterisation of an extracellular lipase from Ophiostoma piliferum.*, The University of Waikato, Hamilton.

Medlin, L., Elwood, H. J., Stickel, S., Sogin, M. L. (1988). The characterization of enzymatically amplified eukaryotic 16S-like rRNA coding regions. *Gene, 71*, 491-499.

Melin, E., & Nannfeldt, J. A. (1934). Researches into the blueing of ground wood-pulp. *Svenska Skogsvårdsföreningens Tidskrift, 32*, 397-616.

Mignone, F., Gissi, C., Liuni, S., & Pesole, G. (2002). Untranslated regions of mRNAs. *Genome Biology, 3*(3), 0004.0001-0004.0010.

Min, X. J., Butler, G., Storms, R., & Tsang, A. (2005). OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research, 33*, W667-W680.

Moller, I. M. (2001). Plant mitochondria and oxidative stress: electron transport, NADPH turnover, and metabolism of reactive oxygen species. *Annual Review of Plant Physiology and Plant Molecular Biology, 52*, 561-591.

Morin, C., Coutrier, S., & Bernier, L. (2007). Pathogenicity of wild-type and albino strains of the fungus *Ceratocystis resinifera*, a potential biocontrol agent against bluestain. *Canadian Journal of Forestry Research, 37*, 919-930.

Morin, M., Monteoliva, L., Insenser, M., Gil, C., & Dominquez, A. (2007). Proteomic analyis reveals metabolic changes during yeast to hypha transition in *Yarrowia lipolytica*. *Journal of Mass Spectrometry 42*(11), 1453-1462.

Moriyama, E. N., & Powell, J. R. (1998). Gene length and codon usage bias in *Drosophila melanogaster, Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Research, 26*, 3188-3193.

Morlet, M. (1992). *Ophiostoma querci* sur chene en France. *Ann Soc Sci Nat Archeol Toulon Var, 44*, 109-112.

Moszer, I., Rocha, E. P. C., & Danchin, A. (1999). Codon usage and lateral gene transfer in Bacillus subtilis *Current Opinion in Microbiology, 2*, 524-528.

Muyzer, G., De Waal, E. C., Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*. *59*, 695-700.

Mouton, M., Wingfield, M. J., & Van Wyk, P. W. (1994). Condidium development in anamorphs of *Ceratocystis sensu lato*: a review. *South African Journal of Science, 90*, 293-298.

Naranda, T., MacMillan, S. E., Donahue, T. F., & Hershey, J. W. B. (1996). SUI1/p16 Is required for the activity of eukaryotic translation initiation factor 3 in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology, 16*(5), 2307-2313.

Nelson, G., Kozlova-Zwinderman, O., Collis, A. J., Knight, M. R., Fincham, J. R. S., Stanger, c. P., Renwick, A., Hessing, J. G. M., Punt, P. J., Van den Hondel, C. A. M. J. J., & Read, N. D. (2004). Calcium measurement in living filamentous fungi expressing codon-

optimized aequorin. *Molecular Microbiology, 52*(5), 1437-1450.

Nevalainen, K. M. H., Bergquist, P. L., Te'o, V. S. J., & Farrell, R. A. (2004). Fungal host for expression and production of recombinant products. *Australian provisional patent 1405967*

Niimi, M. (1996). Dibutyryl cyclic AMP-enhance germ tube formation in exponential growing *Candida albicans* cells. *Fungal Genetics and Biology, 20*, 79-83.

Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G., & Burley, S. K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proceedings of the National Academy of Sciences, 93*, 4862-4867.

Novina, C. D., & Roy, A. L. (1996). Core promoters and transcriptional control. *Trends in Genetics, 12*, 351-355.

Okada, G., Seifert, K. A., Takematsu, A., Yamaoka, Y., Miyazaki, S., & Tubaki, K. (1998). A molecular phylogenetic reappraisal of the *Graphium* complex based on 18S rDNA sequences *Canadian Journal of Botany, 76*, 1495-1506.

Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G., & Goldman, A. (1992). The alpha/beta hydrolase fold. *Protein Engineering, 5*(3), 197-211.

Orlando, C., Pinzani, P., & Pazzagli, M. (1998). Developments in quantitative PCR. *Clinical Chemistry Laboratory Medicine, 36*, 255-269.

Ouellette, G. B., Chamberland, H., Goulet, A., Lachapelle, M., & Lafontaine, J. G. (1999). Fine structure of the extracellular sheath and cell walls in *Ophiostoma novo-ulm*i growing on various substrates. *Canadian Journal of Microbiolgoy 45*(7), 582-597.

Overkamp, K. M., Bakker, B. M., Kotter, P., Luttik, M. A., van Dijken, J. P., & Pronk, J. T. (2002). Metabolic engineering of glycerol production in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology, 68*, 2814-2821.

Palade, G. (1975). Intracellular aspects of the process of protein synthesis. *Science, 189*, 347-357.

Paoletti, M., Buck, K. W., & Braisier, C. M. (2005). Cloning and sequence analysis of the MAT-B (MAT-2) genes from the three Dutch elm disease pathogens, *Ophiostoma ulmi*, *O. novo-ulmi*, and *O. himal-ulmi Mycological Research 109*(9), 983-991.

Pavlova, K., L., K., Kratchanova, M., & Panchev, I. (2004). Production and characterization of an exopolysaccharide by yeast. *World Journal of Microbiology and Biotechnology, 20*, 435-439.

Payne, G. A., Nierman, W. C., Wortman, J. R., Pritchard, B. L., Brown, D., Dean, R. A., Bhatnager, D., Cleveland, T. E., Machida, M., & Yu, J. (2006). Whole genome comparison of *Aspergillus flavus* and *A. oryzae*. *Medical mycology: official publication of the International Society for Hunam and Animal Mycology, 44*, 9-11.

Peberdy, J. F. (1999). Extracellular proteins in fungi: a cytological and molecular perspective. *Acta Microbiologica et Immunologica Hungarica, 46*(2-3), 165-174.

Pel, H. J., de Winde, J. H., Archer, D. B., Dyer, P. S., Hofmann, G., Schapp, P. J., Turner, G., de

Vries, R. P., Albang, R., Albermann, K., Andersen, M. R., Bendtsen, J. D., Benen, J. A. E., van den Berg, M., Breestraat, S., Caddick, M. X., Contreras, R., Cornell, M., Coutinho, P. M., Danchin, E. G. J., Debets, A. J. M., Dekker, P., van Dijck, P. W. M., van Dijck, A., Dijkhuizen, L., Driessen, A. J. M., dÉnfert, C., Geysens, S., Goosen, C., Groot, G. S. P., de Groot, P. W. J., Guillemette, T., Henrissat, B., Herweijer, M., van den Hombergh, J. P. T. W., van den Hondel, C. A. M. J. J., van der Heijden, R. T. J. M., van der Kaaij, R. M., Klis, F. M., Kools, H. J., Kubicek, C. P., van Kuyk, P. A., Lauber, J., Lu, X., van der Maarel, M. J. E. C., Meulenberg, R., Menke, H., Mortimer, M. A., Neilsen, J., Oliver, S. G., Olsthoorn, M., Pal, K., van Peij, N. N. M. E., Ram, A. F. J., Rinas, U., Roubos, J. A., Sagt, C. M. J., Schmoll, M., Sun, J., Ussery, D., Varga, J., Vervecken, W., van de Vondervoort, P. J. J., Wedler, H., Wösten, H. A. B., Zeng, A., van Ooyen, A. J. J., Visser, J., & Stam, H. (2007). Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nature Biotechnology, 25*(2).

Pereira, V., Royer, J. C., Hintz, W. E., Field, D., Bowden, C., Kokurewicz, K., Hubbes, M., & Horgen, P. A. (2000). A gene associated with filamentous growth in *Ophistoma novo-ulmi* has RNA-binding motifs and is similar to a yeast gene involved in mRNA splicing. *Current Genetics, 37*, 94-103.

Perez-Leblic, M. I., R., F. Marines, M. J., & Lahoz, R. (1982). Cell wall degradation in the autolysis of filamentous fungi. *Mycopathologia, 80*(3), 147-155.

Pikaart, M. J., & Villeponteau, B. (1993). Suppresion of PCR amplification by high level of RNA. *Biotechniques, 18*, 225-226.

Pipe, N. D., Buck, K. W., & Brasier, C. M. (1995). Genomic fingerprinting supports the separation of *Ophiostoma piceae* into two species. *Mycological Research, 99*, 1182-1186.

Pleiss, J., Fischer, M., Peiker, M., Thiele, C., & Schmid, R. D. (2000). Lipase engineering database understanding and exploiting sequence-structure-function relationships. *Journal of Molecular Catalysis B: Enzymatic, 10*, 491-508.

Pongsunk, S., Andrianopoulos, A., & C., C. S. (2005). Conditional lethal disruption of TATA-binding protein gene in *Penicillium marneffei Fungal Genetics and Biology, 42*, 893-903.

Pradet-balade, B., Boulmé, F., Beug, H., Müllner, E. W., & Garcia-Sanz, J. A. (2001). Translation control: bridging the gap between genomics and proteomics. *TRENDS in Biochemical Sciences. , 26*(4), 225-229.

Prak, E. T. L., & Kazazian, H. H. (2000). Mobile elements and the human genome. *Nature Reviews 1*, 134-144.

Preiss, T., & Hentze, M. W. (1999). From factor to mechanisms: translation and translational control in eukaryotes. *Current Opinion in Genetics and Development, 9*, 515-521.

Pronk, M. E. J., Verger, P., & Walker, R. (2008). *Phospholipase A1 from Fusarium venenatum expressed in Aspergillus oryzae*: World Health Organization - Food additives series.

Proud, C. G. (1994). Peptide-chain elongation in eukaryotes. *Molecular Biology Reports, 19*(3), 161-170.

Quin, L., Overmars, H., Helder, J., Popeijus, H., van der Voort, J. R., Groenink, W., van Koert,

P., Schots, A., Bakker, J., & Smant, G. (2000). An efficient cDNA-AFLP-based strategy for the identification of putative pathogenicity factors from the potato cyst nematode *Globodera rostochiensis*. *Molecular Plant-Microbe Interactions, 13*(8), 830-836.

Reeslev, M., Strom, T., Jensen, B., & Olsen, J. (1997). The ability of the yeast form of *Aureobasidium pullullans* to elaborate exopolysaccharide in chemostat culture at various pH values. *Mycological Research 101*, 650-652.

Rhoads, R. E., Dinkova, T. D., & Korneeva, N. L. (2006). Mechanism and regulation of translation in *C. elegans*. In Wormbook (Ed.): The *C. elegans* Research Community, Wormbook.

Riley, M. (1993). Functions of the gene products of *Escericia coli*. *FEMS Microbiology Reviews, 57*, 862-952.

Ririe, K. M., Rasmussen, R. P., & Wittwer, C. T. (1997). Product differentiation by analysis of DNA melting curves during the polymerase chain reaction *Analytical Biochemistry, 245*(2), 154-160.

Robinson-Jeffrey, R. C., & Davidson, R. W. (1968). *Canadian Journal of Botany., 46*, 1523.

Robson, G. D., Huang, J., Wortman, J., & Archer, D. B. (2005). A preliminary analysis of the process of protein secretion and the diversity of putative secreted hydrolases encodedin *Aspergillus fumigatus*: insights from the genome. *Medical Mycology Supplement, 43*, S41-S47.

Rosse , L., Norskov, P., Holmstrom, K., & Rasmussen, O. F. (1992). Inhibition of PCR components of food samples, microbial diagnostic assays and DNA-extraction solutions. *International Journal of Food Microbiology, 17*, 37-45.

Ryan, K. J., & (editors), R. C. G. (2004). *Sherris Medical Microbiology* (4th ed.): McGraw Hill.

Samuels, G. J. (1993). *The case for distinguishing Ceratocystis and Ophiostoma.*: APS Press, St. Paul, Minnesota.

Sánchez-Martinez, C., & Pérez-Martin, J. (2001). Dimorphism in fungal pathogens: *Candida albicans* and *Ustilago maydis* - similar inputs, different outputs. *Current Opinion in Microbial Biotechnology, 4*, 214-221.

Schauwecker, F., Wanner, G., & Kahmann, R. (1995). Filament-specific expression of a cellulase gene in the dimorphic fungus *Ustilago madis*. *Biol. Chem. Hoppe Seyler, 376*, 617-625.

Schirp, A. (2002). *Effect of sapstaining fungi on structural wood integrity of radiate pine. PhD Thesis.* The University of Waikato, New Zealand, Hamilton.

Schirp, A., Farrell, R. L., & Kreber, B. (2003a). Effects of New Zealand Sapstaining Fungi on Structural Integrity of Unseasoned Radiata Pine. *Holz als Rohund Werkstoff, 61*, 369-376.

Schirp, A., Farrell, R. L., Kreber, B., & Singh, A. P. (2003b). Advances in understanding the ability of sapstaining fungi to produce cell-wall degrading enzymes. *Wood and Fiber Science, 35*, 434-444.

Schmittgen, T. D., Zakrajsek, B. A., Mills, A. G., Gorn, V., Singer, M. J., & Reed, M. W. (2000). Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: Comparison of endpoint and real-time methods. *Analytical Biochemistry, 285*, 194-204.

Schroeder, S., Kim, S. H., Cheung, W. T., Sterflinger, K., & Breuil, C. (2001). Phylogenetic relationshio of *Ophiostoma piliferum* to other sapstain fungi based on the nuclear rRNA gene. *FEMS Microbiological Letters, 195*(2), 163-167.

Scung, J. (2003). *Using TESS to predict transcription factor binding sites in DNA sequence*: J. Wiley and Sons.

Seifert, K. A., & Okada, G. (1993). *Graphium anamorphs of Ophiostoma species and similar anamorphs of other ascomycetes. In Ceratocystis and Ophiostoma, taxonomy, ecology and pathogenicity*. St. Paul, Minnesota: The American Phytopathological Society.

Selbmann, L., Stingele, F., & Petruccioli, M. (2003). Exopolysaccharide production by filamentous fungi: the example of *Botryosphaeria rhodina*. *Antonie van Leeuwenhoek, 84*, 135-145.

Selker, E. U., Nikolaos, A. T., Cross, S. H., Margolin, B. S., Murphy, J. G., Bird, A. P., & Freitag, M. (2003). The methylated component of the *Neurospora crassa* genome. *Nature, 422*, 893-897.

Seo, Y. S., Kim, E. Y., Mang, H. G., & Kim, W. T. (2008). Heterologous expression, and biochemical and cellular characterization of *CaPLA1* encoding a hot pepper phospholipase A1 homolog. *The Plant Journal, 53*(6), 895-908.

Sharp, P. M., & Li, W. H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *Journal of Molecular Evolution., 24*, 28-38.

Sharp, P. M., & Li, W. H. (1987). The codon adaptation index – a measure of directional synonymous codond usage bias and its potential applications *Nucleic Acids Research, 15*, 1281-1295.

Sims, A. H., Dunn-Coleman, N. S., Robson, G. D., & Oliver, S. G. (2004). Glutamic protease distribution is limited to filamentous fungi. *FEMS Microbiology Letters 239*(1), 95-101.

Sinclair, G., & Choy, F. Y. M. (2002). Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*. *Protein Expression and Purification, 26*, 96-105.

Singh, A., Kumar, P. K. R., & Schugerl, K. (1992). Bioconversion of cellulosic material to ethanol by filamentous fungi. *Advances in Biochemical Engineering/Biotechnology 45*, 29-55.

Skory, C. D., Freer, S. N., & Bothast, R. J. (1997). Screening for ethanol-producing filamentous fungi. *Biotechnology Letters 19*, 203-206.

Small, J. L., & McAlister-Henn, L. (1998). Identification of a cytosolically directed NADH dehydrogenase in mitochondria of *Saccharomyces cerevisiae*. *Journal of Bacteriology, 180*, 4051-4055.

Solheim, H., & Krokene, P. (1998). Growth and virulence of mountain pine beetle associated

blue-stain fungi, *Ophiostoma clavigerum* and *Ophiostoma montium*. *Canadian Journal of Botany, 76*(4), 561-566.

Solheim, H., & Krokene, P. (1998). Growth and virulence of *Ceratocystis rufipenni* and three blue-stain fungi isolated from the Douglas-fir beetle. *Canadian Journal of Botany, 76*, 1763-1769.

Sonenberg, N., Hershey, J. W. B., & Mathews, M. B. e. (2000). *Translational control of gene expression*: Cold Spring Harbor Laboratory Press.

Soundararajan, S., Jedd, G., Li, X., Ramos-Pamplona, M., Chua, N. H., & Naqvi, N. I. (2004). Woronin body function in *Magnaporthe grisea* is essential for efficient pathogenesis and for survival during nitrogen starvation stress. *Plant Cell, 16*(6), 1564-1574.

Spatafora, J. W., & Blackwell, M. (1994). The polyphyletic origins of opiostomatoid fungi. *Mycological Research, 98*, 1-9.

Suárez, B., Vizcaíno, J. A., Llobell, A., & Monte, E. (2007). Characterization of genes encoding novel peptidases in the biocontrol fungus *Trichoderma harzianum* CECT 2413 using the TrichEST functional genomics approach. *Current Genetics, 51*(5), 331-342.

Swift, G. H., Peyton, M. J., & MacDonald, R. J. (2000). Assessment of RNA quality by semi-quantitative RT-PCR of multiple regions of a long ubiquitous mRNA. *Biotechniques, 28*, 524-531.

Swire, J., Judson, O. P., & Burt, A. (2005). Mitochondrial genetic codes evolves to match amino acid requirements of proteins. *Journal of Molecular Evolution, 60*, 128-139.

Tadesse, Y., Bernier, L., Hintz, W. E., & Horgen, P. A. (2003). Real time RT-PCR quantification and Northern analysis of *Cerato ulmin* (CU) gene transcription in different strains of the phytopathogens *Ophiostoma ulm*i and *O. novo ulmi*. *Molecular Genetics and Genomics MGG, 269*, 789-796.

Talbot, N. J. (2003). On the trail of a cereal killer. Exploring the biology of *Magnaporthe grisea*. *Annual Review of Microbiology, 57*, 177-202.

Tanguay, P., Bozza, S., & Breuil, C. (2006). Assessing RNAi frequency and efficiency in *Ophiostoma floccosum* and *O. piceae*. *Fungal Genetics and Biology, 43*(12), 804-812.

Tanguay, P., & Breuil, C. (2003). Transforming the sapstaining fungus *Ophiostoma piceae* with *Agrobacterium tumefaciens*. *Canadian Journal of Microbiology 49*(4), 301-304.

Tanguay, P., Tangen, K., & Breuil, C. (2007). Identifying pigmentation related genes in *Ophiostoma piceae*. *Phytopathology, 97*, 1040-1048.

Tatsuta, T., Augustin, S., Nolden, M., Friedrichs, B., & Langer, T. (2007). *m*-AA protease-driven membrane dislocation allows intramembrane cleavage by rhomboid in mitochondria. *The EMBO Journal, 26*, 325-335.

Tenney, K., Hunt, I., Sweigard, J., Pounder, J. I., Mclain, C., Bowman, E. J., & Bowman, B. J. (2000). *Hex-1*, a gene unique to filamentous fungi, encodes the major protein of the Woronin body and functions as a plug for septal pores. *Fungal Genetics 3*, 205-217.

Thacker, D. G., & Good, H. M. (1952). The composition of air in trunks of sugar maple in

relation to decay. *Canadian Journal of Botany 30*, 475-485.

Thomson, C. A., Delaquis, P. J., & Mazza, G. (1999). Detection and measurement of microbial lipase activity: a review. *Critical Reviews in Food Science and Nutrition, 39*(2), 165-187.

Thwaites, J. M. (2003). *Black and White: Investigation on the ecology, biology and biological control of sapstain fungi in New Zealand.* Unpublished Doctor of Philosophy, The University of Waikato, Hamilton.

Thwaites, J. M., Farrell, R. L., Duncan, S. M., Reay, S., Blanchette, R. A., Hadar, E., Hadar, Y., Harrington, T. C., & McNew, D. (2005). Survey of potential sapstain fungi on *Pinus radiata* in New Zealand. *New Zealand Journal of Botany, 43*, 653-663.

Umzurike, G. M. (1969). Cellulolytic activitis of *Botryodiplodia theobromae* pat. *Annals of Botany, 33*, 451-462.

Upadhyay, H. P. (1981). *A monograph of Ceratocystis and Ceratocystiopsis*. Athens, Georgia. : University of Georgia Press.

Upadhyay, H. P. (1993). *Classification of the Ophiostomatoid fungi*: APS Press, St. Paul Minnesota.

Upadhyay, H. P., & Kendrick, W. B. (1975). Prodromus for the revision of *Ceratocystis* (Microascales, Ascomycetes) and its condidial states. *Mycologia, 67*, 798-805.

Uzunovic, A., Yang, D. Q., Gagne, P., Breuil, C., Bernier, L., Byrne, A., Gignac, M., & Kim, S. H. (1999). Fungi that cause sapstain in Canadian softwoods. *Canadian Journal of Microbiology., 45*, 914-922.

Van Pouderoyen, G., Eggert, T., Jaeger, K. E., & Dijkstra, B. W. (2001). The crystal structure of *Bacillus subtilis* lipase: a minimal alpha/beta hydrolase fold enzyme. *Journal of Molecular Biology, 309*, 103-119.

Van Wyk, P. W., & Wingfield, M. J. (1992). Ascospore development in *Ophiostma piceae*. *Canadian Journal of Botany, 70*, 2170-2176.

Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Hieter, P., Vogelstein, B., & Kinzler, K. W. (1997). Charaterization of the yeast transcriptome. *Cell, 88*, 243-251.

Videira, A. (1998). Complex I from the fungus *Nerospora crassa*. *Biochimica et Biophysica Acta, 1364*, 89-100.

Viret, O., Toti, L., Chapela, I. H., & Petrini, O. (1994). The role of the extracellular sheath in recognition and attachment of conidia of *Discula umbrinella* (Berk. & Br.) Morelet to the host surface. *New Phytologist 127*, 123-131.

Wada, K. S., Aota, R., Tsuchiya, F., Ishibashi, T., Gojobori, T., & Ikemure, T. (1990). Codon usage tabulated from GenBank genetic sequence data *Nucleic Acids Research, 18*, 2367-2411.

Waites, M. J., Rockey, N. L., & Higton, J. S. (2001). *Industrial Microbiology: an Introduction* Blackwell Science, Ltd.

Wang, C. Q., Eufemi, M., Turano, M., & Giartosio, A. (1996). Influence of the carbohydrate moiety on the stbility of glycoproteins. *Biochemisty, 35*, 7299-7307.

Wendler, P. A., Brush, T. S., Iverson, S., Krisa, K., Zimmerman, W., & Farrell, R. L. (1992). Biological control of pitch problems in paper mills. *Kemia kemi, 19*, 262-264.

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J., & Bahler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature Letters*.

Wingfield, M. J., Seifert, K. A., & Webber, J. J. (1993). *Ceratocystis and Ophiostoma: Taxonomy ecology and pathogenicity*. St. Paul, Minnesota: APS Press.

Winkler, U. K., & Stuckmann, M. (1979). Glycogen, hyaluronate, and some other polysaccharides greatly enhance the formation of exoplipase by *Serratia marcescens*. *Journal of Bacteriology, 138*, 663-670.

Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., & Lockhart, D. J. (1997). Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature, Biotechnology, 15*, 1359-1367.

Wolner, B. S., & Gralla, J. D. (2000). Roles for non-TATA core promoter sequences in transcription factor binding. *Molecular and Cellular Biology 20*(10), 3608-3615.

Wu, C., Te'o, V. S. J., Farrell, R. L., Bergquist, P. L., & Nevalainen, K. M. H. (2006). Improvement of the secretion of extracellular proteins and isolation and characterization of the amylase I (*amy1*) gene from *Ophiostoma floccosum*. *Gene, 384*, 96-103.

Wu, C., Xu, Q., Liu, F., & Nevalainen, K. M. H. (2007). Activity-based identification of secreted serine proteases of the filamentous fungus, *Ophiostoma*. *Biotechnology Letters, 29*, 937-943.

Yoda, K., & Noda, Y. (2001). Vesicular transport and the golgi apparatus in yeast. *Journal of Bioscience and Bioengineering, 91*(1), 1-11.

Yu, J., Hu, S., Wang, J., Wong, G. K., Lis, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, A., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., ZXu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L., & Yang, H. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science, 296*(5565), 79-92.

Zabel, R. A., & Morrell, J. J. (1992). *Wood Microbiology: Decay and its prevention.* Sandiego, California: Academic Press Inc.

Zhang, J., & Byrne, C. D. (1999). Differential priming of RNA templates during cENA synthesis markedly affects both accuracy and reproducibility of quantitative competitive reverse-transcriptase PCR. *Analytical Biochemistry, 285*, 194-204.

Zimmerman, W., Blanchette, R. A., Burnes, T. A., & Farrell, R. L. (1995). Melanin and perithecial development in *Ophiostoma piliferum*. *Mycologia, 87*(6), 857-863.

Zipfel, R. D., de Beer, Z. W., Jacobs, K., Wingfield, B. D., & Wingfield, M. J. (2006). Multi-gene phylogenies define *Ceratocystiopsisi* and *Grosmannia* distinct from *Ophiostoma*. *Studies in Mycology, 55*, 75-97.

Appendix 1: Annotation of Significantly Aligned Unique Putative Transcripts from *Ophiostoma floccosum* EST dataset OF48

FULL-LENGTH

ID	Predicted Function	*E*-value	Start
OF0002	(Q9CR62) Mitochondrial 2-oxoglutarate/malate carrier protein (OGCP) (Solute carrier family 25 member 11)	2.00E-29	62_64
OF0007	(P22151) Glucose-repressible gene protein	3.00E-20	133_135
OF0017	(Q01520) Elongation factor 1-alpha (EF-1-alpha)	0.00E+00	134_136
OF0018	(Q01554) ATP synthase protein 9, mitochondrial (EC 3.6.3.14) (Lipid-binding protein)	1.00E-17	140_142
OF0026	(O00086) Probable inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205) (IMP dehydrogenase) (IMPDH) (IMPD)	9.00E-36	29_31
OF0032	(P15959) NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3) (NADH dehydrogenase subunit 6)	4.00E-35	149_151
OF0034	(O94083) Eukaryotic translation initiation factor 5A (eIF-5A) (eIF-4D)	2.00E-51	306_308
OF0045	(P03945) Cytochrome c oxidase subunit 1 (EC 1.9.3.1) (Cytochrome c oxidase polypeptide I)	2.00E-99	130_132
OF0046	(P36149) Transport protein particle 22 kDa subunit (TRAPP 22 kDa subunit)	2.00E-50	243_245
OF0073	(Q9RA05) Carveol dehydrogenase (EC 1.1.1.275) ((+)-trans-carveol dehydrogenase) (CDH)	0.00E+00	558_560
OF0077	(Q8X132) Histone H2A	2.00E-48	28_30
OF0080	(P15994) ATP synthase a chain (EC 3.6.3.14) (ATPase protein 6)	4.00E-63	227_229
OF0085	(P54745) Heat-responsive suppressor hrsA (Putative PTS system EIIABC component)	0.00E+00	322_324
OF0097	(P76577) Penicillin-binding protein 1C (PBP-1c) (PBP1c)	E-1OF_07	60_62
OF0102	(P18775) Anaerobic dimethyl sulfoxide reductase chain A precursor (EC 1.8.99.-) (DMSO reductase)	E-127	31_33
OF0108	(P20681) Cytochrome c oxidase subunit 1 (EC 1.9.3.1) (Cytochrome c oxidase polypeptide I)	3.00E-15	232_234
OF0111	(P59302) Acetylglutamate kinase (EC 2.7.2.8) (NAG kinase) (AGK) (N-acetyl-L-glutamate 5-phosphotransferase)	2.00E-50	36_38
OF0112	(P36938) Phosphoglucomutase (EC 5.4.2.2) (Glucose phosphomutase) (PGM)	E-168	264_266
OF0113	(P0A6A6) 3-isopropylmalate dehydratase large subunit (EC 4.2.1.33) (Isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI)	2.00E-72	49_51
OF0118	(P37773) UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso- diaminopimelate ligase (EC 6.3.2.-) (Murein peptide ligase)	E-112	58_60
OF0124	(Q9X5X4) HTH-type transcriptional regulator hmrR (Copper efflux regulator) (Copper export regulator)	1.00E-18	9_11
OF0125	(Q3YZN3) Penicillin-insensitive murein endopeptidase precursor (EC 3.4.24.-) (D-alanyl-D-alanine-endopeptidase) (DD-endopeptidase)	E-117	135_137
OF0127	(P77690) UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase (EC 2.6.1.-) (UDP-(beta-L-threo-pentapyranosyl-4''-ulose diphosphate) aminotransferase) (UDP-Ara4O aminotransferase) (Polymyxin resistance protein pmrH)	6.00E-56	17_19
OF0129	(P35627) Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase) (Rotamase) (Cyclophilin) (Cyclosporin A-binding protein)	1.00E-33	72_74
OF0130	(P0AFG3) 2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2) (Alpha-ketoglutarate dehydrogenase)	E-118	64_66
OF0134	(P38054) Cation efflux system protein cusA	E-150	39_41
OF0135	(Q88DX4) Glutamyl-Q tRNA(Asp) synthetase (EC 6.1.1.-) (Glu-Q-RSs)	2.00E-08	385_387
OF0137	(P37002) Protein crcB	2.00E-41	85_87

ID	Predicted Function	E-value	Start
OF0141	(P53720) tRNA-dihydrouridine synthase 2 (EC 1.-.-.-) (SMM1 protein)	9.00E-33	26_28
OF0144	(Q99LE6) ATP-binding cassette sub-family F member 2	1.00E-20	151_153
OF0148	(P0A9F6) Glycine cleavage system transcriptional activator (Gcv operon activator)	4.00E-14	112_114
OF0152	(P0AEH4) Protein elaA	1.00E-39	170_172
OF0154	(P21647) Outer membrane usher protein mrkC precursor	4.00E-21	113_115
OF0156	(O67610) 3-oxoacyl	7.00E-31	181_183
OF0157	(P52OF_073) Glycolate oxidase subunit glcE	9.00E-74	13_15
OF0158	(P25666) Protein htrL	5.00E-23	150_152
OF0162	(P0AC23) Probable formate transporter 1 (Formate channel 1)	2.00E-63	234_236
OF0164	(Q83LU7) Probable ATP-dependent helicase dinG (EC 3.6.1.-) (DNA-damage-inducible protein G)	4.00E-71	55_57
OF0167	(P11989) Cryptic beta-glucoside bgl operon antiterminator	2.00E-37	136_138
OF0168	(P03769) Protein ninF	1.00E-27	87_89
OF0171	(P00892) Acetolactate synthase isozyme II large subunit (EC 2.2.1.6) (AHAS-II) (Acetohydroxy-acid synthase II large subunit) (ALS-II)	4.00E-86	62_64
OF0172	(P30860) Arginine-binding periplasmic protein 2 precursor	E-127	364_366
OF0176	(P20449) ATP-dependent RNA helicase DBP5 (EC 3.6.1.-) (DEAD box protein 5) (Helicase CA5/6) (Ribonucleic acid trafficking protein 8)	0.00E+00	142_144
OF0186	(P19624) 4-hydroxythreonine-4-phosphate dehydrogenase (EC 1.1.1.262) (4-(phosphohydroxy)-L-threonine dehydrogenase)	3.00E-95	104_106
OF0189	(P59405) Anaerobic nitric oxide reductase flavorubredoxin (FlRd) (FlavoRb)	1.00E-65	395_397
OF0190	(P18776) Anaerobic dimethyl sulfoxide reductase chain B (DMSO reductase iron-sulfur subunit)	1.00E-89	70_72
OF0193	(P62594) Beta-lactamase TEM precursor (EC 3.5.2.6) (Penicillinase)	E-101	49_51
OF0195	(P24174) Mannose-1-phosphate guanylyltransferase	E-117	418_420
OF0197	(P18186) Ornithine carbamoyltransferase (EC 2.1.3.3) (OTCase)	9.00E-25	21_23
OF0200	(P53514) Outer membrane usher protein pmfC precursor	4.00E-31	177_179
OF0202	(P75925) Cytochrome b561 homolog 2	2.00E-36	757_759
OF0204	(Q98KT9) Probable malate--CoA ligase beta chain (EC 6.2.1.9) (Malyl-CoA synthetase) (Malate thiokinase) (MTK-beta)	4.00E-89	159_161
OF0207	(P09147) UDP-glucose 4-epimerase (EC 5.1.3.2) (Galactowaldenase) (UDP-galactose 4-epimerase)	E-109	295_297
OF0209	(P32721) D-allose transport ATP-binding protein alsA (EC 3.6.3.17)	5.00E-71	31_33
OF0210	(P0AFB7) Nitrogen regulation protein NR(II) (EC 2.7.3.-)	3.00E-70	29_31
OF0211	(Q46629) Amylovoran export outer membrane protein amsH precursor	4.00E-70	1_3
OF0212	(P12985) ATP synthase alpha chain (EC 3.6.3.14)	E-103	136_138
OF0214	(P22523) Chromosome partition protein mukB (Structural maintenance of chromosome related protein)	0.00E+00	152_154
OF0221	(P18200) Phosphatidylglycerophosphatase A (EC 3.1.3.27)	1.00E-93	111_113
OF_07_A1-GX	(Q02854) NADH-ubiquinone oxidoreductase 21 kDa subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-21KD) (CI-21KD)	7.00E-60	250_252
OF_07_A9-GX	(P02723) ADP,ATP carrier protein (ADP/ATP translocase) (Adenine nucleotide translocator) (ANT)	E-115	119_121
OF_07_B6-EX	(Q6FJ04) mRNA turnover protein 4 homolog	2.00E-06	65_67
OF_07_B8-FX	(P48024) Eukaryotic translation initiation factor 1 (eIF1) (Protein translation factor SUI1 homolog)	1.00E-27	120_122
OF_07_C5-EX	(Q9Y237) Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 (EC 5.2.1.8) (Rotamase Pin4) (PPIase Pin4) (Parvulin 14) (Par14) (Peptidyl-prolyl cis/trans isomerase EPVH) (hPar14)	9.00E-22	4_6
OF_07_D10-FX	(Q9LJD7) Constitutive photomorphogenesis protein 10	6.00E-21	243_245
OF_07_D12-BX	(Q18885) Transcription factor BTF3 homolog (Inhibitor of cell death 1)	6.00E-34	3_5
OF_07_D9-AX	(Q9HAU6) Apoptosis inhibitor FKSG2	1.00E-21	116_118
OF_07_E1-BX	(P37210) Histone H2B	4.00E-55	3_5

ID	Predicted Function	*E*-value	Start
OF_07_E1-BX	(P37210) Histone H2B	4.00E-55	3_5
OF_07_E3-CX	(Q5RBI3) 26S proteasome non-ATPase regulatory subunit 12 (26S proteasome regulatory subunit p55)	2.00E-48	121_123
OF_07_E5-DX	(Q96VN3) Nuclear transport factor 2 (NTF-2)	9.00E-39	85_87
OF_07_F11-A-SP6	(P40303) Proteasome component PRE6 (EC 3.4.25.1) (Macropain subunit PRE6) (Proteinase YSCE subunit PRE6) (Multicatalytic endopeptidase complex subunit PRE6)	4.00E-89	117_119
OF_07_F1-H-SP6	(P32378) Para-hydroxybenzoate--polyprenyltransferase, mitochondrial precursor (EC 2.5.1.-) (PHB:polyprenyltransferase)	7.00E-43	80_82
OF_07_G10-A-SP6	(Q01302) Clock-controlled protein 6	1.00E-10	132_134
OF_07_G11-B-SP6	(P32495) H/ACA ribonucleoprotein complex subunit 2 (H/ACA snoRNP protein NHP2) (High mobility group-like nuclear protein 2)	6.00E-11	164_166
OF_07_G11-H-SP6	(P37202) Mitotic control protein dis3	1.00E-49	448_450
OF_07_G5-A-SP6	(Q9VJ33) NEDD8 precursor (Ubiquitin-like protein Nedd8) (Neddylin)	2.00E-20	19_21
OF_07_G6-E-SP6	(Q12702) Protein phosphatase PP2A regulatory subunit B (PR55) (Protein phosphatase 2A 55 kDa regulatory subunit)	8.00E-48	2OF_07_209
OF_07_G8-D-SP6	(Q9Z223) Molybdenum cofactor synthesis protein 2 large subunit (Molybdopterin synthase large subunit) (MPT synthase large subunit) (MOCS2B)	6.00E-19	222_224
OF_07_H11-E-SP6	(Q8X097) Probable ATP-citrate synthase subunit 1 (EC 2.3.3.8) (ATP-citrate (pro-S-)-lyase 1) (Citrate cleavage enzyme subunit 1)	E-120	187_189
OF_07_H1-H-SP6	(P20654) Serine/threonine-protein phosphatase PP1 (EC 3.1.3.16)	E-145	9_11
OF_07_H3-D-SP6	(P53444) Fructose-bisphosphate aldolase (EC 4.1.2.13)	2.00E-79	30_32
OF_07_I10-E-SP6	(Q5QXH8) 3-ketoacyl-CoA thiolase (EC 2.3.1.16) (Fatty oxidation complex beta subunit) (Beta-ketothiolase) (Acetyl-CoA acyltransferase)	3.00E-26	61_63
OF_07_I11-H-SP6	(P04914) Histone H4	1.00E-39	135_137
OF_07_I2-C-SP6	(P42058) Minor allergen Alt a 7 (Alt a VII)	8.00E-31	112_114
OF_07_I5-A-SP6	(P32637) Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH)	E-108	129_131
OF_07_I5-C-SP6	(P04914) Histone H4	3.00E-45	160_162
OF_07_I8-E-SP6	(Q9JK72) Copper chaperone for superoxide dismutase (Superoxide dismutase copper chaperone)	1.00E-28	132_134
OF_11D-y	(Q08268) Probable transporter MCH4	1.00E-11	210_212
OF_12B-x	(Q01610) Putative transcriptional regulator	8.00E-07	290_292
OF_12G-x	(Q83QZ5) Deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP deaminase)	2.00E-74	56_58
OF_14D-x	(P0ABB3) ATP synthase alpha chain (EC 3.6.3.14)	3.00E-56	104_106
OF_4F-y	(P15578) NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3) (NADH dehydrogenase subunit 2)	6.00E-49	73_75
OF_05_2H-y	(P15578) NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3) (NADH dehydrogenase subunit 2)	9.00E-56	121_123
OF_05_3A-y	(P15578) NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3) (NADH dehydrogenase subunit 2)	4.00E-61	94_96

ID	Predicted Function	E-value	Start
OF_05_5F-x	(P53595) Malate--CoA ligase alpha chain (EC 6.2.1.9) (Malyl-CoA synthetase) (Malate thiokinase) (MTK-alpha)	1.00E-43	386_388
OF_1-H-y	(Q27546) Inosine-uridine preferring nucleoside hydrolase (EC 3.2.2.1) (IU-nucleoside hydrolase) (IU-NH) (Purine nucleosidase) (Nonspecific nucleoside hydrolase)	1.00E-07	80_82
OF_14-H-x	(P28861) Ferredoxin--NADP reductase (EC 1.18.1.2) (FNR) (Flavodoxin reductase) (FLXR) (FLDR) (Methyl viologen resistance protein A) (DA1)	2.00E-55	14_16

SHORT FULL-LENGTH

ID	Predicted Function	*E*-value	Start
OF0023	(Q9P6C8) Alcohol dehydrogenase 1 (EC 1.1.1.1) (Alcohol dehydrogenase I)	1.00E-98	29_31
OF0072	(O42993) FK506-binding protein (FKBP) (Peptidyl-prolyl cis-trans isomerase) (PPIase) (EC 5.2.1.8)	4.00E-15	85_87
OF0089	(P0AAG3) Glutamate/aspartate transport ATP-binding protein gltL	E-109	184_186
OF0099	(P0AFG3) 2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2) (Alpha-ketoglutarate dehydrogenase)	E-103	44_46
OF0121	(P36771) Probable HTH-type transcriptional regulator lrhA	E-134	103_105
OF0149	(P24488) Transcriptional activator hap2	0.0	135_137
OF0161	(P81549) Probable tonB-dependent receptor bfrD precursor (Virulence-associated outer membrane protein Vir-90)	2.00E-25	59_61
OF0174	(P40151) DNA-dependent ATPase MGS1 (Maintenance of genome stability protein 1)	1.00E-33	13_15
OF0175	(Q8Q0M5) Tetrahydromethanopterin:alpha-L-glutamate ligase (EC 6.3.2.-) (H(4)MPT:alpha-L-glutamate ligase)	3.00E-18	129_131
OF0179	(Q8PY84) CoB--CoM heterodisulfide reductase 1 iron-sulfur subunit C (EC 1.8.98.1)	5.00E-06	125_127
OF0182	(P0A9H9) Chemotaxis protein cheZ	1.00E-88	80_82
OF0196	(P80668) Phenylacetaldehyde dehydrogenase (EC 1.2.1.39) (PAD)	9.00E-78	46_48
OF_07_B5-HX	(Q9C291) Double-strand break repair protein mus-23 (Recombinational repair protein mus-23)	E-136	129_131
OF_07_E6-FX	(Q9D099) Alkaline phytoceramidase (EC 3.5.1.-) (aPHC) (Alkaline ceramidase)	2.00E-24	14_16
OF_07_H2-C-SP6	(P01097) ATPase inhibitor, mitochondrial precursor	8.00E-06	112_114

PARTIAL

ID	Predicted Function	*E*-value	Start
OF0024	(P71377) Cell division protein ftsH homolog 1 (EC 3.4.24.-)	4.00E-69	NO
OF0035	(P05510) NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3) (NADH dehydrogenase subunit 5)	5.00E-29	9_11
OF0065	(Q8NK50) L-xylulose reductase (EC 1.1.1.10) (XR)	4.00E-41	1_3
OF0067	(O74173) Elongation factor 1-beta (EF-1-beta)	3.00E-31	334_336
OF0083	(P09201) Fructose-1,6-bisphosphatase (EC 3.1.3.11) (D-fructose-1,6-bisphosphate 1-phosphohydrolase) (FBPase)	1.00E-57	206_208
OF0084	(P30015) Probable ATP-dependent helicase lhr (EC 3.6.1.-) (Large helicase-related protein)	E-113	NO
OF0086	(P03960) Potassium-transporting ATPase B chain (EC 3.6.3.12) (Potassium-translocating ATPase B chain) (ATP phosphohydrolase	9.00E-79	24_26
OF0094	(P00864) Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPCase) (PEPC)	1.00E-50	71_73
OF0095	(Q8NRV6) Peptidyl-tRNA hydrolase 2 (EC 3.1.1.29) (PTH 2)	6.00E-17	NO
ID	Predicted Function	E-value	Start
OF0105	(P42086) Xanthine permease	5.00E-17	58_60

ID	Predicted Function	*E*-value	Start
OF0109	(Q9UTH3) Probable serine/threonine-protein kinase C1805.01c (EC 2.7.1.37)	5.00E-67	60_62
OF0115	(P31134) Putrescine transport ATP-binding protein potG	5.00E-72	227_229
OF0120	(P0ACD9) Hydrogenase-1 large chain (EC 1.12.99.6) (NiFe hydrogenase) (Membrane-bound hydrogenase 1 large subunit) (HYD1)	4.00E-80	130_132
OF0139	(Q08484) GTPase-activating protein GYP1 (GAP for YPT1)	6.00E-22	87_89
OF0155	(P0AFA4) Nitrate/nitrite sensor protein narX (EC 2.7.3.-)	2.00E-95	121_123
OF0177	(P24137) Oligopeptide transport ATP-binding protein oppF	2.00E-47	NO
OF0180	(P62594) Beta-lactamase TEM precursor (EC 3.5.2.6) (Penicillinase)	5.00E-83	90_92
OF0183	(P0AEE7) D-galactose-binding periplasmic protein precursor (GBP) (D-galactose/ D-glucose-binding protein) (GGBP)	0.00000000 OF_07	NO
OF0191	(P23930) Apolipoprotein N-acyltransferase (EC 2.3.1.-) (ALP N-acyltransferase) (Copper homeostasis protein cutE)	5.00E-58	129_131
OF0205	(Q45978) Holdfast attachment protein C (Protein hfaC)	3.00E-21	115_117
OF0206	(P54745) Heat-responsive suppressor hrsA (Putative PTS system EIIABC component)	6.00E-37	135_137
OF0218	(Q9UKM7) Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase (EC 3.2.1.113) (ER alpha-1,2-mannosidase) (Mannosidase alpha class 1B member 1) (Man9GlcNAc2-specific-processing alpha-mannosidase)	8.00E-30	90_92
OF0224	(P0AFG6) Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex (EC 2.3.1.61) (E2) (Dihydrolipoamide succinyltransferase component of 2-oxoglutarate dehydrogenase complex)	E-120	60_62
OF_07_A10-FX	(Q4I8Q4) Acyl-protein thioesterase 1 (EC 3.1.2.-)	4.00E-10	72_74
OF_07_A12-EX	(Q99144) Peroxisomal targeting signal receptor (Peroxisomal protein PAY32) (Peroxin-5) (PTS1 receptor)	4.00E-12	54_56
OF_07_A6-EX	(P37211) ATP synthase alpha chain, mitochondrial precursor (EC 3.6.3.14)	2.00E-47	211_213
OF_07_A7-DX	(P41816) NADPH dehydrogenase 3 (EC 1.6.99.1) (Old yellow enzyme 3)	1.00E-22	158_160
OF_07_A9-BX	(Q9CWD8) Nucleotide-binding protein-like	6.00E-31	142_144
OF_07_B11-DX	(P53621) Coatomer subunit alpha (Alpha-coat protein) (Alpha-COP) (HEPCOP) (HEP-COP)	9.00E-32	65_67
OF_07_B12-GX	(Q04894) NADP-dependent alcohol dehydrogenase 6 (EC 1.1.1.2) (NADP-dependent alcohol dehydrogenase VI) (ScADHVI)	1.00E-16	185_187
OF_07_B6-GX	(P40292) Heat shock protein 90 (Heat shock protein hsp1) (65 kDa IgE-binding protein) (Allergen Asp f 12)	4.00E-55	244_246
OF_07_C12-FX	(Q88DX4) Glutamyl-Q tRNA(Asp) synthetase (EC 6.1.1.-) (Glu-Q-RSs)	2.00E-09	NO
OF_07_C3-BX	(P17423) Homoserine kinase (EC 2.7.1.39) (HSK) (HK)	8.00E-26	201_203
OF_07_C3-DX	(P40495) Homoisocitrate dehydrogenase, mitochondrial precursor (EC 1.1.1.87)	3.00E-81	NO
OF_07_C3-FX	(Q9ZAE8) dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)	2.00E-12	NO
OF_07_C4-GX	(Q96X46) Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-glycerate hydro-lyase) (Allergen Pen c 22)	E-122	26_28
OF_07_D4-FX	(O70152) Dolichol-phosphate mannosyltransferase (EC 2.4.1.83) (Dolichol-phosphate mannose synthase) (Dolichyl-phosphate beta-D-mannosyltransferase) (Mannose-P-dolichol synthase) (MPD synthase) (DPM synthase)	3.00E-28	226_228
OF_07_D9-HX	(Q01369) Guanine nucleotide-binding protein beta subunit-like protein (Cross-pathway control WD-repeat protein cpc-2)	1.00E-71	149_151
OF_07_E1-CX	(P24590) Phosphoglycerate kinase (EC 2.7.2.3)	2.00E-84	164_166
OF_07_E2-BX	(POF_07337) Beta-glucosidase precursor (EC 3.2.1.21) (Gentiobiase) (Cellobiase) (Beta-D-glucoside glucohydrolase)	1.00E-24	3_5
OF_07_E4-AX	(P42058) Minor allergen Alt a 7 (Alt a VII)	7.00E-28	82_84

ID	Predicted Function	*E*-value	Start
OF_07_E4-CX	(P34244) Probable serine/threonine-protein kinase YKL101W (EC 2.7.1.37)	2.00E-08	81_83
OF_07_E5-EX	(Q86ZF9) Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS)	4.00E-87	62_64
OF_07_E8-BX	(P47943) Eukaryotic initiation factor 4A (EC 3.6.1.-) (ATP-dependent RNA helicase eIF4A) (eIF-4A)	4.00E-94	25_27
OF_07_F1-C-SP6	(Q8K4G6) Protein LRP16 (Fragment)	1.00E-24	22_24
OF_07_F3-D-SP6	(P49426) Glucan 1,3-beta-glucosidase precursor (EC 3.2.1.58) (Exo-beta 1,3 glucanase) (1,3-beta-D-glucanohydrolase)	1.00E-06	30_32
OF_07_F8-A-SP6	(Q96X45) Elongation factor 2 (EF-2) (Colonial temperature-sensitive 3)	E-126	NO
OF_07_F9-B-SP6	(Q04182) ATP-dependent permease PDR15	5.00E-68	77_79
OF_07_G12-G-SP6	(Q10361) Electron transfer protein 1, mitochondrial precursor	2.00E-14	58_60
OF_07_H10-H-SP6	(Q99PT0) Probable ATP-dependent RNA helicase DDX52 (EC 3.6.1.-) (DEAD box protein 52) (ATP-dependent RNA helicase ROK1-like) (rROK1L)	0.00000000000000OF_07	25_27
OF_07_H12-B-SP6	(P38715) NADPH-dependent aldose reductase GRE3 (EC 1.1.1.21) (NADPH-dependent aldo-keto reductase GRE3) (NADPH-dependent methylglyoxal reductase GRE3) (EC 1.1.1.-) (Xylose reductase) (EC 1.1.1.-) (Genes de respuesta a estres protein 3)	3.00E-32	NO
OF_07_H2-G-SP6	(P21375) Osmotic growth protein 1	4.00E-24	173_175
OF_07_H3-G-SP6	(P40989) 1,3-beta-glucan synthase component GLS2 (EC 2.4.1.34) (1,3-beta-D-glucan-UDP glucosyltransferase)	3.00E-33	91_93
OF_07_H4-B-SP6	(P28873) Benomyl/methotrexate resistance protein	1.00E-30	134_136
OF_07_H5-F-SP6	(Q12618) Acyl-CoA desaturase (EC 1.14.19.1) (Stearoyl-CoA desaturase) (Fatty acid desaturase) (Delta(9)-desaturase)	4.00E-28	7_9
OF_07_I10-A-SP6	(O13366) Pyruvate dehydrogenase E1 component alpha subunit, mitochondrial precursor (EC 1.2.4.1) (PDHE1-A)	1.00E-62	29_31
OF_07_I10-C-SP6	(P38720) 6-phosphogluconate dehydrogenase, decarboxylating 1 (EC 1.1.1.44)	E-108	86_88
OF_07_I10-D-SP6	(P09437) Cytochrome b2, mitochondrial precursor (EC 1.1.2.3) (L-lactate dehydrogenase	3.00E-45	152_154
OF_07_I10-F-SP6	(Q9VJ33) NEDD8 precursor (Ubiquitin-like protein Nedd8) (Neddylin)	1.00E-10	NO
OF_07_I3-D-SP6	(P53704) Glucosamine--fructose-6-phosphate aminotransferase	3.00E-45	119_121
OF_07_I3-F-SP6	(O74173) Elongation factor 1-beta (EF-1-beta)	3.00E-10	165_167
OF_07_I8-G-SP6	(P35571) Glycerol-3-phosphate dehydrogenase, mitochondrial precursor (EC 1.1.99.5) (GPD-M) (GPDH-M)	1.00E-28	NO
OF_07_I9-C-SP6	(P19449) Cellulose synthase catalytic subunit	6.00E-19	48_50
OF_7-H-x	(Q9Y713) Elongation factor 1-alpha (EF-1-alpha)	1.00E-38	NO
OF_12-H-y	(P40151) DNA-dependent ATPase MGS1 (Maintenance of genome stability protein 1)	7.00E-24	4_6
OF_16-C-x	(P0A9I1) Citrate lyase beta chain (EC 4.1.3.6) (Citrase beta chain) (Citrate (pro-3S)-lyase beta chain) (Citryl-CoA lyase subunit) (EC 4.1.3.34)	E-113	NO
OF_16-C-y	(P75793) Putative formate acetyltransferase 3 (EC 2.3.1.54) (Pyruvate formate-lyase 3)	E-1OF_07	38_40
OF0014	(Q8SHP7) NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3)	0.00E+00	24_26
OF0074	(P10255) Peptidyl-prolyl cis-trans isomerase, mitochondrial precursor (EC 5.2.1.8) (PPIase) (Rotamase) (Cyclophilin) (Cyclosporin A-binding protein) (CPH)	6.00E-60	201_203
OF0098	(P0AEG1) Dipeptide transport system permease protein dppC	8.00E-08	NO
OF0101	(P08622) Chaperone protein dnaJ (Heat shock protein J) (HSP40)	3.00E-55	50_52
OF0110	(P75905) Biofilm PGA synthesis N-glycosyltransferase pgaC (EC 2.4.-.-)	8.00E-67	NO
OF0117	(P28248) Deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP deaminase)	5.00E-29	66_68
OF0119	(P12996) Biotin synthase (EC 2.8.1.6) (Biotin synthetase)	2.00E-51	55_57

ID	Predicted Function	*E*-value	Start
OF0122	(P0AB91) Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-sensitive (EC 2.5.1.54) (Phospho-2-keto-3-deoxyheptonate aldolase) (DAHP synthetase) (3-deoxy-D-arabino-heptulosonate 7-phosphate synthase)	7.00E-48	NO
OF0133	(P59663) Malate synthase G (EC 2.3.3.9)	7.00E-58	171_173
OF0160	(P0A714) 2-keto-3-deoxygluconate permease (KDG permease)	3.00E-50	68_70
OF0163	(P24137) Oligopeptide transport ATP-binding protein oppF	2.00E-43	NO
OF0165	(P17117) Oxygen-insensitive NADPH nitroreductase (EC 1.-.-.-) (Modulator of drug activity A)	6.00E-73	97_99
OF0173	(P41442) Putative general secretion pathway protein G precursor (Protein transport protein hofG)	1.00E-31	174_176
OF0187	(P0AB95) Arsenical pump membrane protein (Arsenic efflux pump protein)	2.00E-71	NO
OF0188	(P21865) Sensor protein kdpD (EC 2.7.3.-)	2.00E-56	116_118
OF0198	(P76016) PTS-dependent dihydroxyacetone kinase operon regulatory protein	1.00E-78	122_124
OF0208	(P33136) Glucans biosynthesis protein G precursor	E-124	88_90
OF0216	(P0AC23) Probable formate transporter 1 (Formate channel 1)	2.00E-65	222_224
OF_07_B11-EX	(P45856) Probable 3-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157) (Beta-hydroxybutyryl-CoA dehydrogenase) (BHBD)	6.00E-50	NO
OF_07_C5-GX	(Q4I8Q4) Acyl-protein thioesterase 1 (EC 3.1.2.-)	1.00E-10	119_121
OF_07_I5-F-SP6	(Q9S735) Inorganic phosphate transporter 1-9 (AtPht1;9) (H(+)/Pi cotransporter)	2.00E-08	NO
OF_07_I7-H-SP6	(P28349) Nitrogen assimilation transcription factor nit-4	3.00E-06	138_140
ID	Predicted Function	E-value	Start
OF_07_I8-D-SP6	(Q12398) Probable transcription factor HMS1 (High-copy MEP suppressor protein 1)	0.00E+00	NO
OF_07_I9-A-SP6	(O66557) Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.62) (7,8-diamino-pelargonic acid aminotransferase) (DAPA aminotransferase)	2.00E-28	40_42
OF_07_I9-H-SP6	(P33416) Heat shock protein 78, mitochondrial precursor	3.00E-76	31_33
OF_05_5A-x	(P36938) Phosphoglucomutase (EC 5.4.2.2) (Glucose phosphomutase) (PGM)	7.00E-91	NO
OF_05_5A-y	(Q52675) Dimethyl sulfoxide/trimethylamine N-oxide reductase precursor (EC 1.7.2.3) (DMSO reductase) (DMSOR)	2.00E-46	NO
OF_05_5E-y	(P39188) Alu subfamily J sequence contamination warning entry	0.00E+00	24_26
OF_2-F-x	(P52697) 6-phosphogluconolactonase (EC 3.1.1.31) (6-P-gluconolactonase) (Pgl)	3.00E-21	NO
OF_14-F-y	(P09805) Killer toxin alpha/beta subunits precursor (RF2 protein)	1.00E-33	NO

AMBIGUOUS

ID	Predicted Function	*E*-value	Start
OF0064	(P15582) NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3) (NADH dehydrogenase subunit 4)	9.00E-72	NO
OF_07_A9-DX	(Q96X46) Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-glycerate hydro-lyase) (Allergen Pen c 22)	3.00E-63	NO
OF_12D-x	(P25553) Aldehyde dehydrogenase A (EC 1.2.1.22) (Lactaldehyde dehydrogenase) (Glycolaldehyde dehydrogenase) (EC 1.2.1.21)	2.00E-36	NO
OF_8B-y	(Q12150) Protein CSF1 (Cold sensitive for fermentation protein 1)	0.0	NO

ID represents the sequence identifier given to the UPT. Predicted function was based on TargetIdentifier annotation using BLASTx alignment to determine the most homologous annotated gene. The subsequent E-value of the BLASTx alignment is given. Start indicates the predicted ORF start site.

Appendix 2: Best Match and Putative Function of *Ophiostoma* Unique Putative Transcripts With Greater Than 0.4 Percent Frequency

O. floccosum

ID	ESTs	Bp	Accession	Description - best hit	Score	*E*-value	Putative Function
OF480064	59	692	YP_667826	NADH dehydrogenase subunit 4 [Verticillium dahliae] >gb|ABC60423.1| NADH dehydrogenase subunit 4 [Verticillium dahliae]	270	9.00E-71	NADH dehydrogenase subunit 4
OF480045	39	851	CAA32799.1	cytochrome oxidase subunit I [Neurospora crassa]	360	2.00E-98	cytochrome oxidase subunit I
OF480014	38	1293	NP_570156	NADH-ubiquinone oxidoreductase chain 5 [Hypocrea jecorina]	652	0.00E+00	NADH-ubiquinone oxidoreductase chain 5
OF480035	30	950	AAT74903.1	NADH dehydrogenase subunit 5 [Neurospora crassa] >gb|AAT74905.1| NADH dehydrogenase subunit 5 [Neurospora crassa]	125	1.00E-27	NADH dehydrogenase subunit 5
OF480174	30	1003	YP_001725660	AAA ATPase central domain protein [Escherichia coli ATCC 8739] >gb|ACA78333.1| AAA ATPase central domain protein [Escherichia coli ATCC 8739]	381	4.00E-104	ATPase central domain protein
OF480018	18	994	YP_667831.1	ATP synthase F0 subunit 9 [Verticillium dahliae] >gb|ABC60428.1| ATP synthase F0 subunit 9 [Verticillium dahliae]	90.5	4.00E-17	ATP synthase F0 subunit 9
OF480217	16	525	BAB33421.1	putative senescence-associated protein [Pisum sativum]	105	4.00E-21	putative senescence-associated protein
OF480223	14	622	YP_691358.1	DNA-directed RNA polymerase subunit beta' [Shigella flexneri 5 str. 8401]	125	2.00E-27	DNA-directed RNA polymerase subunit beta
OF480008	12	502	XP_001727264	hypothetical protein [Aspergillus oryzae RIB40]	35.4	1.40E+00	No significant match
OF480055	11			No significant match			No significant match
OF480007	10	523	XP_961100.1	hypothetical protein NCU03753 [Neurospora crassa OR74A] >sp|P22151|GRG1_NEUCR Glucose-repressible gene protein	96.7	5.00E-19	Possible glucose-repressible gene protein
OF480011	10			No significant match			No significant match
OF480069	10	407	CAD71042.1	putative protein [Neurospora crassa]	67.8	1.00E-09	No significant match
OF480136	10	649	NP_074923.1	orf301 [Podospora anserina] >emb|CAA38775.1| GIY Cytb i2 grp ID protein [Podospora anserina] >prf||1703266C cytochrome b intronic ORF 2	107	7.00E-22	cytochrome b intronic ORF
OF480134	9	2804	NP_415107.1	copper/silver efflux system, membrane component [Escherichia coli str. K-12 substr. MG1655]	555	1.00E-149	copper/silver efflux system, membrane component
OF480162	8	582	YP_688432.1	formate transporter [Shigella flexneri 5 str. 8401]	239	1.00E-61	formate transporter
OF480108	7	2069	1703265G	cytochrome oxidase I intronic ORF 11	495	5.00E-109	cytochrome oxidase I intronic ORF 11
OF480079	6	872	NP_074923.1	orf301 [Podospora anserina] >emb|CAA38775.1| GIY Cytb i2 grp ID protein [Podospora anserina] >prf||1703266C cytochrome b intronic ORF 2	340	7.00E-85	endonuclease
OF480163	6	1672	YP_539850.1	Heme-binding protein precursor (hemin-binding lipoprotein) [Escherichia coli UTI89]	403	7.00E-111	Heme-binding protein precursor
OF480194	6	409	XP_538689.2	similar to transcription elongation factor B polypeptide 3 binding protein 1 isoform 1 [Canis familiaris]	35.8	3.80E+00	No significan match
OF480004	5	835	XP_001228475	40s ribosomal protein S5 [Chaetomium globosum CBS 148.51] >gb|EAQ84144.1|	352	7.00E-96	40s ribosomal protein

OF480017	5	1826	XP_001907437	unnamed protein product [Podospora anserina] >sp|Q01520|EF1A_PODAN Elongation factor 1-alpha (EF-1-alpha)	808	0.00E+00	Elongation factor 1-alpha (EF-1-alpha)
OF480022	5	199	AAK67169.1	NADP-dependent mannitol dehydrogenase [Cladosporium fulvum]	38.9	1.20E-01	No significant match
OF480031	5	638	BAB33421.1	putative senescence-associated protein [Pisum sativum]	139	1.00E-21	putative senescence-associated protein
OF480222	5	313	ZP_00710871.1	COG1501: Alpha-glucosidases, family 31 of glycosyl hydrolases [Escherichia coli B171]	148	4.00E-34	Alpha-glucosidases, family 31 of glycosyl hydrolases

O. piliferum

ID	ESTs	Bp	Accession	Description - best hit	Score	*E*-value	Putative Function
OP0015	45	1535	XP_755489.1	F-box domain protein [Aspergillus fumigatus Af293]	111	6E-23	F-box domain protein
OP0018	67	1288	XP_363356.1	ubiquitin [Magnaporthe grisea 70-15] >ref|XP_001227036.1| polyubiquitin [Chaetomium globosum CBS 148.51]	590	4E-167	ubiquitin
OP0062	41	2297	AAA76693.1	phosphoenolpyruvate carboxykinase	735	0	phosphoenolpyruvate carboxykinase
OP0172	57	1517	XP_001909550	unnamed protein product [Podospora anserina] >emb|CAP70683.1| unnamed protein product [Podospora anserina]	568	4e-160	Unknown
OP0193	73	2607	XP_961298.1	heat shock protein 90 [Neurospora crassa OR74A] >emb|CAC28765.1|	999	0.0	heat shock protein 90
OP0209	39	3017	XP_962286.2	elongation factor 2 [Neurospora crassa OR74A] >sp|Q96X45.3|EF2_NEUCR Elongation factor 2 (EF-2) (Colonial temperature-sensitive 3)	1444	0.0	elongation factor 2
OP0370	50	1390	YP_001834698	Alcohol dehydrogenase GroES domain protein [Beijerinckia indica subsp. indica ATCC 9039]	455	2e-126	Alcohol dehydrogenase GroES domain protein
OP0410	40	1850	XP_361105.1	hypothetical protein MGG_03648 [Magnaporthe grisea 70-15] >emb|CAI67591.1| D-xylose reductase [Magnaporthe grisea]	441	3e-122	xylose reductase
OP0411	178	2621	XP_001910788	unnamed protein product [Podospora anserina] >emb|CAP72612.1| unnamed protein product [Podospora anserina]	245	3e-63	mismatched base pair and cruciform DNA recognition protein
OP0461	46	1472	XP_001228473	ADP,ATP carrier protein [Chaetomium globosum CBS 148.51]	551	2e-155	ADP,ATP carrier protein
OP0621	60	1445	XP_381616.1	hypothetical protein FG01440.1 [Gibberella zeae PH-1]	269	2e-70	Unknown
OP0741	40	2217	XP_001910057	unnamed protein product [Podospora anserina] >emb|CAP71191.1|	903	0	ATP synthase beta chain, mitochondrial precursor
OP0906	65	2052	XP_363864.1	conserved hypothetical protein [Magnaporthe grisea 70-15]	799	0	acyl-CoA desaturase 1
OP1172	43	2637	XP_001905894	unnamed protein product [Podospora anserina] >emb|CAP66560.1| unnamed protein product [Podospora anserina]	526	4E-147	2-nitropropane dioxygenase family oxidoreductase
OP1239	52	2168	XP_001909556	>emb|CAP70689.1| unnamed protein product [Podospora anserina]	690	2E-165	zinc finger protein zpr1
OP1359	168	2485	XP_001225918	heat shock 70 kDa protein [Chaetomium globosum CBS 148.51] >gb|EAQ87009.1|	1051	0	heat shock 70 kDa protein
OP1402	47	1851	AAR16425.1	translation elongation factor 1 alpha [Metarhizium anisopliae]	846	0	translation elongation factor 1 alpha

O. clavigera

ID	ESTs	Bp	Accession	Description - best hit	Score	*E*-value	Putative Function
7	80	1587	XP_956174.2	cytoplasmic ribosomal protein subunit S3 [Neurospora crassa OR74A]	446	3E-123	cytoplasmic ribosomal protein subunit S3
CV0511	33	952	XP_965678.1	hypothetical protein NCU02547 [Neurospora crassa OR74A] >gb|EAA36442.1	233	3E-59	Unknown
CV0814	60	1175	XP_389688.1	hypothetical protein FG09512.1 [Gibberella zeae PH-1]	217	2E-54	branched-chain amino acid aminotransferase, cytosolic
CV0914	34			No significant match			
CV1124	38	1039	XP_962348.1	ubiquinol-cytochrome c reductase iron-sulfur subunit, mitochondrial [Neurospora crassa OR74A]	345	4E-93	ubiquinol-cytochrome c reductase iron-sulfur subunit, mitochondrial
CV1135	29			No significant match			

O. novo-ulmi

ID	ESTs	Bp	Accession	Description - best hit	Score	*E*-value	Putative Function
ANU0004	26	1224	XP_750388.2	translation elongation factor EF-1 alpha subunit [Aspergillus fumigatus Af293]	330	2E-88	translation elongation factor EF-1 alpha subunit
ANU0126	27	1341	XP_386504.1	hypothetical protein FG06328.1 [Gibberella zeae PH-1]	612	2E-173	xylulose-5-phosphate phosphoketolase
ANU0144	21	866	AAK58048.1	unknown [Ophiostoma novo-ulmi]	246	8E-57	Unknown
ANU0214	26	976	AAK58048.1	unknown [Ophiostoma novo-ulmi]	246	8E-57	Unknown
ANU0279	20	1606	XP_368203.1	hypothetical protein MGG_01041 [Magnaporthe grisea 70-15]	100	3E-08	No significant match
ANU0306	33	998	Q7S045.1	Non-histone chromosomal protein 6	174	2E-41	Non-histone chromosomal protein 6
ANU0308	20	830	XP_370262.2	hypothetical protein MGG_06759 [Magnaporthe grisea 70-15] >gb|EDJ98363.1| hypothetical protein MGG_06759 [Magnaporthe grisea 70-15]	213	2E-53	molecular chaperone Mod-E/Hsp90
ANU0360	30	1242	XP_001402218	hypothetical protein An04g08600 [Aspergillus niger] >emb|CAK38947.1| unnamed protein product [Aspergillus niger]	37.4	2.4	No significant match
ANU0367	20	922	XP_368812.1	hypothetical protein MGG_00432 [Magnaporthe grisea 70-15]	194	2E-47	Unknown
ANU0397	23			No significant match			
ANU0435	36	1592	ABF84060.1	C-4 sterol methyl oxidase [Chaetomium globosum] >gb|ABF84061.1| C-4 sterol methyl oxidase [Chaetomium globosum]	279	3E-73	C-4 sterol methyl oxidase
ANU0444	39	966	XP_386433.1	G3P_COLGL Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) [Gibberella zeae PH-1]	288	6E-76	glyceraldehyde-3-phosphate dehydrogenase
ANU0466	50			No significant match			
ANU0485	40	978	Q06153.1	Cerato-ulmin precursor (CU) (Dutch elm disease toxin)	201	7E-50	cerato-ulmin
ANU0548	22	927	XP_962543.1	hypothetical protein NCU08330 [Neurospora crassa OR74A]	258	9E-67	Unknown
ANU0553	23	911	XP_001224695	hypothetical protein CHGG_07039 [Chaetomium globosum CBS 148.51]	104	1E-20	Unknown
ANU0590	32	914	CAJ44125.1	LEA dehydrin-like protein [Medicago sativa subsp. falcata]	57	0.000003	No significant match
ANU0656	31	1134	XP_001228673	hypothetical protein CHGG_02157 [Chaetomium globosum CBS 148.51]	164	2E-38	Unknown
ANU0680	33	1248	XP_001219658	pyruvate decarboxylase [Chaetomium globosum CBS 148.51]	405	2E-104	pyruvate decarboxylase

ID represents the sequence identifier given to the UPT. ESTs is the number of EST fragments in each UPT. Note: *O. piceae* EST data was not included in the BLASTx analysis.

Appendix 3: Putative Proteases In *Ophiostoma* Datasets

	Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
Aspartic Peptidases	CV0967	A01A	MER000928	2.50E-44	endothiapepsin	*Cryphonectria parasitica*	<D/H, <F/S/Y, D225	
	ANU_435	A01A	MER019988	1.30E-39	saccharopepsin	*Aspergillus oryzae*	<D/H, <F/S/Y, <D	
	YLMW23_A01_1_001	A01A	MER019988	1.50E-34	saccharopepsin	*Aspergillus oryzae*	<D/H, <F/S/Y,	
	CV0453	A01A	MER071452	1.80E-44	subfamily A1A unassigned peptidases	*Magnaporthe grisea*	<D/H, <F/S/Y, D384	
	CV0011	A01A	MER075981	8.10E-68	PepAa peptidase (*Aspergillus niger*)	*Magnaporthe grisea*	<D/H, <F/S/Y, D267	
	ANU_334	A01A	MER000928	7.00E-59	endothiapepsin	*Cryphonectria parasitica*	<D/H, <F/S/Y, D370	
	CV1129	A01A	MER019988	1.60E-79	saccharopepsin	*Aspergillus oryzae*	<D/H, <F/S/Y, D312	
	ANU_310	A01A	MER019988	1.20E-81	saccharopepsin	*Aspergillus fumigatus*	<D/H, <F/S/Y, D456	
	ANU_38	A01A	MER082513	2.60E-28	subfamily A1A non-peptidase homologues	*Aspergillus fumigatus*	<D/H, <F/S/Y,	
	ANU_19	A01A	MER090758	3.20E-23	PepAb peptidase (*Aspergillus niger*)	*Aspergillus niger*	<D/H, <F/S/Y,	
	LMW_47_D03_20_027	A01A	MER019988	8.60E-28	saccharopepsin	*Aspergillus oryzae*	<D/H, <F/S/Y, A293D	
	Clv_114217757_gb_EE724431	A22B	MER065262	1.30E-16	subfamily A22B unassigned peptidases	*Gibberella zeae*	<D, D443	
Cysteine Peptidases	LMW_60_F_C01_3_002	A24A	MER083047	1.70E-41	subfamily A24A unassigned peptidases	*Ralstonia solanacearum*	X27D, D88	
	Clv_114219322_gb_EE725996	C12	MER014615	1.90E-39	Uch2 peptidase (*Schizosaccharomyces pombe*)	*Neurospora crassa*	Q436, C442, H523, >D	
	CV0421	C12	MER014615	5.40E-41	Uch2 peptidase (*Schizosaccharomyces pombe*)	*Neurospora crassa*	<Q, C422, H503, D518	
	Clv_114219797_gb_EE726471	C12	MER033298	4.10E-31	family C12 unassigned peptidases	*Neurospora crassa*	Q173, C179, H254, E270D	
	07_F2-H-SP6	C14	MER039477	5.00E-23	metacaspase-1	*Gibberella zeae*	<H, <C	
	OF480033	C14	MER039477	1.70E-37	metacaspase-1	*Gibberella zeae*	<H, <C	
	Clv_114221981_gb_EE728655	C14B	MER039472	4.30E-63	metacaspase-1	*Podospora anserina*	H468, >C	
	LMW_58_F_A08_57_053	C14B	MER093173	4.40E-15	metacaspase-1	*Aspergillus niger*	<H, <C	
	LMW_57_F_E01_5_003	C14B	MER039477	1.30E-12	metacaspase-1	*Gibberella zeae*	<H, <C	
	CV0947	C15	MER065415	3.10E-11	family C15 unassigned peptidases	*Gibberella zeae*	<E, C70, >H	

Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
Clv_114222742_gb_EE729416	C19	MER031259	2.50E-59	family C19 non-peptidase homologues	*Podospora anserina*	<N, <C, H603, >D/N	
Clv_114220647_gb_EE727321	C19	MER090088	3.00E-06	family C19 non-peptidase homologues	*Emericella nidulans*	<N, <C, T396H, Q430D/N	
PiL_90613116_gb_EB046559	C26	MER060647	2.40E-24	dihydro-orotase (N-terminal unit) (*Homo sapiens*-type)	*Homo sapiens*	>C, >H	
LMW_29_F11_86_091	C26	MER064212	3.20E-40	family C26 unassigned peptidases	*Gibberella zeae*	C297, H383	
OP1086	C26	MER064637	1.40E-79	family C26 unassigned peptidases	*Gibberella zeae*	C154, H246	
YeastLMW8_F05_38_043	C26	MER069897	2.10E-12	family C26 unassigned peptidases	*Bacillus clausii*	<C, H412	
OP1126	C26	MER076861	1.70E-57	family C26 unassigned peptidases	*Marinobacter hydrocarbonoclas-ticus*	C222, H316	
Clv_114218507_gb_EE725181	C26	MER090327	3.70E-26	family C26 unassigned peptidases	*Geobacillus thermodenitrificans*	C382, >H	
Clv_114222869_gb_EE729543	C26	MER101518	3.00E-06	family C26 unassigned peptidases	*Acaryochloris marina*	<C, H366	
Clv_114218507_gb_EE725181	C26	MER066916	2.90E-06	family C26 unassigned peptidases	*Pyrococcus furiosus*	<C, H217	
OP0284	C26	MER069897	4.20E-46	family C26 unassigned peptidases	*Bacillus clausii*	C90, H222	
OF480177	C44	MER029814	8.50E-12	family C44 non-peptidase homologues	*Escherichia coli*	C844	
PIL_90615213_gb_EB048656	C44	MER033304	6.50E-112	family C44 non-peptidase homologues	*Neurospora crassa*	C193	
PIL_90614304_gb_EB047747	C44	MER033362	6.30E-29	family C44 non-peptidase homologues	*Neurospora crassa*	<C	
LMW_38_F08_62_063	C54	MER050319	1.90E-19	ATG4 peptidase	*Cryphonectria parasitica*	<Y, <C, D359, H361	
CV0127	C56	MER028876	2.70E-06	family C56 non-peptidase homologues	*Bacillus cereus*	A591E, C624, S625H	
Clv_114218875_gb_EE725549	C56	MER029114	3.80E-38	family C56 unassigned peptidases	*Neurospora crassa*	C335E, H336C	
CV0581	C56	MER064735	4.90E-36	family C56 non-peptidase homologues	*Gibberella zeae*	<E, F627C, S628H	
OP1181	C56	MER064735	2.60E-52	family C56 non-peptidase homologues	*Gibberella zeae*	A545E, F579C, S580H	
YeastLMW7_A12_89_085	C56	MER065250	1.10E-29	family C56 unassigned peptidases	*Gibberella zeae*	M232E, C259, H260	
CV0679	C88	MER116559	1.30E-20	OUT-1 peptidase (*Homo sapiens*-type)	*Homo sapiens*	D375, C378, H480	

	Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
Glutamic Peptidases	Clv_114222566_gb_EE729240	G01	MER090150	2.30E-64	aspergilloglutamic peptidase	*Emericella nidulans*	Q280, E365	
Metallo Peptidases	Clv_114223446_gb_EE730120	M01	MER028450	8.30E-70	family M1 unassigned peptidases	Neurospora crassa	E89, >Y H88, H92, E111	
	Clv_114220681_gb_EE727355	M13	MER109135	6.60E-59	family M13 unassigned peptidases	*Magnaporthe grisea*	<E, D165	<H, A101H, E161
	PiL_90613267_gb_EB046710	M13	MER109135	6.70E-50	family M13 unassigned peptidases	*Magnaporthe grisea*	E200, D267	P199H, H203, E263
	LMW_46_D03_20_027	M14A	MER065441	6.80E-21	subfamily M14A non-peptidase homologues	*Gibberella zeae*	<R, K349E	<H, <E, <H
	YLMW13_D10_76_078	M16B	MER043990	5.80E-63	mitochondrial processing peptidase beta subunit domain 2	*Aspergillus fumigatus*		
	CV0486	M18	MER100435	2.40E-111	aspartyl aminopeptidase	*Magnaporthe grisea*	<D, E176	<H, D139, E177, D235, H329
	CV0465	M19	MER033336	6.90E-104	family M19 unassigned peptidases	*Neurospora crassa*	<H, <D, <E, H199, H231	
	OP0409	M19	MER033336	7.90E-83	family M19 unassigned peptidases	*Neurospora crassa*	H347, D349, E463, H538, >H	
	PiL_90612160_gb_EB045603	M19	MER093118	7.60E-85	family M19 unassigned peptidases	*Aspergillus niger*	H201, D203, E317, >H, >H	
	Clv_114221107_gb_EE727781	M20A	MER064205	3.10E-55	subfamily M20A unassigned peptidases	*Gibberella zeae*	<D, X290E	<H, D256, X291E, E317, >H
	Clv_114219777_gb_EE726451	M20A	MER093085	3.70E-114	carnosine dipeptidase II	*Aspergillus niger*	<D, <E <H, <D, <E, <D/E, H316	
	OP0428	M20A	MER093085	5.90E-218	carnosine dipeptidase II	*Aspergillus niger*	D238, E304	H236, D270, E305, D333, H583
	ANU_424	M20A	MER093085	3.00E-106	carnosine dipeptidase II	*Aspergillus niger*	<D, <E <H, <D, <E, <D/E, P355H	

Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
07_C8-AX	M20D	MER092183	3.10E-32	subfamily M20D non-peptidase homologues	*Chaetomium globosum*	<D, <E <D/E, <D, <E, H117	
CV0793	M20D	MER092128	4.30E-65	subfamily M20D non-peptidase homologues	*Phaeosphaeria nodorum*	<D, <E <D/E, <D, <E, H588	
PiL_90611981_gb_EB045424	M20D	MER092183	4.10E-70	subfamily M20D non-peptidase homologues	*Chaetomium globosum*	D104, X152E	E102, N116D, X153E, >H
PiL_90608793_gb_EB042238	M20X	MER033110	4.30E-27	family M20 non-peptidase homologues	*Magnaporthe grisea*	D364	H362, N373D, >E, >D/E, >H
PIL_90615832_gb_EB049275	M22	MER097754	1.90E-27	NO MATCH		<H, <H	
CV0583	M23B	MER060269	3.40E-52	subfamily M23B non-peptidase homologues	*Gibberella zeae*	X656H	V574H, R579D, X658H
PiL_90608588_gb_EB042033	M23B	MER060276	2.30E-07	subfamily M23B non-peptidase homologues	*Gibberella zeae*	<H	<H, <D, <H
PiL_90610376_gb_EB043819	M23B	MER060370	4.10E-08	subfamily M23B non-peptidase homologues	*Metallosphaera sedula*	F329H	G250H, D256, F329H
PiL_90612645_gb_EB046088	M23B	MER064182	5.40E-08	subfamily M23B non-peptidase homologues	*Gibberella zeae*	D133H	<H, I27D, L135H
PiL_90608750_gb_EB042195	M23B	MER064183	6.60E-09	subfamily M23B non-peptidase homologues	*Gibberella zeae*	N323H	A213H, I220D, K325H
PiL_90609686_gb_EB043129	M23B	MER064183	4.70E-24	subfamily M23B non-peptidase homologues	*Gibberella zeae*	>H	E283H, K290D, >H
PiL_90609686_gb_EB043129	M23B	MER064183	1.30E-18	subfamily M23B non-peptidase homologues	*Gibberella zeae*	A661H <H, <D, K663H	
PIL_90615401_gb_EB048844	M23B	MER064185	4.50E-22	subfamily M23B non-peptidase homologues	*Gibberella zeae*	<H	<H, <D, <H
PiL_90612529_gb_EB045972	M23B	MER065286	7.80E-46	subfamily M23B non-peptidase homologues	*Gibberella zeae*	<H	<H, <D, <H
YLMW25_C07_51_050	M23B	MER065598	1.20E-32	subfamily M23B non-peptidase homologues	*Escherichia coli*	<H	<H, <D, <H
PIL_90614064_gb_EB047507	M23B	MER068425	4.90E-06	subfamily M23B non-peptidase homologues	*Natronomonas pharaonis*	G79H	<H, <D, L81H
PiL_90610136_gb_EB043579	M23B	MER090283	3.60E-06	subfamily M23B non-peptidase homologues	*Emericella nidulans*	Q209H	<H, <D, T211H
PiL_90612118_gb_EB045561	M23B	MER114044	1.70E-07	subfamily M23B non-peptidase homologues	*Neisseria meningitidis*	<H	<H, <D, <H
PiL_90613607_gb_EB047050	M23B	MER114487	1.90E-09	subfamily M23B non-peptidase homologues	*Emericella nidulans*	<H	<H, <D, <H
CIv_114219275_gb_EE725949	M24A	MER033339	1.20E-120	methionyl aminopeptidase 1	*Neurospora crassa*	H203	D220, D231, H297, E330, E361
LMW_34_C07_51_050	M24A	MER033339	1.40E-62	methionyl aminopeptidase 1	*Neurospora crassa*	<H	<D, <D, H268, E301, E332

Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
PIL_90614805_gb_EB048248	M24A	MER033339	3.20E-63	methionyl aminopeptidase 1	*Neurospora crassa*	H479	D496, D507, >H, >E, >E
CV0833	M24A	MER033340	2.10E-71	methionyl aminopeptidase 2	*Neurospora crassa*	<H	<D, <D, H150, E183, E278
PIL_90615783_gb_EB049226	M24A	MER033340	5.80E-95	methionyl aminopeptidase 2	*Neurospora crassa*	<H	<D, <D, H276, E309, E405
OP0604	M24B	MER033341	3.80E-24	aminopeptidase P1	Neurospora crassa	>H, >H, >H	>D, >D, >H, >E, >E
PiL_90608406_gb_EB041851	M24B	MER090160	5.30E-39	subfamily M24B non-peptidase homologues	*Emericella nidulans*	<H, R62H, K73H	<D, <D, Y66H, T96E, T117E
Clv_114220626_gb_EE727300	M24B	MER090240	5.90E-33	aminopeptidase P homologue	*Emericella nidulans*	<H, H243, Q254H <D, <D, H247, E274, E296	
Clv_114219182_gb_EE725856	M24X	MER064643	3.30E-05	proliferation-associated protein 2G4, 38kDa (*Rattus norvegicus*)	*Rattus norvegicus*	<H	<D, <D, W410H, -448E, >E
CV0986	M24X	MER064643	2.50E-23	proliferation-associated protein 2G4, 38kDa (*Rattus norvegicus*)	*Rattus norvegicus*	<H	<D, <D, W319H, -357E, L447E
Clv_114219182_gb_EE725856	M24X	MER079635	3.30E-05	proliferation-association protein 1	*Caenorhabditis briggsae*	<H	<D, <D, <H, <E, L151E
ANU_457	M24X	MER064643	3.10E-18	proliferation-associated protein 2G4, 38kDa (*Rattus norvegicus*)	*Rattus norvegicus*	<H	<D, <D, <H, <E, L298E
PIL_90615103_gb_EB048546	M28A	MER090048	2.70E-30	aminopeptidase Y	*Emericella nidulans*	<D, <E <H, <D, <E, <D/E, H56	
PiL_90613088_gb_EB046531	M28B	MER079120	3.50E-14	subfamily M28B unassigned peptidases	*Aspergillus oryzae*	<D, H97E	<H, <D, R98E, D130, >H
OP1031	M28B	MER079120	6.70E-23	subfamily M28B unassigned peptidases	*Aspergillus oryzae*	<D, <E <H, <D, <E, <D/E, H40	
Clv_114219594_gb_EE726268	M28E	MER033345	1.90E-106	Mername-AA063 peptidase	*Neurospora crassa*	D205, E260	H203, D222, E261, D288, H370

Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
CV0973	M28X	MER033123	6.60E-63	Mername-AA103 peptidase	*Magnaporthe grisea*	<D, E239	<H, <D, E240, D286, H373
OP0664	M38	MER033125	5.20E-112	family M38 non-peptidase homologues	*Magnaporthe grisea*	A584H, F586H	
OP1161	M38	MER033407	4.50E-105	family M38 non-peptidase homologues	*Neurospora crassa*	<H, <H, K5, D49H, Y98H	
PiL_90612045_gb_EB045488	M38	MER058115	7.00E-05	family M38 non-peptidase homologues	*Sphingomonas wittichii*	>D	<H, <H, L405K, R437H, H451
YLMW25_G05_39_036	M38	MER060262	6.20E-31	family M38 non-peptidase homologues	*Gibberella zeae*	>D	H201, H203, >K, >H, >H
CV0834	M38	MER065489	3.80E-25	Pro-Hyp dipeptidase	*Acidobacteria bacterium*	D279	<H, <H, <K, Q194H, H203
CV1069	M38	MER075387	6.10E-20	family M38 non-peptidase homologues	*Azoarcus sp. BH72*	D71	<H, <H, <K, <H, H1
07_I7-F-SP6	M38	MER076986	1.60E-43	family M38 non-peptidase homologues	*Acidovorax sp. JS42*	S270D	<H, <H, A125K, L166H, K200H
PiL_90613650_gb_EB047093	M38	MER090098	1.10E-07	family M38 non-peptidase homologues	*Emericella nidulans*	<D	<H, <H, <K, <H, <H
OP1326	M38	MER090098	1.70E-53	family M38 non-peptidase homologues	*Emericella nidulans*	D158	<H, <H, <K, H42, H83
YeastLMW7_C11_83_082	M38	MER090310	7.80E-66	urease	*Emericella nidulans*	<D	<H, <H, <K, <H, <H
PiL_90610577_gb_EB044020	M38	MER090310	2.30E-118	urease	*Emericella nidulans*	D145	<H, <H, L3K, H57, H105
PIL_90614535_gb_EB047978	M38	MER090322	1.10E-74	family M38 non-peptidase homologues	*Emericella nidulans*	>D	H201, H203, K292, K340H, >H
YLMW23_G08_63_056	M38	MER096733	1.40E-06	family M38 non-peptidase homologues	*Maricaulis maris*	>D	<H, <H, S92K, K125H, H156
YeastLMW7_C11_83_082	M38	MER090310	3.60E-05	urease	*Emericella nidulans*	<D	<H, <H, <K, <H, <H
OF480024	M41	MER014133	7.40E-50	i-AAA peptidase	*Neurospora crassa*	E515	H514, H518, >D
OP0100	M49	MER113012	3.00E-81	dipeptidyl-peptidase III	*Gibberella zeae*	>E	>H, >H, >E
PiL_90612935_gb_EB046378	M67A	MER025109	7.60E-07	family M67 non-peptidase homologues	*Bombyx mori*	T105E >H, >H, >D	

	Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
Serine Peptidases	Clv_114221869_gb_EE728543	M67A	MER033353	2.30E-51	Mername-AA063 peptidase	*Neurospora crassa*	<E	<H, <H, <D
	OP0912	M67A	MER033384	1.30E-52	subfamily M67A non-peptidase homologues	*Neurospora crassa*	R308E	T369H, S371H, S378D
	LMW_53_F_D08_60_062	M67A	MER060221	1.90E-38	Jab1/MPN domain metalloenzyme	*Gibberella zeae*	<E	<H, <H, <D
	LMW_43_G03_23_021	M67A	MER033353	3.00E-44	Mername-AA063 peptidase	*Neurospora crassa*	<E	<H, <H, <D
	LMW_32_D01_4_010	M67C	MER090223	5.20E-49	subfamily M67C unassigned peptidases	*Emericella nidulans*	K351E	H406, H408, D419
	PiL_90612935_gb_EB046378_1_EB046378	M67X	MER021885	3.40E-24	Mername-AA168 protein	*Mus musculus*	<E	T347H, S348H, F359D
	CV0458	M67X	MER030133	1.10E-05	eukaryotic translation initiation factor 3 subunit 5	*Homo sapiens*	<E	<H, <H, <D
	OF480125	M74	MER01298	3.30E-119	murein endopeptidase	*Escherichia coli*	H605	H509, D516, H607
	CV0823	S08A	MER080485	6.40E-55	cerevisin	*Chaetomium thermophilum*	<D, <H, <N, S45	
	CV1049	S08A	MER080485	1.50E-35	cerevisin	*Chaetomium thermophilum*	<D, <H, <N, <S	
	YLMW12_D08_60_062	S08A	MER082928	7.60E-21	subfamily S8A non-peptidase homologues	*Acidiphilium cryptum*	<D, <H, <N, G40S	
	PIL_90614095_gb_EB047538	S09X	MER033025	3.90E-19	family S9 unassigned peptidases	*Magnaporthe grisea*	<S, <D, G265H	
	Clv_114219978_gb_EE726652	S09X	MER033307	1.70E-15	family S9 unassigned peptidases	*Neurospora crassa*	<S, D211, H241	
	Clv_114219978_gb_EE726652	S09X	MER033307	3.00E-09	family S9 unassigned peptidases	*Neurospora crassa*	X682S, >D, >H	
	CV0479	S09X	MER033307	2.20E-29	family S9 unassigned peptidases	*Neurospora crassa*	S402, >D, >H	
	CV0479	S09X	MER033307	3.60E-13	family S9 unassigned peptidases	*Neurospora crassa*	<S, D979, H1009	
	CV0407	S09X	MER034143	1.70E-37	family S9 unassigned peptidases	*Mycobacterium bovis*	S240, D357, H387	
	CV0683	S09X	MER035051	3.60E-33	family S9 non-peptidase homologues	*Gibberella zeae*	<S, <D, H876	
	07_C5-GX	S09X	MER047694	7.10E-09	family S9 unassigned peptidases	*Trichodesmium erythraeum*	V555S, D596, >H	
	ANU_631	S09X	MER065419	2.00E-52	family S9 unassigned peptidases	*Gibberella zeae*	<S, D186, H216	
	PiL_90608471_gb_EB041916	S09X	MER066433	1.70E-05	family S9 unassigned peptidases	*Ignicoccus hospitalis*	<S, D685, H713	
	OP0482	S09X	MER067920	5.50E-05	family S9 non-peptidase homologues	*Myxococcus xanthus*	>S, >D, >H	
	LMW_47_G03_23_021	S09X	MER072467	5.40E-05	family S9 unassigned peptidases	*Rickettsia akari*	S90, >D, >H	

Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
CV0283	S09X	MER075602	1.30E-08	family S9 non-peptidase homologues	*Nocardioides sp. JS614*	S649, >D, >H	
Ylmw3_B07_50_057	S09X	MER078361	3.80E-09	family S9 unassigned peptidases	*Methylibium petroleiphilum*	<S, D305, H337	
PiL_90610459_gb_EB043902	S09X	MER093135	3.70E-11	family S9 non-peptidase homologues	*Aspergillus niger*	C319S, E367D, >H	
PiL_90613484_gb_EB046927	S09X	MER101613	7.80E-13	family S9 non-peptidase homologues	*Acaryochloris marina*	T268S, G375D, >H	
PIL_90615672_gb_EB049115	S09X	MER101613	8.30E-05	family S9 non-peptidase homologues	*Acaryochloris marina*	<S, >D, >H	
PiL_90617248_gb_EB050691	S09X	MER101613	1.90E-15	family S9 non-peptidase homologues	*Acaryochloris marina*	E261S, >D, >H	
PiL_90610896_gb_EB044339	S09X	MER101638	1.90E-06	family S9 non-peptidase homologues	*Acaryochloris marina*	E67S, >D, >H	
CV0482	S09X	MER101643	1.20E-15	family S9 non-peptidase homologues	*Acaryochloris marina*	A404S, S480D, D526H	
PiL_90613512_gb_EB046955	S09X	MER101644	2.80E-10	family S9 non-peptidase homologues	*Acaryochloris marina*	<S, A476D, G513H	
07_D9-HX	S09X	MER101652	7.30E-14	family S9 unassigned peptidases	*Acaryochloris marina*	G88S, >D, >H	
LMW_41_B05_34_042	S09X	MER101668	3.50E-07	family S9 unassigned peptidases	*Acaryochloris marina*	<S, G66D, G105H	
CV0277	S09X	MER101719	1.60E-06	family S9 unassigned peptidases	*Acaryochloris marina*	<S, V730D, >H	
OP1158	S09X	MER101719	4.00E-05	family S9 unassigned peptidases	*Acaryochloris marina*	G359S, >D, >H	
PiL_90613888_gb_EB047331	S09X	MER101720	8.20E-05	family S9 unassigned peptidases	*Acaryochloris marina*	<S, A629D, >H	
OP0584	S09X	MER101724	8.20E-07	family S9 unassigned peptidases	*Acaryochloris marina*	-492S, >D, >H	
CV0303	S09X	MER101750	1.50E-13	family S9 unassigned peptidases	*Acaryochloris marina*	<S, A550D, >H	
PIL_90615478_gb_EB048921	S09X	MER101750	1.30E-12	family S9 unassigned peptidases	*Acaryochloris marina*	A23S, >D, >H	
ANU_366	S09X	MER101750	5.40E-08	family S9 unassigned peptidases	*Acaryochloris marina*	432-734 <S, V489D, >H	
CV0971	S09X	MER101958	2.90E-05	family S9 unassigned peptidases	*Cyanothece sp. ATCC 51142*	<S, >D, >H	
Clv_114218419_gb_EE725093	S09X	MER114060	1.10E-25	family S9 unassigned peptidases	*Coxiella burnetii*	S270, D327, >H	
PiL_90611895_gb_EB045338	S09X	MER114216	1.50E-05	family S9 non-peptidase homologues	*Sorangium cellulosum*	<S, I61D, >H	
Clv_114221825_gb_EE728499	S09X	MER114216	2.80E-16	family S9 non-peptidase homologues	*Sorangium cellulosum*	P185S, V293D, >H	
PiL_90616198_gb_EB049641	S09X	MER114216	4.00E-06	family S9 non-peptidase homologues	*Sorangium cellulosum*	>S, >D, >H	

Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
PiL_90616748_gb_EB050191	S09X	MER114450	8.20E-05	family S9 non-peptidase homologues	*Sorangium cellulosum*	<S, >D, >H	
PiL_90609108_gb_EB042553	S09X	MER114493	2.60E-17	family S9 non-peptidase homologues	*Emericella nidulans*	E137S, G221D, >H	
PiL_90609805_gb_EB043248	S09X	MER114556	3.30E-07	family S9 non-peptidase homologues	*Emericella nidulans*	X477S, >D, >H	
OP0667	S09X	MER115415	4.70E-18	family S9 non-peptidase homologues	*Microcystis aeruginosa*	G143S, I255D, >H	
LMW_40_E06_45_040	S09X	MER064110	5.50E-61	esterase D (*Homo sapiens*)	*Gibberella zeae*	S120, D207, H242	
PIL_90614450_gb_EB047893	S09X	MER068641	3.00E-05	family S9 non-peptidase homologues	*Sinorhizobium meliloti*	C217S, >D, >H	
LMW_49_D09_68_074	S09X	MER101719	6.40E-12	family S9 unassigned peptidases	*Acaryochloris marina*	G192S, >D, >H	
OP1057	S09X	MER114450	2.30E-21	family S9 non-peptidase homologues	*Sorangium cellulosum*	A493S, L602D, >H	
ANU_666	S09X	MER033309	3.20E-30	family S9 unassigned peptidases	*Neurospora crassa*	<S, D393, Y416H	
LMW_60_F_E05_37_035	S09X	MER033309	1.70E-56	family S9 unassigned peptidases	*Neurospora crassa*	S116, D228, Y251H	
OP0004	S09X	MER114250	2.90E-06	family S9 non-peptidase homologues	*Sorangium cellulosum*	A847S, >D, >H	
LMW_135_C07_51_050	S09X	MER116474	2.90E-16	family S9 non-peptidase homologues	*Caulobacter sp. K31*	S19, >D, >H	
CV1090	S10	MER028379	4.50E-61	Mername-AA085 peptidase	Neurospora crassa	<S, D463, H532	
CV1146	S10	MER073560	4.00E-105	carboxypeptidase Y	*Magnaporthe grisea*	<S, D323, H393	
LMW_38_E02_13_007	S10	MER073561	6.20E-13	kex carboxypeptidase	*Magnaporthe grisea*	<S, <D, H25	
OF480180	S11	MER088022	5.40E-28	family S11 unassigned peptidases	*Escherichia coli*	<S, <K, L83S	
OF480193	S11	MER088022	1.50E-66	family S11 unassigned peptidases	*Escherichia coli*	V468S, C471K, L532S	
Clv_114222046_gb_EE728720	S12	MER064162	4.80E-09	family S12 unassigned peptidases	*Gibberella zeae*	<S, <K, Y231	
onuper_gi_117166525_gb_EG355998	S12	MER090172	1.40E-17	family S12 unassigned peptidases	*Emericella nidulans*	<S, <K, Y204	
PIL_90614349_gb_EB047792	S16	MER059848	9.10E-30	family S16 non-peptidase homologues	*Methanococcus vannielii*	A217S, P240K/R	
Clv_114218478_gb_EE725152	S16	MER090002	2.50E-69	PIM1 peptidase	*Emericella nidulans*	S55, K98	
PiL_90609231_gb_EB042676	S16	MER101293	1.00E-25	family S16 non-peptidase homologues	*Methanococcus maripaludis*	A146S, P169K/R	
LMW_49_C03_19_018	S16	MER116274	1.90E-52	family S16 non-peptidase homologues	*Pseudomonas putida*	<S, G134K/R	
OP0871	S16	MER059848	1.50E-31	family S16 non-peptidase homologues	*Methanococcus vannielii*	A536S, A559K/R	

Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
LMW_27_A04_25_021	S16	MER090002	1.60E-07	PIM1 peptidase	*Emericella nidulans*	<S, <K/R	
Clv_114219275_gb_EE725949_1_EE725949	S24	MER021263	2.00E-07	family S24 unassigned peptidases	*Staphylococcus aureus*	<S, C350K	
YLMW24_A05_33_033	S26A	MER064143	2.30E-35	subfamily S26A unassigned peptidases	*Gibberella zeae*	<S, K213	
ANU_642	S26B	MER032881	2.10E-25	signalase (eukaryote) 21 kDa component	*Magnaporthe grisea*	<S, <H	
CV0938	S28	MER093133	6.10E-72	family S28 unassigned peptidases	*Aspergillus niger*	<S, D330, H363	
PiL_90616912_gb_EB050355	S28	MER093133	2.30E-92	family S28 unassigned peptidases	*Aspergillus niger*	<S, D252, H285	
CV0248	S33	MER033047	1.20E-42	family S33 non-peptidase homologues	*Magnaporthe grisea*	<S, G368D, T390H	
YeastLMW9_C07_51_050	S33	MER033052	7.50E-15	family S33 non-peptidase homologues	*Magnaporthe grisea*	<S, D207, H234	
PiL_90609682_gb_EB043125	S33	MER033314	6.50E-52	family S33 unassigned peptidases	*Neurospora crassa*	<S, D174, >H	
PiL_90617084_gb_EB050527	S33	MER033314	1.20E-75	family S33 unassigned peptidases	*Neurospora crassa*	S407, >D, >H	
ANU_610	S33	MER033425	6.70E-37	family S33 unassigned peptidases	*Neurospora crassa*	T268S, D310, H348	
CV0603	S33	MER033425	6.20E-45	family S33 unassigned peptidases	*Neurospora crassa*	T437S, D481, >H	
Clv_114220529_gb_EE727203	S33	MER033429	2.60E-33	family S33 unassigned peptidases	*Neurospora crassa*	<S, D377, H410	
Clv_114219839_gb_EE726513	S33	MER039326	3.60E-53	family S33 unassigned peptidases	*Emericella nidulans*	<S, D279, H308	
CV0129	S33	MER044624	1.50E-35	family S33 unassigned peptidases	*Schizosaccharomyces pombe*	<S, D235, H272	
OP0776	S33	MER044624	6.20E-50	family S33 unassigned peptidases	*Schizosaccharomyces pombe*	>S, >D, >H	
Clv_114223346_gb_EE730020	S33	MER064109	5.30E-30	family S33 unassigned peptidases	*Gibberella zeae*	S397, >D, >H	
Clv_114223438_gb_EE730112	S33	MER064746	5.70E-25	family S33 unassigned peptidases	*Gibberella zeae*	<S, D585, H613	
PiL_90608471_gb_EB041916	S33	MER064746	8.70E-69	family S33 unassigned peptidases	*Gibberella zeae*	S70, D235, H263	
ANU_486	S33	MER065263	9.10E-28	family S33 non-peptidase homologues	*Gibberella zeae*	<S, Q390D, H429	
LMW_47_G03_23_021	S33	MER088684	1.50E-27	family S33 unassigned peptidases	*Saccharopolyspora erythraea*	S83, >D, >H	
OP0867	S33	MER115440	3.60E-07	family S33 non-peptidase homologues	*Microcystis aeruginosa*	G1146S, >D, >H	

	Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
Threonine Peptidases	LMW_52_H04_32_032	S33	MER033047	1.70E-30	family S33 non-peptidase homologues	*Magnaporthe grisea*	<S, D365, A390H	
	ANU_260	S33	MER033052	5.60E-24	family S33 non-peptidase homologues	*Magnaporthe grisea*	<S, D393, H420	
	OP0482	S33	MER096447	5.10E-20	family S33 unassigned peptidases	*Kineococcus radiotolerans*	X228S, D355, >H	
	ANU_486	S33	MER076920	1.00E-04	family S33 non-peptidase homologues	*Acidovorax sp. JS42*	<S, >D, >H	
	Clv_114219417_gb_EE726091	S53	MER087653	2.70E-84	grifolisin	*Neosartorya fischeri*	<E, <D, D112, S230	
	YLMW24_F11_86_091	S53	MER049772	2.90E-36	aorsin	*Aspergillus fumigatus*		
	YeastLMW8_D03_20_026	S54	MER030047	7.00E-11	Rhomboid-7 peptidase	*Homo sapiens*	S352, H411	
	PiL_90611745_gb_EB045188	S54	MER033060	8.60E-59	family S54 non-peptidase homologues	*Magnaporthe grisea*	I105S, X163H	
	Clv_114221090_gb_EE727764	S54	MER033317	2.80E-83	family S54 unassigned peptidases	*Neurospora crassa*	S308, H361	
	LMW_47_A03_17_018	S54	MER090286	7.30E-15	family S54 unassigned peptidases	*Emericella nidulans*	S405, >H	
	PiL_90611350_gb_EB044793	S59	MER035056	2.60E-28	*Emericella nidulans*	family S59 non-peptidase homologues	H124, T126S	
	Clv_114223700_gb_EE730374	S66	MER079100	4.70E-06	family S66 unassigned peptidases	*Clostridium botulinum*	<S, <E, H440	
	CV1065	T01A	MER049775	1.30E-89	proteasome subunit beta 2	*Aspergillus fumigatus*		
	LMW_58_F_D07_52_058	T01A	MER064094	1.50E-69	proteasome catalytic subunit 3	*Gibberella zeae*	<T	
	CV0782	T01A	MER064166	8.10E-99	proteasome catalytic subunit 1	*Gibberella zeae*	T226	
	CV0252	T01A	MER083526	9.70E-84	proteasome catalytic subunit 2	*Chaetomium globosum*	<T	
	YeastLMW9_B07_50_057	T01A	MER083526	9.40E-21	proteasome catalytic subunit 2	*Chaetomium globosum*	<T	
	Clv_114222659_gb_EE729333	T01A	MER086130	3.50E-103	proteasome subunit alpha 7	*Aspergillus terreus*	T214	
	07_F11-A-SP6	T01A	MER086130	5.30E-102	proteasome subunit alpha 7	*Aspergillus terreus*	T202	
	PiL_90609540_gb_EB042985	T01A	MER087520	3.80E-60	proteasome subunit beta 3	*Aspergillus clavatus*	G338T	
	CV0079	T01A	MER087527	2.80E-83	proteasome subunit alpha 3	*Aspergillus clavatus*	G426T	
	Clv_114218930_gb_EE725604	T01A	MER087625	1.00E-85	proteasome subunit alpha 6	*Neosartorya fischeri*	M30T	
	CV0932	T01A	MER089994	3.80E-77	proteasome subunit alpha 1	*Emericella nidulans*	<T	
	Ylmw21_F04_30_041	T01A	MER091363	6.20E-58	proteasome subunit alpha 4	*Neosartorya fischeri*	<T	
	Clv_114222302_gb_EE728976	T01A	MER093154	1.80E-44	proteasome subunit alpha 2	*Aspergillus niger*	<T	
	Ylmw21_F04_30_041	T01A	MER091363	4.60E-18	proteasome subunit alpha 4	*Neosartorya fischeri*	<T	

Identifier	Family	MEROPS ID	*E*-value	Peptidase	Species	Active Site	Metal Ligands
LMW_30_A07_49_049	T01A	MER093154	4.70E-45	proteasome subunit alpha 2	*Aspergillus niger*	<T	
ANU_223	T01A	MER087575	5.20E-97	proteasome subunit beta 1	*Aspergillus clavatus*	S191T	
PiL_90616231_gb_EB049674	T02	MER064132	9.00E-55	isoaspartyl dipeptidase (threonine type)	*Gibberella zeae*	>T	
PiL_90616231_gb_EB049674	T02	MER064132	2.20E-17	isoaspartyl dipeptidase (threonine type)	*Gibberella zeae*	T743	
Clv_114219144_gb_EE725818	T03	MER091405	3.50E-08	gamma-glutamyltransferase 2 (bacterial)	*Aspergillus niger*	>S/T	
PiL_90608649_gb_EB042094	T03	MER107516	3.50E-81	gamma-glutamyltransferase 2 (bacterial)	*Magnaporthe grisea*	>S/T	
CV1075	T03	MER107518	2.10E-54	gamma-glutamyltransferase (*Schizosaccharomyces pombe*)	*Magnaporthe grisea*	T879	
PiL_90616005_gb_EB049448	T03	MER107518	8.60E-85	gamma-glutamyltransferase (*Schizosaccharomyces pombe*)	*Magnaporthe grisea*	T91	
PiL_90617043_gb_EB050486	T05	MER100905	3.30E-46	ornithine acetyltransferase precursor	*Magnaporthe grisea*	<T	

Appendix 4: Transcriptional Factors Predicted by TESS Analysis

5' Region of NADH

Motif	Factor	Model	Pos.	Length
CGATTA	Prd	I00252 (Prd)	29	6
ATTATT	HMG	J00134 (MF0011)	113	6
TTATCT	GATA-3	I00108 (GATA-3)	69	6
TTATAT	T00302 GAL4	R00497 ()	245	6
TTATCT	T00267 GATA-1 T00305 GATA-1 T00306 GATA-1 T01302 GATA-2 T02716 GATA-4	R02892 () R03206 () R03438 () R03884 () R08288 ()	69	6
TATAGTTT	FOXL1	J00033 (MA0033)	248	8
CTTTTG	HOXA5	I00184 (HOXA5)	73	6
CCTATATTG	T00378 HOXA3	M00395 (V$HOXA3_01)	213	9
ACTTATATAG	MADS	J00131 (MF0008)	243	10
TTTTTATCT	T00798 TBP	Q00171 (-)	66	9
GCTGT	GT-IIBa	I00160 (GT-IIBa)	78	5
TACTTATA	FOXC1	J00032 (MA0032)	242	8
TATCTTTTGCT	T02691 Dof3	M00354 (P$DOF3_01)	70	11
ATTGATTAT	SRY	J00084 (MA0084)	109	9
TAWWWWTA	T01009 RSRFC4	R03607 ()	64	8
TTTATTTG	HNF-5	I00100 (HNF-5)	50	8
TTTTT	T00395 Hb	Q00091 (-)	48	5
TATATAGTTTAT	T00997 T00996 SRY	M00160 (V$SRY_02)	246	12
TGGAAG	PEA3	I00212 (PEA3)	146	6
GTTAAAA	GT-1	I00347 (GT-1)	173	7
TATATTGC	FOXL1	J00033 (MA0033)	215	8
TACTTAT	T01675 Nkx2-5	M00240 (V$NKX25_01)	242	7
ACTAAATA	SGF-1	I00253 (SGF-1)	101	8
AATTATTT	HOXD10	I00179 (HOXD10)	43	8
TTTTTATT	SGF-1	I00253 (SGF-1)	48	8
AATTATT	Homeobox	J00133 (MF0010)	43	7
ATTGATTA	GHF-1	I00106 (GHF-1)	109	8
ATTAAAG	T01484 Cdx-1	M00100 (V$CDXA_01)	116	7
ATATTGCCTAAAGC	T00459 T00581 T00017 C/EBPbeta	M00109 (V$CEBPB_01)	216	14
CTATATGAGG	AGL3	J00001 (MA0001)	91	10
ATTTTTATCT	T00395 Hb Hunchback	J00049 (MA0049) M00022 (I$HB_01)	65	10
GTAAAACA	MNB1b	I00351 (MNB1b)	16	8
TGGAAG	SPI1	J00080 (MA0080)	146	6
CGTTGAATCT	RAR	I00040 (RAR)	128	10
TTTGGA	T00535 NF-1	Q00112 (-)	54	6
TTATTTTTAg	T00505 MEF-2 T01004 MEF-2 T01005 MEF-2 T01006 aMEF-2 T01784 MEF-2 (516 AA)	R00244 ()	45	10
GACTAAATAATT	T04166 T04167 T02290 FOXD3	M00130 (V$FOXD3_01)	100	12
GTAAAAAC	FOXD1	J00031 (MA0031)	81	8
CAATTATTTT	T01429 Sox-5	M00042 (V$SOX5_01)	42	10
AGTTTATAT	Forkhead	J00128 (MF0005)	183	9
TTATTAAA	T00691 Pit-1a	Q00144 (-)	114	8
TATTAA	TFIID	I00259 (TFIID)	115	6
ATAACACAATTATTTT	T01483 S8	M00099 (V$S8_01)	36	16
AAAGGC	Dof2	J00020 (MA0020)	119	6
TATATA	TFIID	I00259 (TFIID)	246	6
ATTATTTTT	SRY	J00084 (MA0084)	44	9
ATTGCCTAA	AP-3	I00271 (AP-3)	218	9
TGTAAAAA	SEF1	I00354 (SEF1)	80	8
GATTATATAACA	HLF	J00043 (MA0043)	30	12
TATCTT	T00627 NIT2	Q00133 (-)	70	6
TCTTTT	TCF-1	I00029 (TCF-1)	72	6
TGCTGTAAAA	T00487 MATalpha2	M00031 (F$MATALPHA2_01)	77	10
ACTAGTTTAT	T00138 T00139 T00137 c-Myb	M00183 (V$MYB_Q6)	180	10
CTAAATAAT	SRY	J00084 (MA0084)	102	9
ATTTTTAT	Oct-4	I00055 (Oct-4)	47	8
TGAATCTCCT	T01042 T01525 HSF1_(long) T01044 HSF1	M00146 (V$HSF1_01)	131	10
CGGGCTGG	LBP-1	I00280 (LBP-1)	9	8
ACTCCT	Sp1	I00295 (Sp1)	87	6
TAWWWWTA	T01009 RSRFC4	R03607 ()	46	8
AATTGATTA	Kruppel	I00249 (Kruppel)	108	9

Motif	Factor	Model	Pos.	Length
TATAACACAA	T00487 MATalpha2	M00031 (F$MATALPHA2_01)	35	10
TGGTATATG	T02072 Msx-1	M00394 (V$MSX1_01)	191	9
AATTATT	Homeobox	J00133 (MF0010)	43	7
ATTATATAA	T00798 TBP	Q00171 (-)	31	9
CACGTT	Arnt	J00004 (MA0004)	126	6
TTATCT	GATA3	J00037 (MA0037)	69	6
TATTTTT	T01484 Cdx-1	M00101 (V$CDXA_02)	46	7
TAATTGA	Nkx2-5	J00063 (MA0063)	107	7
TCCTATAT	Broad-complex_2	J00011 (MA0011)	89	8
CCCGGGCT	GCF	I00152 (GCF)	7	8
ACTAGTTTATA	Foxq1	J00040 (MA0040)	180	11
GCCTA	T-Ag	I00215 (T-Ag)	221	5
TTGAATC	DBP	I00131 (DBP)	130	7
GTTTATG	IHF	I00375 (IHF)	252	7
GGACTAA	c-Jun	I00010 (c-Jun)	99	7
ATATGGTA	Broad-complex_2	J00011 (MA0011)	188	8
ACTAAATA	FOXL1	J00033 (MA0033)	101	8
TTTATA	TFIID	I00217 (TFIID)	185	6
TTTATC	T00267 GATA-1 T00305 GATA-1 T00794 TBP T00821 TFIID	R01944 ()	68	6
TTTATATGG	opaque	I00359 (opaque)	185	9
TATGAGG	E12	I00274 (E12)	94	7
ACTAATT	Homeobox	J00133 (MF0010)	203	7
TATTTTTA	FOXL1	J00033 (MA0033)	64	8
ACAATTA	Nkx2-5	J00063 (MA0063)	41	7
TAAATAATTGATTATT	T01483 S8	M00099 (V$S8_01)	103	16
RArYMAAYAWTB	T01049 HNF-3B	R02177 ()	100	12
AAAGC	MNB1A	J00053 (MA0053)	225	5
ATTATTAAA	T01481 Pbx-1a	M00096 (V$PBX1_01)	113	9
TTCAATTAATT	T02983 Pax-4a	M00377 (V$PAX4_02)	230	11
ATTATATAAC	T00881 VBP	M00228 (V$VBP_01)	31	10
TTATTAAAG	T00798 TBP	Q00171 (-)	114	9
AAATAATTG	SOX9	J00077 (MA0077)	104	9
GTgGAATCTC	T00572 NF-GMa	R02683 ()	129	10
TTTTATCTT	T00311 GATA-3	M00077 (V$GATA3_01)	67	9
ATATAACACAATTA	T00104 C/EBPalpha T00105 T00107 T00108 T01388 C/EBP	M00190 (V$CEBP_Q2)	34	14
TAATNNNNCTTA	T03978 Cart-1	R09498 ()	236	12
CTTTAGG	T01484 Cdx-1	M00100 (V$CDXA_01)	139	7
CTGGTAA	IL-6.RE-BP	I00218 (IL-6.RE-BP)	13	7
AGTGTTA	Sox5	J00087 (MA0087)	170	7
CTCCTATAT	ILF	I00093 (ILF)	88	9
RTTTTTR	T01101 SEF4	R03650 ()	65	7
AATTAATT	HOXD10	I00179 (HOXD10)	233	8
CGATTATATAACAC	T00104 C/EBPalpha T00105 T00107 T00108 T01388 C/EBP	M00116 (V$CEBPA_01)	29	14
AACAAATTTCG	T02983 Pax-4a	M00377 (V$PAX4_02)	20	11
TAATT	Prrx2	J00075 (MA0075)	236	5
TGATTA	GATA3	J00037 (MA0037)	111	6
ATTA	Ubx	J00094 (MA0094)	31	4
TAACNNNATTA	T03978 Cart-1	R09500 () R09516 ()	37	11
TTTTTATTT	T00798 TBP	Q00171 (-)	48	9
TATATA	TFIID	I00259 (TFIID)	33	6
CAATTAT	Nkx2-5	J00063 (MA0063)	42	7
GTAAAAC	T00582 NF-IL-2A T00641 POU2F1 T00646 POU2F2 (Oct-2.1)	R00936 ()	16	7
CGGCCCGG	GCF	I00152 (GCF)	4	8
ATTATATAAC	T01071 Hlf	M00260 (V$HLF_01)	31	10
GCCTAAAGCTT	T01059 MNB1a	M00352 (P$DOF1_01)	221	11
TAATTG	Prd	I00252 (Prd)	107	6
CTCCTTTAGGT	T02690 Dof2	M00353 (P$DOF2_01)	136	11
AATTATTTTTAT	T04166 T04167 T02290 FOXD3	M00130 (V$FOXD3_01)	43	12
TCGATTATATA	T00794 TBP	Q00170 (-)	28	11
TATATGGTATATG	T00104 C/EBPalpha T00105 T00107 T00108 T01388 C/EBP	M00159 (V$CEBP_01)	187	13
GACTAATT	Broad-complex_2	J00011 (MA0011)	202	8
TAWWWWTA	T01009 RSRFC4	R03607 ()	64	8
ATTATATAAC	T01071 Hlf	M00260 (V$HLF_01)	31	10

Motif	Factor	Model	Pos.	Length
GGTAAAA	GT-1	I00347 (GT-1)	15	7
CGGCCCGcGC	T00320 GCF	R02641 ()	4	10
ATTATTT	Sox5	J00087 (MA0087)	44	7
TATTAAAGGCT	T01059 MNB1a	M00352 (P$DOF1_01)	115	11
TTAAAAACTAGTTTA	T01479 BR-C_Z3	M00093 (I$BRCZ3_01)	174	15
ATATAA	T00302 GAL4	R00497 ()	34	6
ATAACACA	AP-3	I00150 (AP-3)	36	8
TTTTACTT	Broad-complex_2	J00011 (MA0011)	239	8
TTTATA	T00794 TBP T00796 TBP T00798 TBP T00820 TFIID	R01014 () R02247 () R03172 () R03173 () R03840 () R08480 () R08513 ()	185	6
TaTTTATTTG	T04169 FOXJ2 (long isoform)	R09761 ()	48	10
ACTAAATAA	T01481 Pbx-1a	M00096 (V$PBX1_01)	101	9
TTATTTG	T00306 GATA-1	Q00069 (-)	51	7
TATTTTTAT	Forkhead	J00128 (MF0005)	46	9
TCTCCTTTAGG	Agamous	J00005 (MA0005)	135	11
GCCTAAAGCTT	T02690 Dof2	M00353 (P$DOF2_01)	221	11
ACCTATAT	Broad-complex_2	J00011 (MA0011)	212	8
TCCTTT	Dof2	J00020 (MA0020)	137	6
ATTATTAA	GHF-1	I00106 (GHF-1)	113	8
ATTATATAAC	T00881 VBP	M00228 (V$VBP_01)	31	10
TGATTA	T00267 GATA-1	Q00060 (-)	111	6
TCTTTTGCTG	CUP2	I00306 (CUP2)	72	10
TATCTTT	T00997 T00996 SRY	M00148 (V$SRY_01)	70	7
TATGGT	T02256 AML1a	M00271 (V$AML1_01)	189	6
TCCTATAT	T00337 GR alpha T01920 GR beta	R03555 ()	89	8
ATTATT	T00015 AFP1 T00048 ATBF1-B	R09386 ()	44	6
TTCAATT	EcR	I00238 (EcR)	230	7
TTGATTA	DBP	I00131 (DBP)	110	7
AAAACTAGTTTATAT	T01479 BR-C_Z3	M00093 (I$BRCZ3_01)	177	15
GCGGC	T00788 T-Ag	Q00168 (-)	162	5
GGCCCGGGC	TFAP2A	J00003 (MA0003)	5	9
CTAGTTT	T00997 T00996 SRY	M00148 (V$SRY_01)	181	7
AAACTAGTTT	T01806 T00671 p53	M00272 (V$P53_02)	178	10
AAAACAAATTT	Broad-complex_3	J00012 (MA0012)	18	11
CAATTAT	Dfd	I00236 (Dfd)	42	7
TTTTTATCTT	Hb	I00248 (Hb)	66	10
AATTAATT	SGF-2/3/4	I00254 (SGF-2/3/4)	233	8
AATTTTAC	Tll	I00260 (Tll)	206	8
TAAATA	TFIID	I00259 (TFIID)	103	6
AATAATT	Homeobox	J00133 (MF0010)	105	7
TTAATTT	Nkx2-5	J00063 (MA0063)	235	7
RTKAYgTAAY	T01428 E4BP4	R04100 ()	31	10
CAATTATTTT	Hb	I00248 (Hb)	42	10
TTGATTATTAAAGGCT	T00193 Dfd	M00019 (I$DFD_01)	110	16
ATAATTGATTATTAA	T00100 CUTL1	M00102 (V$CDP_02)	106	15
ATTGCCTAA	T00107 C/EBPalpha	Q00023 (-)	218	9
TTTACTTA	HNF-5	I00100 (HNF-5)	240	8
TAACACAAT	bZIP	J00129 (MF0006)	37	9
TTAAAAAC	SBF-1	I00353 (SBF-1)	174	8
AACACAATTATTTTTA	T00193 Dfd	M00019 (I$DFD_01)	38	16
AACAAAT	Sn	I00250 (Sn)	20	7
TATATGTAG	T00120 CF2-II	M00013 (I$CF2II_02)	194	9
TAATTGATT	SOX9	J00077 (MA0077)	107	9
TTATTTTTAa	T00505 MEF-2 T01005 MEF-2 T01006 aMEF-2 T01784 MEF-2 (516 AA)	R03584 ()	45	10
TATCT	GATA2	J00036 (MA0036)	70	5
AAACTAGTTTA	Broad-complex_3	J00012 (MA0012)	178	11
CTCCTTTAGGT	T02691 Dof3	M00354 (P$DOF3_01)	136	11
TAGGTGGAA	Macho-1	J00118 (MA0118)	142	9
ATTTTACTTATATAG	T00796 T00797 T00794 TBP	M00252 (V$TATA_01)	238	15
TATATGTAG	T00120 CF2-II	M00012 (I$CF2II_01)	194	9
ATAATTGATTATTAA	T01485 Clox	M00103 (V$CLOX_01)	106	15
ATATGG	T00865 YY1	R03000 ()	188	6
TTTATG	Cad	I00232 (Cad)	253	6
TATATA	TFIID	I00338 (TFIID)	246	6

Motif	Factor	Model	Pos.	Length
TTTACTTAT	Bapx1	J00122 (MA0122)	240	9
TAAAAAC	GT-1	I00347 (GT-1)	82	7
ACTCCTATATGAG	T01599 LCR-F1	M00285 (V$TCF11_01)	87	13
TAAT	Ubx	J00094 (MA0094)	205	4
TACTTATA	T00061 B factor	R00674 ()	242	8
TGATTATT	SGF-1	I00253 (SGF-1)	111	8
AGTTTATG	HOXD10	I00179 (HOXD10)	251	8
TTAAAAAC	Gt	I00246 (Gt)	174	8
TTATAT	T00302 GAL4	R00497 ()	32	6
TATATTGCCTAA	cEBP	J00102 (MA0102)	215	12
ATTTTAC	T04286 GT-1b	R09846 ()	207	7
AAAAA	T00395 Hb	Q00091 (-)	176	5
TTTTTATC	RUSH1-alfa	J00109 (MA0109)	66	8
CTCCTTTAG	ILF	I00093 (ILF)	136	9
TAATTGATTATTAA	Broad-complex_1	J00010 (MA0010)	107	14
TGTAAA	EFII	I00275 (EFII)	80	6
TGAGGA	ZNF42_1-4	J00056 (MA0056)	96	6
AATAAT	T00015 AFP1 T00048 ATBF1-B	R09386 ()	105	6
ATTATA	TFIID	I00217 (TFIID)	31	6
CAATTATTT	SOX9	J00077 (MA0077)	42	9
TAGGTG	T01467 deltaEF1	R08520 ()	142	6
GAATCTCC	HSF-2	I00211 (HSF-2)	132	8
TCACGTTG	NHP-1	I00285 (NHP-1)	125	8
AATTAATT	HOXD10	I00179 (HOXD10)	233	8
TTTTTA	TFIID	I00338 (TFIID)	48	6
TATATAAC	Elf-1/NTF-1	I00239 (Elf-1/NTF-1)	33	8
ATTTTTATCTTTT	T00304 T00305 GATA-1A T00306 T00267 GATA-1	M00128 (V$GATA1_04)	65	13
ATTATATAA	SRY	J00084 (MA0084)	31	9
TATTTTTAT	Forkhead	J00128 (MF0005)	64	9
CGATTATATAA	NFIL3	J00025 (MA0025)	29	11
AGGTGG	Sp1	I00295 (Sp1)	143	6
TTATCT	T00267 GATA-1	Q00060 (-)	69	6
ACCTATATTGCCTA	T01476 Abd-B	M00090 (I$ABDB_01)	212	14
TAACACA	T00997 T00996 SRY	M00148 (V$SRY_01)	37	7
CAATTAATT	T02072 Msx-1	M00394 (V$MSX1_01)	232	9
TTACCT	Nuclear	J00127 (MF0004)	210	6
TATGGTAT	FOXL1	J00033 (MA0033)	189	8
AGTGNNNNNNNNNAGT	T02325 ZPT2-1 T02447 ZPT2-2 T02448 EPF2-4	R05005 ()	170	16
ATATTGCCTAAAGC	T00459 T00581 T00017 C/EBPbeta	M00117 (V$CEBPB_02)	216	14
TATTTTTA	FOXL1	J00033 (MA0033)	46	8
GATTATATAACA	T01428 E4BP4	M00045 (V$E4BP4_01)	30	12
ATAATTG	T02016 En-1	M00396 (V$EN1_01)	106	7
ATAGTGTTAA	T01429 Sox-5	M00042 (V$SOX5_01)	168	10
TTTTATTT	Broad-complex_2	J00011 (MA0011)	49	8
TCAATTAATT	engrailed	I00240 (engrailed)	231	10
ATTA	Ubx	J00094 (MA0094)	158	4
TATATA	TFIID	I00338 (TFIID)	33	6
TATCTT	T00627 NIT2	M00142 (F$NIT2_01)	70	6
AGTGTT	HMG	J00134 (MF0011)	170	6
TATTTTT	T01484 Cdx-1	M00101 (V$CDXA_02)	64	7
CTATATGAGG	MADS	J00131 (MF0008)	91	10
TATTTTTATT	Hb	I00248 (Hb)	46	10
TAAAAA	TFIID	I00338 (TFIID)	175	6
AGTATAGTATT	Broad-complex_3	J00012 (MA0012)	150	11
TATAAC	TFIID	I00217 (TFIID)	35	6
AAAAA	T00395 Hb	Q00091 (-)	83	5
TTTTTATT	DEF	I00073 (DEF)	48	8
ATATAGTT	Broad-complex_2	J00011 (MA0011)	247	8
TTTTATTTGGAG	FOXI1	J00042 (MA0042)	49	12
GTATATGTA	T00120 CF2-II	M00012 (I$CF2II_01)	193	9
TATATA	TFIID	I00338 (TFIID)	33	6
ATTTTTATT	T01481 Pbx-1a	M00096 (V$PBX1_01)	47	9
TATAGTA	T00798 TBP	R03842 ()	152	7
TGAATCTCCT	Gfi	J00038 (MA0038)	131	10
TCAATTAATT	T00253 En	Q00056 (-)	231	10
ATTATTAA	SBF-1	I00353 (SBF-1)	113	8
GCTGG	LBP-1	I00191 (LBP-1)	12	5
TCGATTAT	Bcd	I00231 (Bcd)	28	8

Motif	Factor	Model	Pos.	Length
GGAAGTA	erg	I00003 (erg)	147	7
TTACTTA	DBP	I00131 (DBP)	241	7
TCACGTTGAA	T00487 MATalpha2	M00031 (F$MATALPHA2_01)	125	10
ATGGTATA	FOXL1	J00033 (MA0033)	190	8
TAGTGTTAAA	T00487 MATalpha2	M00031 (F$MATALPHA2_01)	169	10
TAWWWWTA	T01009 RSRFC4	R03607 ()	236	8
TAGTATT	T01484 Cdx-1	M00101 (V$CDXA_02)	154	7
ATAATTG	Nkx2-5	J00063 (MA0063)	106	7
CGGGCTGGTA	T00759 Sp1	M00008 (V$SP1_01)	9	10
TATAGTATTAGCGGCT	T00193 Dfd	M00019 (I$DFD_01)	152	16
TAAT	Ubx	J00094 (MA0094)	236	4
ATTATATA	T00759 Sp1 T00794 TBP T00798 TBP T00820 TFIID	R01732 () R01764 ()	31	8
CACAATTATTT	T02983 Pax-4a	M00377 (V$PAX4_02)	40	11
TCAATTA	Nkx2-5	J00063 (MA0063)	231	7
TAATTGATTATTAAAG	T00193 Dfd	M00019 (I$DFD_01)	107	16
CCTAAAG	T01484 Cdx-1	M00100 (V$CDXA_01)	222	7
CTATATGAGG	MADS	J00131 (MF0008)	91	10
AATTTTAC	Tll	I00260 (Tll)	237	8
CAATTAA	Nkx2-5	J00063 (MA0063)	232	7
CAATTAAT	T01675 Nkx2-5	M00241 (V$NKX25_02)	232	8
GACTAAATA	Forkhead	J00128 (MF0005)	100	9
AAAGGC	Dof3	J00021 (MA0021)	119	6
ATTGATT	Homeobox	J00133 (MF0010)	109	7
ATAACcCAAT	T00456 Kr	R02486 () R02797 () R02829 () R02882 () R02890 () R03443 ()	36	10
TTATCT	T00305 GATA-1 T00306 GATA-1 T00308 GATA-2 T00310 GATA-3 T00311 GATA-3	R03723 () R08243 () R08272 ()	69	6
CAATTA	Prd	I00252 (Prd)	42	6
ATTATTTT	Tll	I00260 (Tll)	44	8
TGGTAAAACAAA	ID1	J00120 (MA0120)	14	12
TTTTGCTG	Zen	I00262 (Zen)	74	8
TATAGTTTATG	Foxq1	J00040 (MA0040)	248	11
TTATTAAA	SEF1	I00354 (SEF1)	114	8
GCCTAAAGC	TFAP2A	J00003 (MA0003)	221	9
CTATAGTGT	SOX9	J00077 (MA0077)	166	9
TATCTTTT	FOXL1	J00033 (MA0033)	70	8
GATTATAT	Gt	I00246 (Gt)	30	8
ATTGATTATTAAAGGC	T01524 T00630 POU3F2	M00145 (V$BRN2_01)	109	16
CCTATATGAGG	Agamous	J00005 (MA0005)	90	11
CACGTTGA	ABF-2	I00389 (ABF-2)	126	8
ACTAAATAA	T00371 HNF-3	Q00084 (-)	101	9
TGATTA	GATA-1	I00109 (GATA-1)	111	6
GAGGAC	H-2RIIBP	I00178 (H-2RIIBP)	97	6
GTTTATA	GT-1	I00347 (GT-1)	184	7
TAAAACAAA	SRY	J00084 (MA0084)	17	9
TGGAGG	Sp1	I00295 (Sp1)	56	6
GTAGACTAATTTTACC	T00193 Dfd	M00019 (I$DFD_01)	199	16
AATAAT	HMG	J00134 (MF0011)	105	6
TGGAGG	T00011 ADR1	M00048 (F$ADR1_01)	56	6
TTATTTTTAT	Hb	I00248 (Hb)	45	10
AAACAAAT	fl-fII	I00222 (fl-fII)	19	8
ATTATTAA	Oct-4	I00055 (Oct-4)	113	8
GAGGAGTATT	MalT	I00377 (MalT)	58	10
GGAAGT	MAF	I00394 (MAF)	147	6
ATTTTAC	T04286 GT-1b	R09846 ()	238	7
CAATTA	Prd	I00252 (Prd)	232	6
TACTTATAT	T00120 CF2-II	M00012 (I$CF2II_01)	242	9
GTTTATG	T01484 Cdx-1	M00100 (V$CDXA_01)	252	7
TTTATATGG	T00798 TBP	Q00171 (-)	185	9
TTATTTTTAa	T00505 MEF-2 T01005 MEF-2 T01006 aMEF-2 T01772 D-MEF2 T01784 MEF-2 (516 AA)	R03587 () R09145 ()	45	10
TGAATC	GCN4	I00312 (GCN4)	131	6
TATTTTT	SEF4	I00355 (SEF4)	46	7
CACGTTGAA	CG-1	I00345 (CG-1)	126	9
TATTTTTATCT	T00304 GATA-3 T00305 GATA-1A T00306 GATA-2 T00307 NF-E1b T00308 T00309 T00310 T00311 T00312 T00313 T00314 T00567 T01302 T00267 GATA-1	M00203 (V$GATA_C)	64	11
ATTA	Ubx	J00094 (MA0094)	113	4
ATTAAA	TFIID	I00291 (TFIID)	116	6

Motif	Factor	Model	Pos.	Length
ATTAATTT	Athb-1	J00008 (MA0008)	234	8
TATTAAA	T01484 Cdx-1	M00101 (V$CDXA_02)	115	7
TATTTTTATTT	Broad-complex_4	J00013 (MA0013)	46	11
TCTCCTTT	PEB1	I00052 (PEB1)	135	8
TAATTGATTATTA	T00104 C/EBPalpha T00105 T00107 T00108 T01388 C/EBP	M00159 (V$CEBP_01)	107	13
TTATTTGGA	T00371 HNF-3	Q00084 (-)	51	9
TTATTTTTATTT	T04166 T04167 T02290 FOXD3	M00130 (V$FOXD3_01)	45	12
TAACACAATTAT	T00997 T00996 SRY	M00160 (V$SRY_02)	37	12
TTTATT	TFIID	I00217 (TFIID)	50	6
TCTCCTT	MNF1	I00352 (MNF1)	135	7
TTATATAAC	bZIP	J00129 (MF0006)	32	9
TTCAATTAATTTTACT	T00193 Dfd	M00019 (I$DFD_01)	230	16
TCCTTT	Dof3	J00021 (MA0021)	137	6
ATTAATTTT	60k-protein	I00343 (60k-protein)	234	9
AATTA	Prrx2	J00075 (MA0075)	43	5
ATTATT	T00015 AFP1 T00048 ATBF1-B	R09386 ()	113	6
AAACTAGTTT	T01806 T00671 p53	M00272 (V$P53_02)	178	10
GGGCTGGTAAAA	T02983 Pax-4a	M00378 (V$PAX4_03)	10	12
TTTATA	TFIID	I00259 (TFIID)	185	6
GTATATGTA	T00120 CF2-II	M00013 (I$CF2II_02)	193	9
AAAACAAA	HNF-5	I00100 (HNF-5)	18	8
CACAATTAT	Sox17	J00078 (MA0078)	40	9
TTATTAA	T03461 Crx	R09109 ()	114	7
CGATTAT	T02016 En-1	M00396 (V$EN1_01)	29	7
AGTATTA	Ftz.2	I00243 (Ftz.2)	155	7
TTTTTA	TFIID	I00217 (TFIID)	48	6
TAATNNATTA	T02967 Alx-4 T03999 Cart-1	R09590 ()	107	10
TTATAT	T00302 GAL4	R00497 ()	186	6
ATTTTTAT	Gt	I00246 (Gt)	47	8
TGAGGACT	AT-BP1/AT-BP2	I00141 (AT-BP1/AT-BP2)	96	8
ATTWNNATK	T00630 POU3F2 T01524 POU3F2 T01873 POU3F2	R04327 ()	47	9
AAACAAAT	Oct-2	I00056 (Oct-2)	19	8
GATTA	Prrx2	J00075 (MA0075)	30	5
TTTTTA	TFIID	I00338 (TFIID)	66	6
TCCTATATG	T02072 Msx-1	M00394 (V$MSX1_01)	89	9
TAATT	Prrx2	J00075 (MA0075)	205	5
TCTTT	MNB1A	J00053 (MA0053)	72	5
TTTATC	TFIID	I00291 (TFIID)	68	6
ATAACACAAT	T01071 Hlf	M00260 (V$HLF_01)	36	10
AAATAATTGAT	T00295 Ftz	Q00064 (-)	104	11
TTTATCTTTTGCT	T00305 GATA-1	Q00068 (-)	68	13
AATCTCC	T00997 T00996 SRY	M00148 (V$SRY_01)	133	7
TGGAAGTA	FOXC1	J00032 (MA0032)	146	8
TCAATTAA	Ftz.1	I00242 (Ftz.1)	231	8
ATTATTAAA	T01481 Pbx-1a	M00096 (V$PBX1_01)	113	9
CAATTAATTTT	Broad-complex_3	J00012 (MA0012)	232	11
TTGCTGTA	E74A	I00237 (E74A)	76	8
TTTATCTT	Eryf1	I00278 (Eryf1)	68	8
ATTAATTT	Tll	I00260 (Tll)	234	8
TTTATT	TFIID	I00291 (TFIID)	50	6
CTTATATAGT	T00796 TFIID T00820 T00821 T01159 T01175 T00794 TBP	M00216 (V$TATA_C)	244	10
TTACTT	T00851 T3R-beta1	R04771 ()	241	6
TTTTATCTTT	T00267 GATA-1	M00347 (V$GATA1_06)	67	10
TAWWWWTA	T01009 RSRFC4	R03607 ()	46	8
GTATAGTA	Broad-complex_2	J00011 (MA0011)	151	8
CTGTAAA	IL-6.RE-BP	I00218 (IL-6.RE-BP)	79	7
GTgGAATCTCC	T00590 NF-kappaB	R00605 ()	129	11
TTAATTTTACTT	T02099 Zen-2	Q00253 (-)	235	12
TTCAATTAA	T01481 Pbx-1a	M00096 (V$PBX1_01)	230	9
TAAAAA	TFIID	I00217 (TFIID)	82	6
ATTATT	HMG	J00134 (MF0011)	44	6
TTATATA	TBP	I00408 (TBP)	245	7
ATTAATT	Homeobox	J00133 (MF0010)	234	7
TTTTTATTT	Forkhead	J00128 (MF0005)	48	9
TTTTGCTGT	Eve	I00241 (Eve)	74	9
CTTATATAG	T00798 TBP	Q00171 (-)	244	9
ATTGAT	HMG	J00134 (MF0011)	109	6
TTTAGG	MBF-1	I00083 (MBF-1)	140	6
TTTTACC	GT-1	I00347 (GT-1)	208	7

Motif	Factor	Model	Pos.	Length
TAACACAATTA	En1	J00027 (MA0027)	37	11
TTTATCTT	RUSH1-alfa	J00109 (MA0109)	68	8
CAATTAAT	Athb-1	J00008 (MA0008)	232	8
GTTTATATGGT	T00339 GT-1	Q00078 (-)	184	11
ATATTGCCTAAAGC	T00104 C/EBPalpha T00105 T00107 T00108 T01388 C/EBP	M00190 (V$CEBP_Q2)	216	14
ATTWNNATK	T00630 POU3F2 T01524 POU3F2 T01873 POU3F2	R04327 ()	25	9
AAACAAATT	MYB.ph3	J00054 (MA0054)	19	9
GATTATATAACA	HLF	J00043 (MA0043)	30	12
GTTTATATGG	MADS	J00131 (MF0008)	184	10
TTTTTA	TFIID	I00217 (TFIID)	66	6
TATGGT	YY1	J00095 (MA0095)	189	6
CGATTAT	Nkx2-5	J00063 (MA0063)	29	7
TAATTTTACTT	Broad-complex_4	J00013 (MA0013)	236	11
TCGATTA	T00063 Bcd	Q00016 (-)	28	7
AAACTCC	T00997 T00996 SRY	M00148 (V$SRY_01)	85	7
TATAGTAT	FOXL1	J00033 (MA0033)	152	8
GTATATaTAG	T00505 MEF-2 T01005 MEF-2 T01006 aMEF-2 T01784 MEF-2 (516 AA)	R03589 ()	193	10
TATTTTT	SEF4	I00355 (SEF4)	64	7
TTTACTTATAT	Broad-complex_4	J00013 (MA0013)	240	11
ATTATATAA	bZIP	J00129 (MF0006)	31	9
TAACACAAT	SRY	J00084 (MA0084)	37	9
CAATTAAT	GHF-1	I00106 (GHF-1)	232	8
AACTAGTT	Broad-complex_2	J00011 (MA0011)	179	8
TGGAGGAGTATTTT	T01427 p300	M00033 (V$P300_01)	56	14
TAAAAA	TFIID	I00217 (TFIID)	175	6
CACGTT	deltaEF1	J00103 (MA0103)	126	6
ATTGCCTAA	bZIP	J00129 (MF0006)	218	9
GGAGGA	ZNF42_1-4	J00056 (MA0056)	57	6
TATTTTTAT	T00798 TBP	Q00171 (-)	64	9
AAATAATTGAT	T02983 Pax-4a	M00377 (V$PAX4_02)	104	11
TAAAAAC	GT-1	I00347 (GT-1)	175	7
ATTGCCTAA	NF-IL6	I00088 (NF-IL6)	218	9
TATTTTTAT	T00798 TBP	Q00171 (-)	46	9
TTTTGCTGT	EBP-40	I00121 (EBP-40)	74	9
CAATTAT	T02016 En-1	M00396 (V$EN1_01)	42	7
CTTTAGGTGGAA	T00625 AREB6	M00413 (V$AREB6_02)	139	12
AAATAATT	Broad-complex_2	J00011 (MA0011)	104	8
GTTGAATCT	Eve	I00241 (Eve)	129	9
TTTCGATTATATA	T00642 POU2F1a T00643 T00644 T00959 T01031 T01157 T01466 T00641 POU2F1	M00137 (V$OCT1_03)	26	13
TATTAAAGGCT	T02690 Dof2	M00353 (P$DOF2_01)	115	11
AATTATTT	Broad-complex_2	J00011 (MA0011)	43	8
CAATTATT	Athb-1	J00008 (MA0008)	42	8
TAAAACAAATTT	T00997 T00996 SRY	M00160 (V$SRY_02)	17	12
AGGCTCACG	T00117 CF1	M00111 (I$CF1_01)	121	9
AACACAATTATTT	T00642 POU2F1a T00643 T00644 T00959 T01031 T01157 T01466 T00641 POU2F1	M00137 (V$OCT1_03)	38	13
TTTTATCTTT	T00307 GATA-2	M00348 (V$GATA2_02)	67	10
ATTGCCTAAA	T00108 C/EBPalpha	Q00024 (-)	218	10
TAAAAA	TFIID	I00338 (TFIID)	82	6
TATATGGT	CF1	I00159 (CF1)	187	8
MATWAAT	T00630 POU3F2 T01524 POU3F2 T01873 POU3F2	R04327 ()	233	7
TCGATTAT	T00063 Bcd	M00140 (I$BCD_01)	28	8
AATTAAT	Homeobox	J00133 (MF0010)	233	7
TTATTTTTATTTG	T00642 POU2F1a T00643 T00644 T00959 T01031 T01157 T01466 T00641 POU2F1	M00137 (V$OCT1_03)	45	13
TCCTATATGAGGAC	T01476 Abd-B	M00090 (I$ABDB_01)	89	14
TCAATTAATT	Oct-2	I00251 (Oct-2)	231	10
TAATT	Prrx2	J00075 (MA0075)	107	5
GTTTATATGG	AGL3	J00001 (MA0001)	184	10
AGGAGT	Sp1	I00295 (Sp1)	59	6
TATCTTTTGCT	T02692 PBF	M00355 (P$PBF_01)	70	11
AAAGC	PBF	J00064 (MA0064)	225	5
GCGGC	T00788 T-Ag	Q00168 (-)	162	5
AAGCTTCAATTAATTT	T01483 S8	M00099 (V$S8_01)	226	16
ACTAGTTTATAT	T02288 HFH-1	M00129 (V$HFH1_01)	180	12
AATAATTG	Athb-1	J00008 (MA0008)	105	8

Motif	Factor	Model	Pos.	Length
GGTGG	CAC-binding	I00019 (CAC-binding)	144	5
TTCAAT	T00137 c-Myb	R04082 ()	230	6
TATCTTTTGCT	T02690 Dof2	M00353 (P$DOF2_01)	70	11
TCAATTAATT	AbdB-r	I00226 (AbdB-r)	231	10
TACCTA	deltaEF1	J00103 (MA0103)	211	6
TACTTAT	Nkx2-5	J00063 (MA0063)	242	7
ATATAACACAATTATT	T01019 Elf-1	M00110 (I$ELF1_01)	34	16
AACAAAT	T00997 SRY	R03581 ()	20	7
TATCTTTTGCT	T01059 MNB1a	M00352 (P$DOF1_01)	70	11
CAATTAA	T03461 Crx	R09105 ()	232	7
TgTTTATTTG	T04169 FOXJ2 (long isoform)	R09783 ()	48	10
GTGTTAAAA	T00798 TBP	Q00171 (-)	171	9
ATAATTGATTAT	T00997 T00996 SRY	M00160 (V$SRY_02)	106	12
GATTATATAACA	T01428 E4BP4	M00045 (V$E4BP4_01)	30	12
TTTATA	TFIID	I00338 (TFIID)	185	6
CCTATATGAGG	Agamous	J00005 (MA0005)	90	11
GGCTGGT	T00076 CAC-binding_protein	Q00018 (-)	11	7
ATAATTGAT	T01481 Pbx-1a	M00096 (V$PBX1_01)	106	9
TTGATTATT	T01481 Pbx-1a	M00096 (V$PBX1_01)	110	9
GTATATGTAG	CF2-II	J00015 (MA0015)	193	10
GTTTATAT	Gt	I00246 (Gt)	184	8
GACTAATTTTA	T02983 Pax-4a	M00377 (V$PAX4_02)	202	11
AGGACTA	Ftz.2	I00243 (Ftz.2)	98	7
AAATAATT	HOXD10	I00179 (HOXD10)	104	8
TATATA	TFIID	I00259 (TFIID)	246	6
GTATTTT	T02016 En-1	M00396 (V$EN1_01)	63	7
CAATTAA	T04297 Nkx6-1	R09835 () R09836 ()	232	7
TAATNNATTA	T02967 Alx-4 T03999 Cart-1	R09590 ()	107	10
TAAT	Ubx	J00094 (MA0094)	107	4
TAWWWWTA	T01009 RSRFC4	R03607 ()	205	8
ATTA	Ubx	J00094 (MA0094)	234	4
TTATCT	GATA-1	I00109 (GATA-1)	69	6
ATTTTA	T00691 POU1F1a T01505 F2F T01506 F2F	R01276 ()	238	6
ATTATTTTT	T01481 Pbx-1a	M00096 (V$PBX1_01)	44	9
ACACAATTATTTT	T00642 POU2F1a T00643 T00644 T00959 T01031 T01157 T01466 T00641 POU2F1	M00137 (V$OCT1_03)	39	13
TAAATAATTG	T01429 Sox-5	M00042 (V$SOX5_01)	103	10
TTAATTTTACTT	T00917 Zen-1	Q00193 (-)	235	12
TTATATA	TBP	I00408 (TBP)	32	7
TAAATAA	TBP	I00408 (TBP)	103	7
ACTAAATAATTGATTA	T00193 Dfd	M00019 (I$DFD_01)	101	16
YAAAAAY	T01101 SEF4	R03650 ()	82	7
ATTATATA	FOXL1	J00033 (MA0033)	31	8
TGATTATTAA	T01429 Sox-5	M00042 (V$SOX5_01)	111	10
ATTATTAA	T00015 AFP1	R00080 ()	113	8
TTTTATC	T04286 GT-1b	R09847 ()	67	7
AAACAAA	T00997 T00996 SRY	M00148 (V$SRY_01)	19	7
AATTGATTA	T00456 Kr	Q00099 (-)	108	9
ATATTGCCTAAAGC	T00104 C/EBPalpha T00105 T00107 T00108 T01388 C/EBP	M00116 (V$CEBPA_01)	216	14
TTATATAACAC	NFIL3	J00025 (MA0025)	32	11
TATGTAGA	FOXL1	J00033 (MA0033)	196	8
AAACAAA	T00371 HNF-3alpha T01049 HNF-3B T01050 HNF-3gamma	R05073 ()	19	7
ATATAGTT	TRP(MYB)	J00132 (MF0009)	247	8
CACAAT	HMG	J00134 (MF0011)	40	6
TAGGTG	deltaEF1	J00103 (MA0103)	142	6
GGGCTGGTAA	HAND1-TCF3	J00092 (MA0092)	10	10
TTTTATCTTT	T00314 GATA-3	M00350 (V$GATA3_02)	67	10
TTATTAA	Nkx2-5	J00063 (MA0063)	114	7
TTTATTT	T00798 TBP	R03841 ()	50	7
AACAAA	HMG	J00134 (MF0011)	20	6
AATTAATT	SGF-2/3/4	I00254 (SGF-2/3/4)	233	8
ATTTTTATTT	T00395 Hb Hunchback	J00049 (MA0049) M00022 (I$HB_01)	47	10
ATTGCNNAA	T00017 C/EBPbeta T00108 C/EBPalpha T00216 C/EBPgamma T00459 C/EBPbeta T01386 C/EBP T01420 C/EBPbeta(p20) T02105 C/EBPbeta(p34) T02528 EFII	R01344 ()	218	9
AGGACT	PR	I00288 (PR)	98	6
TTTTATTTGG	MADS	J00131 (MF0008)	49	10

Motif	Factor	Model	Pos.	Length
ATTA	Ubx	J00094 (MA0094)	44	4
TATATAAC	FOXL1	J00033 (MA0033)	33	8
GATTA	Prrx2	J00075 (MA0075)	112	5
GTATATGTAG	CF2-II	J00015 (MA0015)	193	10
TAATNNNNNATTA	T02967 Alx-4 T03999 Cart-1 T04013 DUX1	R09593 () R09709 ()	107	13
AAACTAG	T00997 T00996 SRY	M00148 (V$SRY_01)	178	7
ATTA	Ubx	J00094 (MA0094)	116	4
TaATTATTAAA	T00026 Antp T00377 HOXA5 T00917 Zen-1 T02090 Phox2a T02099 Zen-2 T02403 Barx1	R01678 ()	111	11
TTTAGGTGGAA	T01467 deltaEF1	M00073 (V$DELTAEF1_01)	140	11
RTTTTTR	T01101 SEF4	R03650 ()	47	7
GTAAAAAC	Gt	I00246 (Gt)	81	8
ATTTTTAT	Gt	I00246 (Gt)	65	8
TGGAAG	c-ETS	J00098 (MA0098)	146	6
RTTAcRTMAY	T01428 E4BP4	R04100 ()	31	10
GGTAAAACAAATT	T00104 C/EBPalpha T00105 T00107 T00108 T01388 C/EBP	M00159 (V$CEBP_01)	15	13
ATTAATTT	GHF-1	I00106 (GHF-1)	234	8
AAAACTAGTTT	Broad-complex_3	J00012 (MA0012)	177	11
ATTGCCTAA	C/EBP	I00272 (C/EBP)	218	9
TTTTATCTTT	T00267 GATA-1	M00346 (V$GATA1_05)	67	10
ATTTTACTTATATAG	TBP	J00108 (MA0108)	238	15
TATGGTAT	AP-3	I00150 (AP-3)	189	8
YAAAAAY	T01101 SEF4	R03650 ()	175	7
GCCTAAAGCTT	T02691 Dof3	M00354 (P$DOF3_01)	221	11
TTTTATCTTT	T00307 GATA-2	M00349 (V$GATA2_03)	67	10
TCAATTAAaT	T00253 En T00295 Ftz T00646 POU2F2 (Oct-2.1) T00699 Prd T00917 Zen-1 T00919 Zfh-1 T00920 Zfh-2 T01476 Abd-B T01897	R00414 () R01669 () R01670 () R02507 ()	231	10
TAWWWWTA	T01009 RSRFC4	R03607 ()	236	8
AGGAGT	T00011 ADR1	M00048 (F$ADR1_01)	59	6
TTATTTTTtT	T00395 Hb	R02500 ()	45	10
ATTATTAA	T00691 Pit-1a	Q00144 (-)	113	8
TAAATA	TFIID	I00338 (TFIID)	103	6
ATAATTG	Dfd	I00236 (Dfd)	106	7
TATATA	TFIID	I00259 (TFIID)	33	6
TATTAAAGGCT	T02691 Dof3	M00354 (P$DOF3_01)	115	11
ATTTTA	T00691 POU1F1a T01505 F2F T01506 F2F	R01276 ()	207	6
CCTAAA	MBF-1	I00083 (MBF-1)	222	6
YTATTTWWAr	T00505 MEF-2 T01784 MEF-2 (516 AA)	R02201 ()	45	10
TATATAGTTTA	Broad-complex_3	J00012 (MA0012)	246	11
ATTATTA	Sox5	J00087 (MA0087)	113	7
ACTCCT	T00011 ADR1	M00048 (F$ADR1_01)	87	6
ATTWATK	T00630 POU3F2 T01524 POU3F2 T01873 POU3F2	R04327 ()	234	7
GCTTCAATTAATTTTA	T00193 Dfd	M00019 (I$DFD_01)	228	16
TAWWWWTA	T01009 RSRFC4	R03607 ()	205	8
GATTATT	Homeobox	J00133 (MF0010)	112	7
TTTTTATTTG	T00395 Hb Hunchback	J00049 (MA0049) M00022 (I$HB_01)	48	10
AAAGCT	Dof2	J00020 (MA0020)	225	6
TTTTATCTTT	T00314 GATA-3	M00351 (V$GATA3_03)	67	10
AACTAGTT	Broad-complex_2	J00011 (MA0011)	179	8
TTTATATGGT	T00796 TFIID T00820 T00821 T01159 T01175 T00794 TBP	M00216 (V$TATA_C)	185	10
AAAACAAATa	T04169 FOXJ2 (long isoform)	R09773 ()	18	10
ACTTATA	B-factor	I00229 (B-factor)	243	7
AGTGTTAA	AP-3	I00150 (AP-3)	170	8
TAAATA	TFIID	I00338 (TFIID)	103	6
TAATNNNNNATTA	T02967 Alx-4 T03999 Cart-1 T04013 DUX1	R09593 () R09709 ()	107	13
GTTTATA	T01484 Cdx-1	M00100 (V$CDXA_01)	184	7
TAGTGTT	Ftz.2	I00243 (Ftz.2)	169	7
AGACTAATTTT	Broad-complex_3	J00012 (MA0012)	201	11
ATTGATTA	Athb-1	J00008 (MA0008)	109	8
TTTTTATTTG	Hb	I00248 (Hb)	48	10
TTCAATTAATT	T00295 Ftz	Q00064 (-)	230	11
AAATAAT	Sox5	J00087 (MA0087)	104	7
TTGATTAT	T00063 Bcd	M00140 (I$BCD_01)	110	8

Motif	Factor	Model	Pos.	Length
TGTAGACT	PHO2	I00324 (PHO2)	198	8
TTTCG	T00386 HSTF	M00028 (I$HSF_01)	26	5
GTTAAA	TFIID	I00291 (TFIID)	173	6
TATATAGT	FOXL1	J00033 (MA0033)	246	8
GAGGAGTA	FOXC1	J00032 (MA0032)	58	8
AATAATT	Homeobox	J00133 (MF0010)	105	7
TTTTACTTA	Vmw65	I00017 (Vmw65)	239	9
ATTTTTAT	Oct-4	I00055 (Oct-4)	65	8
TATATA	TFIID	I00338 (TFIID)	246	6
TGATTA	T00305 GATA-1 T00306 GATA-1	R03442 ()	111	6
TTTTACCTATA	T01467 deltaEF1	M00073 (V$DELTAEF1_01)	208	11
AATTGATTATTAA	T00642 POU2F1a T00643 T00644 T00959 T01031 T01157 T01466 T00641 POU2F1	M00137 (V$OCT1_03)	108	13
AAAAACTCCTA	En1	J00027 (MA0027)	83	11
ATTATTAAA	SRY	J00084 (MA0084)	113	9
ATCTTT	Dof2	J00020 (MA0020)	71	6
CGTTGAA	EcR	I00238 (EcR)	128	7
CTCCTTTAGGT	T01059 MNB1a	M00352 (P$DOF1_01)	136	11
TACCTATA	FOXL1	J00033 (MA0033)	211	8
TTATCTTTTGCT	ID1	J00120 (MA0120)	69	12
TATATAA	TBP	I00408 (TBP)	33	7
CGTTGAATCT	T01043 T01045 T00972 HSF2	M00147 (V$HSF2_01)	128	10
TTATCTT	T00306 GATA-1	Q00069 (-)	69	7
TCAATTAAT	T01481 Pbx-1a	M00096 (V$PBX1_01)	231	9

5’ Region of HSP

Motif	Factor	Model	Pos.	Length
GCaGGAAGTG	T00111 c-Ets-1 T00112 c-Ets-1 T00113 c-Ets-2 T00250 Elk-1 T00684 PEA3 T00685 PEA3 T01400 Ets-1 deltaVII T01409 p38erg T01413 Net T02129 p55erg T02130 p49erg	R01224 ()	1677	10
AGGACA	T00698 PR B T01660 PR A	R02996 ()	143	6
AAAGT	MNB1A	J00053 (MA0053)	1921	5
CGTNACG	T00223 E4F1	R00363 ()	375	7
AGGTCA	RAR-beta	I00402 (RAR-beta)	528	6
GGTCAT	RC2	I00329 (RC2)	1076	6
CTTGGC	T00601 NF-1 (-like proteins)	R01192 ()	1427	6
CGAAG	T00386 HSTF	M00028 (I$HSF_01)	1606	5
AAACAAA	T00997 T00996 SRY	M00148 (V$SRY_01)	705	7
TTCAAA	T01159 TFIID	R02749 ()	1964	6
AAAAA	T00395 Hb	Q00091 (-)	1420	5
AAACAAT	Sox5	J00087 (MA0087)	1967	7
AAAAA	T00395 Hb	Q00091 (-)	1880	5
AAACAAA	T00371 HNF-3alpha T01049 HNF-3B T01050 HNF-3gamma	R05073 ()	705	7
CTCTCTCTCTCTC	T00301 GAGA_factor	Q00066 (-)	1712	13
TATATA	TFIID	I00259 (TFIID)	416	6
AAAAAAAAAA	Hb	I00248 (Hb)	1851	10
CCAAT	CP1	I00066 (CP1)	1822	5
AAAGT	PBF	J00064 (MA0064)	1921	5
GATTTTAT	Gt	I00246 (Gt)	583	8
CAACTG	T00137 c-Myb	R04183 ()	1544	6
ATGAAAA	T00691 POU1F1a	R02792 ()	601	7
CAAaCAAACA	T04169 FOXJ2 (long isoform)	R09747 () R09756 ()	1901	10
TACAAATAA	T00371 HNF-3	Q00084 (-)	698	9
AAAGTC	Dof3	J00021 (MA0021)	1921	6
AATAAT	T00015 AFP1 T00048 ATBF1-B	R09386 ()	1759	6
CAGCTG	T00036 AP-4	Q00009 (-)	1675	6
AAAAA	T00395 Hb	Q00091 (-)	1867	5
GGGTCA	T00029 AP-1	R03036 ()	1075	6
TTaTTTTTTT	T00395 Hb	R02500 ()	1738	10
AAAAA	T00395 Hb	Q00091 (-)	1862	5
TGTGCCA	Zta	I00413 (Zta)	99	7
CCTGC	LVc	I00085 (LVc)	950	5
GACCGCA	PEBP2	I00051 (PEBP2)	730	7
CATGTG	muEBP-C2	I00005 (muEBP-C2)	104	6
GTGACGGA	X2BP	I00015 (X2BP)	376	8
AAAAA	T00395 Hb	Q00091 (-)	1874	5
TAAT	Ubx	J00094 (MA0094)	1761	4
TTCCTT	T00113 c-Ets-2	R04339 ()	483	6
TGCAAA	EFII	I00275 (EFII)	651	6
CTTGGC	NF1	I00296 (NF1)	1173	6
TTTGTT	HMG	J00134 (MF0011)	1702	6

Motif	Factor	Model	Pos.	Length
CATTCC	T01085 abaA	R03759 () R03760 () R03761 () R03765 () R03766 () R03768 () R03769 () R03770 () R03771 () R03772 () R03777 () R03784 ()	1952	6
TTTTGG	MBF-1	I00083 (MBF-1)	691	6
AAAAA	T00395 Hb	Q00091 (-)	1878	5
RTCRYNNNNNACG	T00778 TAF	R02579 ()	1523	13
TATgATTTTG	T00108 C/EBPalpha T00216 C/EBPgamma T00459 C/EBPbeta T01420 C/EBPbeta(p20)	R00089 ()	686	10
AcATAAAAAA	T00395 Hb	R02554 ()	1843	10
CCTGC	LVc	I00085 (LVc)	986	5
AAACAAC	T00997 T00996 SRY	M00148 (V$SRY_01)	1928	7
CGGAGT	T00011 ADR1	M00048 (F$ADR1_01)	1247	6
CTCTCT	TFII-I	I00216 (TFII-I)	1501	6
ATCCAGTG	T00150 CP1 T00306 GATA-1 T00641 POU2F1	R04200 ()	346	8
CCTGC	LVc	I00085 (LVc)	1154	5
ATTGGC	T00535 NF-1	Q00112 (-)	1373	6
TGACCT	T00721 RAR-beta	Q00153 (-)	932	6
aAAACAAATA	T04169 FOXJ2 (long isoform)	R09773 ()	704	10
GTGACGtCAC	T00133 c-Jun T00163 CREB T00167 CRE-BP1 T00668 Opaque-2 T00829 TGA1a T00830 TGA1b T01090 TAF-1 T01092 CPRF-2 T01093 CPRF-3 T01380 CREB T01381 deltaCREB T02786 RITA-1	R00455 () R08318 () R08692 ()	1087	10
CAGTTG	T00137 c-Myb T00138 c-Myb T00140 c-Myc	R02392 () R02893 () R04182 ()	1054	6
AAAAA	T00395 Hb	Q00091 (-)	1872	5
GCGGANNNNNNNNNtCCGC	T02848 CHA4	R03817 ()	1000	19
GCAGCCTT	V$CAP_01	M00253 (V$CAP_01)	1783	8
CCAAT	CP2	I00132 (CP2)	1219	5
GAGGC	T-Ag	I00215 (T-Ag)	1794	5
ATTATATT	TII	I00260 (TII)	471	8
GCANNNNNNNNGGC	T02841 FACB	R08607 ()	652	14
TGATGTGAAT	T00488 MATa1	M00030 (F$MATA1_01)	1685	10
TTCAAAG	T00997 SRY T00999 TCF-1A T02918 TCF-4 T02945 TCF-1B	R04297 () R08622 () R08641 ()	1918	7
AAAAAAAAAA	Hb	I00248 (Hb)	1849	10
TTaTTTTTTT	T00395 Hb	R02500 ()	1734	10
TAAACAAATA	T04169 FOXJ2 (long isoform)	R09748 () R09762 ()	704	10
CCACC	CAC-binding	I00019 (CAC-binding)	1563	5
gAAATAAACA	T04169 FOXJ2 (long isoform)	R09776 ()	700	10
CTAATTAG	T00253 En	R01783 ()	1802	8
GGTTAG	T00368 HNF-1A T00369 HNF-1 T01950 HNF-1B T01951 HNF-1C	R00077 ()	1209	6
AGATA	GATA2	J00036 (MA0036)	639	5
CGGNNNNNNGCG	T02846 UAY	R08872 ()	380	12
CCTGG	LBP-1	I00191 (LBP-1)	761	5
TCGNNNNNNNNNNGGA	T02841 FACB	R08595 () R08599 ()	769	16
TGTGTGTA	FOXC1	J00032 (MA0032)	1353	8
GCTTCAGTTt	T00428 ISGF-3	R02188 ()	1385	10
TTATAGCTTt	T00930 LEF-1 T02905 LEF-1	R08634 ()	1433	10
GCCACGGGC	TFAP2A	J00003 (MA0003)	939	9
AAAAA	T00395 Hb	Q00091 (-)	1858	5
AAAAAAAAAA	Hb	I00248 (Hb)	1861	10
TATAAA	TFIID	I00217 (TFIID)	1844	6
GCCTGGCC	T00035 AP-2alphaA T02466 AP-2alphaB	R04379 ()	203	8
GCCGC	T00788 T-Ag	Q00168 (-)	798	5
ATTGCATCA	bZIP	J00129 (MF0006)	1472	9
AAAAA	T00395 Hb	Q00091 (-)	1882	5
CCAAT	CBP/CRF	I00135 (CBP/CRF)	1822	5
CATTCCT	MCBF	I00282 (MCBF)	1952	7
GGTCTT	RC2	I00329 (RC2)	810	6
CTTTGTT	T01109 TCF-1(P) T02905 LEF-1	R08814 () R08819 ()	1701	7
GGGCGT	T00752 Sp1	R01702 ()	1160	6
AGAACAG	T00336 GR	R01659 ()	230	7
AAAGCC	Dof3	J00021 (MA0021)	876	6
TTTCCA	c-ETS	J00098 (MA0098)	1819	6
TATAAAA	TBP	I00408 (TBP)	1844	7
TAAAAAA	T00794 TBP	R04010 ()	1418	7
CCACT	CP2	I00132 (CP2)	1394	5
ATTAGCAT	T00644 POU2F1a T00651 POU5F1 T01862 POU2F1b T01863 POU2F1c	R00305 () R00306 ()	1805	8
TGACCGCA	T00683 PEBP2 T01062 PEBP2alphaA1 T01063	R01222 ()	729	8
GGCTGGG	T00759 Sp1	R04998 ()	187	7

Motif	Factor	Model	Pos.	Length
GAGGC	T00788 T-Ag	Q00168 (-)	1697	5
CCAAC	CP2	I00132 (CP2)	1980	5
AGAAG	T00385 HSF1	M00029 (F$HSF_01)	392	5
CCAAT	CBP/CRF	I00135 (CBP/CRF)	1585	5
tGAGGTCAGGG	T00051 ATF T00163 CREB T01332 RXR-beta	R01053 ()	526	11
TATCGT	T00627 NIT2	R02729 ()	1729	6
AAAAA	T00395 Hb	Q00091 (-)	1853	5
CCCCC	CAC-binding	I00019 (CAC-binding)	1977	5
YTWWaAATAR	T00505 MEF-2 T01784 MEF-2 (516 AA)	R02201 ()	1754	10
TGGGGG	T00011 ADR1	M00048 (F$ADR1_01)	853	6
CTTCTC	T00702 PU.1	R08195 ()	1579	6
TTTCCT	c-ETS	J00098 (MA0098)	482	6
TGACGCA	c-Jun	I00010 (c-Jun)	158	7
AAAAA	T00395 Hb	Q00091 (-)	1847	5
GAGGGgTGGT	T00759 Sp1	R02104 ()	1273	10
GCANNNNNNNNNGGC	T02841 FACB	R08592 () R08608 ()	69	15
ATTGGTT	T00154 CP1A T01804 NF-YA	R00988 ()	517	7
TATAAA	TFIID	I00259 (TFIID)	1844	6
CTCTCTCTCTCTC	T00301 GAGA_factor	Q00066 (-)	1716	13
AGAAT	T00385 HSF1	M00029 (F$HSF_01)	1889	5
AAAAAAAAAA	Hb	I00248 (Hb)	1874	10
ATAATGC	HOXD8	I00182 (HOXD8)	1760	7
CAAATAAA	HNF-5	I00100 (HNF-5)	700	8
GGAGCC	T01944 NF-ATp T01948 NF-ATp	R05082 ()	423	6
CTCTCTCTCTCTc	T00301 GAGA factor	R02059 ()	1718	13
ATTGTT	HMG	J00134 (MF0011)	449	6
GAACAG	T00475 LVa	R01134 () R01135 ()	231	6
AAAAGG	TCF-1	I00029 (TCF-1)	604	6
TCAgCGCGGG	T00036 AP-4	R01038 ()	1078	10
TATATAAA	T00798 TBP	R03162 ()	1842	8
GAGGC	T00788 T-Ag	Q00168 (-)	1794	5
GRGRTTKCaY	T00572 NF-GMa T01217 NF-GMa	R02214 ()	855	10
TGTGGT	T02256 AML1a	M00271 (V$AML1_01)	666	6
GCCGACGG	MafB	J00117 (MA0117)	1038	8
GGGAGT	Sp1	I00295 (Sp1)	755	6
CAtAATGTCT	T04169 FOXJ2 (long isoform)	R09792 ()	1459	10
TTGGCA	T00599 NF-1/L	R00078 ()	1264	6
AAAAAAAAAA	Hb	I00248 (Hb)	1859	10
CTCTCT	TFII-I	I00216 (TFII-I)	1716	6
CCTGG	LBP-1	I00191 (LBP-1)	213	5
TGTGTGcACA	T01388 C/EBP	R04485 ()	1353	10
CCTGG	LBP-1	I00191 (LBP-1)	198	5
AAcAACACTT	T04169 FOXJ2 (long isoform)	R09789 ()	1420	10
CGGAGT	T00011 ADR1	M00048 (F$ADR1_01)	380	6
GTTGGC	T00535 NF-1	Q00112 (-)	870	6
GCTGG	LBP-1	I00191 (LBP-1)	1677	5
GGCAGT	MAF	I00394 (MAF)	1266	6
ATATATAAA	T00798 TBP	Q00171 (-)	1841	9
ACAACA	T00335 GR	R01313 ()	1930	6
AAAAAAAAAA	Hb	I00248 (Hb)	1850	10
AGAAG	T00385 HSF1	M00029 (F$HSF_01)	2	5
AAAAA	T00395 Hb	Q00091 (-)	1879	5
CTTGGC	NF1	I00296 (NF1)	1263	6
TGTGAT	T00333 GR	R01814 ()	1138	6
GGATGC	Gata1	J00035 (MA0035)	1104	6
ACCNNNNNNGGT	T00205 E2	R00177 () R00179 () R00184 ()	1014	12
TGATA	GATA2	J00036 (MA0036)	1839	5
CTCTCT	TFII-I	I00216 (TFII-I)	1718	6
ATCAGCAA	E74A	I00237 (E74A)	28	8
AGAACA	T00335 GR	Q00076 (-)	230	6
AAAGCC	Dof2	J00020 (MA0020)	876	6
GGTGCAA	DBP	I00131 (DBP)	564	7
CTCTCT	TFII-I	I00216 (TFII-I)	1722	6
GCTGG	LBP-1	I00191 (LBP-1)	1277	5
gCYGYYGYCGY	T00008 Adf-1	R04245 ()	1517	11
TGTGCC	T00333 GR	R03535 ()	99	6
TGTTCT	GR/PR	I00104 (GR/PR)	1289	6
GCGGC	T00788 T-Ag	Q00168 (-)	661	5
TCNTACTC	T00104 C/EBPalpha T00105 C/EBPalpha T00107 C/EBPalpha T00108 C/EBPalpha T01388 C/EBP	R04246 ()	1147	8
CACCCG	T04363 DPBF-1 T04364 DPBF-2	R09888 ()	914	6
CAGAGCG	GAGA	I00244 (GAGA)	1601	7
TAAAAAA	T00794 TBP	R04010 ()	1846	7
TCTATC	T00305 GATA-1	R04289 ()	1727	6
CAAATAAAtA	T04169 FOXJ2 (long isoform)	R09761 ()	700	10
ATTGGCT	T00150 CP1 T00174 CTF T00536 NF-1	R01443 () R03038 ()	1373	7
CACGCG	T04363 DPBF-1 T04364 DPBF-2	R09890 ()	92	6
TTCCTCC	Elf-1	I00113 (Elf-1)	673	7
GGGARAWMCm	T00196 Dl	R03463 ()	191	10
CATTCCT	T00499 MCBF T00507 MF3	R02075 () R02200 ()	1952	7

Motif	Factor	Model	Pos.	Length
AAAGTC	Dof3	J00021 (MA0021)	924	6
AGGAGT	T00011 ADR1	M00048 (F$ADR1_01)	272	6
CCTGG	LBP-1	I00191 (LBP-1)	956	5
AGAAG	T00385 HSF1	M00029 (F$HSF_01)	356	5
TAGCTcCACAC	T02983 Pax-4a	R08733 ()	1436	11
AGAAC	T00385 HSF1	M00029 (F$HSF_01)	194	5
CCTGC	LVc	I00085 (LVc)	180	5
TTaTTTTTTT	T00395 Hb	R02500 ()	1737	10
GGGGTCATC	T00117 CF1	M00111 (I$CF1_01)	1074	9
AGATTTTG	ARR10	J00121 (MA0121)	5	8
YRYTGCATaYYY	T00774 su(Hw)	R02385 ()	1395	12
AGGTCAGG	COUP	I00133 (COUP)	528	8
CTTTGTTT	fl-fII	I00222 (fl-fII)	1701	8
TTCCTC	PU.1	I00047 (PU.1)	673	6
WNNANATAAAYA	T02475 FOXL1	R05069 ()	698	12
GCANNNNNNNNGGC	T02841 FACB	R08607 ()	162	14
TTCCTC	T00702 PU.1	R04413 ()	1954	6
GAGGC	T-Ag	I00215 (T-Ag)	295	5
ATATCG	T00305 GATA-1	R08194 ()	784	6
TGACCT	T00264 ER-alpha T04014 SF-1	R03566 () R03567 () R03569 () R09684 () R09686 ()	932	6
GACGC	T00788 T-Ag	Q00168 (-)	159	5
AAAAA	T00395 Hb	Q00091 (-)	1863	5
AAAAAAAAAA	Hb	I00248 (Hb)	1868	10
GGGGGT	T00011 ADR1	M00048 (F$ADR1_01)	854	6
TgTTTGGTAT	T02880 MYB1	R08528 ()	461	10
AGTTCT	T00335 GR	R01313 ()	359	6
TATCGT	T00627 NIT2	R02729 ()	821	6
CTCTCT	TFII-I	I00216 (TFII-I)	1957	6
AAAAAAAAAA	Hb	I00248 (Hb)	1853	10
CAGCTG	RFX2	I00406 (RFX2)	1675	6
TTGAGCG	T02100 Zeste	Q00254 (-)	657	7
CAAAcAAACA	T04169 FOXJ2 (long isoform)	R09747 () R09756 ()	700	10
CTCTCT	TFII-I	I00216 (TFII-I)	1710	6
CCCCC	CAC-binding	I00019 (CAC-binding)	1648	5
AAAAA	T00395 Hb	Q00091 (-)	1877	5
TATTTGGTATATT	T00104 C/EBPalpha T00105 T00107 T00108 T01388 C/EBP	M00159 (V$CEBP_01)	461	13
ACAACA	T00335 GR	R01313 ()	1986	6
GCCAcGTGGC	T00668 Opaque-2 T00829 TGA1a T00830 TGA1b T00937 HBP-1a T01090 TAF-1 T01091 CPRF-1 T01092 CPRF-2 T01093 CPRF-3 T02669 EmBP-1a T02672 GBF1 T02674 GBF2A T02786 RITA-1 T02999 OCSBF-1	R03654 () R08310 () R08659 () R08680 () R08691 () R09851 ()	102	10
CTGTCA	T03388 Meis-1a T03389 Meis-1b T04076 TGIF T04115 Meis-2a T04116 Meis-2b T04117 Meis-2c T04118 Meis-2d	R09633 () R09647 () R09648 ()	863	6
CAAAGC	TCF-1	I00029 (TCF-1)	875	6
SMaTAAAAAA	T00395 Hb	R02606 ()	1415	10
CGAAG	T00386 HSTF	M00028 (I$HSF_01)	36	5
ACGCCC	T00759 Sp1	R01498 ()	1595	6
GGTTAT	RC2	I00329 (RC2)	520	6
AAAAA	T00395 Hb	Q00091 (-)	1871	5
GGATG	GATA2	J00036 (MA0036)	1104	5
AACAAT	T01429 Sox-5	R04109 ()	1968	6
GGATGC	Gata1	J00035 (MA0035)	909	6
CATTAG	HOXA5	I00184 (HOXA5)	576	6
GGTGACgTGGCC	T02789 bZIP910 T02790 bZIP911	R08361 () R08366 () R08368 () R08371 () R08373 () R08374 () R08376 () R08379 () R08380 () R08382 () R08384 () R08385 ()	930	12
CTCTCT	TFII-I	I00216 (TFII-I)	1724	6
AAAAAAAAAA	Hb	I00248 (Hb)	1852	10
GRGRTTKcAY	T00572 NF-GMa T01217 NF-GMa	R02214 ()	581	10
AAAAAAAAAA	Hb	I00248 (Hb)	1863	10
TGGTGACT	AT-BP1/AT-BP2	I00141 (AT-BP1/AT-BP2)	1307	8
ATTCCTCTgT	T00333 GR T00698 PR B T01660 PR A	R00980 () R00981 ()	1953	10
MATNNNWAAT	T00630 POU3F2 T01524 POU3F2 T01873 POU3F2	R04327 ()	1990	10
GAGCAAGA	NHP-1	I00285 (NHP-1)	1409	8
CATGTG	T00215 muEBP-C2 T00814 TFE3-S T00875 USF-1 T01559 SREBP-1 T02115 USF2	R00844 () R00886 () R04432 ()	104	6
TTACTC	T00321 GCN4	R00647 ()	744	6
GTCATC	TF68	I00025 (TF68)	1077	6
GAGGAGTAGG	MalT	I00377 (MalT)	406	10
TCTGGCAGG	IgPE-1	I00188 (IgPE-1)	170	9
GGAATG	T01085 abaA	R03771 () R03774 () R03775 () R03776 ()	336	6
CTTTATA	T01484 Cdx-1	M00100 (V$CDXA_01)	504	7

Motif	Factor	Model	Pos.	Length
CTGGAA	T01498 IL-6 RE-BP T01499 IL-6 RE-BP	R01907 ()	1678	6
TCCTCTCT	PEB1	I00052 (PEB1)	1955	8
GCCCaCGTGGCCAG	T00140 c-Myc T00489 Max1	R04462 ()	1188	14
CTCTCT	TFII-I	I00216 (TFII-I)	1712	6
CACTCA	T00918 Zeste T02100 Zeste	R01518 () R01519 () R04939 () R04942 () R04943 () R04944 () R04945 () R04953 () R04955 ()	1446	6
TTGGCA	T00599 NF-1/L	R00078 ()	1030	6
AGCGTG	Arnt-Ahr	J00006 (MA0006)	235	6
AAAAAAAAAA	Hb	I00248 (Hb)	1871	10
CYYNATTAKY	T00377 HOXA5	R02638 ()	1801	10
GTCGTACTC	HMG-1	J00044 (MA0044)	1146	9
CTGGGA	T01498 IL-6 RE-BP T01499 IL-6 RE-BP T01580 STAT6	R04252 ()	189	6
CTTGCC	NF1	I00296 (NF1)	813	6
CCTGG	LBP-1	I00191 (LBP-1)	935	5
TGCGGT	T02256 AML1a	M00271 (V$AML1_01)	239	6
TAATTAGC	T04139 Chx10	R09710 ()	1803	8
GACGC	T00788 T-Ag	Q00168 (-)	900	5
CCGGCCATC	T00278 delta factor T00283 factor delta T00915 YY1	R01833 ()	127	9
AATCAAA	T00997 T00996 SRY	M00148 (V$SRY_01)	1902	7
AGAAC	T00385 HSF1	M00029 (F$HSF_01)	230	5
GCANNNNNNNNNCGC	T02841 FACB	R08614 ()	56	15
cTCTCTCTCTCTC	T00301 GAGA factor	R02059 ()	1708	13
AGAAG	T00386 HSTF	M00028 (I$HSF_01)	356	5
AAGTCA	T00321 GCN4	R00831 ()	1922	6
AAAAA	T00395 Hb	Q00091 (-)	1855	5
TATCAA	T00627 NIT2	M00142 (F$NIT2_01)	588	6
AAAAAAAAAA	Hb	I00248 (Hb)	1855	10
AGCCAG	T00535 NF-1	R03309 ()	1115	6
GCCTGGGGG	TFAP2A	J00003 (MA0003)	850	9
GCNCTNNAG	T00856 TTF-1 T00857 TTF-1 T00858 TTF-1 T00859 TTF-1 T02098 TTF-1	R04645 ()	178	9
GTTTTGT	T02016 En-1	R09102 ()	1327	7
GACCGCA	T00683 PEBP2 T01062 PEBP2alphaA1	R01228 ()	730	7
TATTTTTCC	T01944 NF-ATp T01945 NF-ATc T01946 NF-ATx T01948 NF-ATp T02462 NF-AT3	R05053 ()	478	9
CATCAG	HOXA5	I00184 (HOXA5)	558	6
TTGACC	T03718 WRKY1 T03719 WRKY2 T03720 WRKY3	R09215 () R09218 ()	728	6
CGaTTTTTTT	T00395 Hb	R02502 ()	1732	10
CCCAAC	T00108 C/EBPalpha T00581 C/EBPbeta	R00111 ()	1979	6
TGTTCT	T00042 AR T00333 GR T00335 GR T00697 PR B T00698 PR B T01660 PR A	R01101 () R01103 () R01106 () R01107 () R01108 () R01110 () R01117 () R01120 () R01121 () R01122 () R01312 () R01547 () R01549 () R01550 ()	451	6
GGATG	GATA2	J00036 (MA0036)	909	5
GCYGYYGYCGy	T00008 Adf-1	R04245 ()	1520	11
CCTGG	LBP-1	I00191 (LBP-1)	204	5
CTTGGCCT	T01542 E2F-1	R09581 ()	1173	8
CCWTNTTNNNW	T00278 delta factor T00915 YY1	R04142 ()	485	11
GTCCTG	H-2RIIBP	I00178 (H-2RIIBP)	759	6
aAAATAAACA	T04169 FOXJ2 (long isoform)	R09771 () R09772 () R09784 ()	700	10
AAAGC	PBF	J00064 (MA0064)	1481	5
ACCATT	YY1	J00095 (MA0095)	1950	6
CACCC	CAC-binding	I00019 (CAC-binding)	70	5
GCCATT	YY1	J00095 (MA0095)	1637	6
GGAGCC	T01944 NF-ATp T01948 NF-ATp	R05082 ()	1002	6
TTGGCA	T00599 NF-1/L	R00078 ()	871	6
ATTTGGAAA	T00107 C/EBPalpha	Q00023 (-)	611	9
AATTA	Prrx2	J00075 (MA0075)	1804	5
AGATTA	T02416 SpOtx T02532 GATA-4 T02716 GATA-4	R04993 () R08290 ()	1414	6
CATCAG	T04362 IPF1	R04226 ()	558	6
CCAAT	CP2	I00132 (CP2)	1822	5
AAAGT	PBF	J00064 (MA0064)	924	5
TCGCGG	T03218 E2F+p107	R08844 ()	1081	6
TGGAGG	T00011 ADR1	M00048 (F$ADR1_01)	525	6
ATATAA	T00302 GAL4	R00497 ()	1843	6
TCTTGC	TEF	I00026 (TEF)	812	6
CTCTCTCTCTCTC	T00301 GAGA_factor	Q00066 (-)	1714	13
CCAAC	CP2	I00132 (CP2)	266	5
CTCCTGC	T00528 myogenin	Q00111 (-)	1152	7

Motif	Factor	Model	Pos.	Length
GGCGGG	T00261 ER-alpha T00759 Sp1	R00818 () R01172 () R02027 () R04883 () R04996 ()	752	6
CATCCT	c-ETS	J00098 (MA0098)	132	6
GGATA	GATA2	J00036 (MA0036)	782	5
TGACC	ER	I00276 (ER)	932	5
GGGCGT	Sp1	I00295 (Sp1)	1160	6
GGCTGGG	T00076 CAC-binding_protein	Q00018 (-)	187	7
AAAAAAAAAA	Hb	I00248 (Hb)	1877	10
TGCAAT	EFII	I00275 (EFII)	566	6
CTCAGCTGGA	T00036 AP-4	M00175 (V$AP4_Q5)	1673	10
AGATAT	T02711 GATA-4 T02712 GATA-5A T02713 GATA-5B T02714 GATA-6A T02715 GATA-6B	R08277 ()	639	6
TTATCA	T00267 GATA-1	Q00060 (-)	587	6
TCGNNNNNNNNNNGGA	T02841 FACB	R08595 () R08599 ()	1271	16
TAATGC	TEF	I00026 (TEF)	1761	6
CCTGG	LBP-1	I00191 (LBP-1)	135	5
GGGGTCA	T01331 RXR-alpha T01332 RXR-beta	R04810 ()	1074	7
CAGCTG	Tal-1	I00412 (Tal-1)	1675	6
CATGATGGTG	T00305 GATA-1	M00075 (V$GATA1_01)	1302	10
AGAAG	T00386 HSTF	M00028 (I$HSF_01)	392	5
AAAAA	T00395 Hb	Q00091 (-)	1857	5
TGACC	ER	I00276 (ER)	729	5
TTTTCTA	T01484 Cdx-1	M00101 (V$CDXA_02)	1812	7
TTCAAA	T01159 TFIID	R02749 ()	1918	6
TGATGC	Gata1	J00035 (MA0035)	1140	6
GGGTCA	RAR-beta	I00402 (RAR-beta)	1075	6
GGCCACCA	IUF-1	I00392 (IUF-1)	1561	8
ACTGG	LBP-1	I00191 (LBP-1)	1312	5
AAAAAAAAAA	Hb	I00248 (Hb)	1870	10
TTCCTC	PU.1	I00047 (PU.1)	1954	6
TGACGNNNNNNNT	T00829 TGA1a T00830 TGA1b T02659 OBF4 T02661	R00009 ()	1630	17
GAcG	OBF5 T02794 TGA1 T02795 TGA2 T02796 TGA3 T02797 TGA6 T03722 ZAP1			
GTGGTGT	T00601 NF-1 (-like proteins) T00755 Sp1	R00015 ()	667	7
TATTATA	T01484 Cdx-1	M00101 (V$CDXA_02)	470	7
AGGTCA	Nuclear	J00127 (MF0004)	528	6
CCAAT	CP1	I00066 (CP1)	1585	5
GaCATGTGGC	T00204 E12 T00506 MEF1 T00526 MyoD T00814 TFE3-S T00926 SUM-1	R00240 ()	102	10
TGCACCC	MEP-1	I00078 (MEP-1)	912	7
CGAAG	T00386 HSTF	M00028 (I$HSF_01)	255	5
AAAGC	MNB1A	J00053 (MA0053)	1481	5
TATCCT	c-ETS	J00098 (MA0098)	60	6
GACGC	T00788 T-Ag	Q00168 (-)	997	5
CGCAAC	T00903 XF1 T00905 XF2	R02466 ()	1542	6
CTCTCT	TFII-I	I00216 (TFII-I)	1720	6
CCCTGCgGGG	T02468 AP-2gamma	R05044 ()	179	10
AAAAA	T00395 Hb	Q00091 (-)	1849	5
CACCC	CAC-binding	I00019 (CAC-binding)	914	5
AGAAG	T00386 HSTF	M00028 (I$HSF_01)	2	5
CATGTGGC	T00342 GT-IIBalpha T00563 NF-muE3 T00810 TFE3-L T00811 TFE3 T00814 TFE3-S T00874 USF1 T00877 USF-1 T00978 NFdeltaE3C T01215 NF-muE3 T02378 USF-1	R00853 ()	104	8
TTGATATA	FOXL1	J00033 (MA0033)	1838	8
CCACT	CP2	I00132 (CP2)	880	5
CATGTGGC	E-box-factor	I00162 (E-box-factor)	104	8
TTTTGG	MBF-1	I00083 (MBF-1)	488	6
TGAGCG	T00918 Zeste T02100 Zeste	R01518 () R01523 () R01868 () R01869 () R04940 () R04958 ()	658	6
AGGCGT	Sp1	I00295 (Sp1)	296	6
CCTGG	LBP-1	I00191 (LBP-1)	851	5
CGGNNNANNNTNNNCCG	T00302 GAL4	R00504 ()	1507	17
GCTGG	LBP-1	I00191 (LBP-1)	806	5
TGTGTgCACA	T01388 C/EBP	R04485 ()	1355	10
GCANNNNNNNNNGGC	T02841 FACB	R08592 () R08608 ()	314	15
CACTCG	T00918 Zeste T02100 Zeste	R01518 () R01525 ()	1909	6
TATAAAA	B-factor	I00229 (B-factor)	1844	7
gGGARAWMCM	T00196 Dl	R03463 ()	1652	10
AAAAA	T00395 Hb	Q00091 (-)	1852	5
TATATAA	TBP	I00408 (TBP)	1842	7
CTCTCTCTCTCTC	T00301 GAGA_factor	Q00066 (-)	1710	13
AAAAAAAAAA	Hb	I00248 (Hb)	1848	10
GGATGC	Gata1	J00035 (MA0035)	278	6
CAGAGCG	T00301 GAGA factor	R02066 ()	1601	7
TGTTCT	PR	I00288 (PR)	1289	6
RRCCAAtSRG	T00084 CBF (2) T00150 CP1 T00537 NF-1	R05057 ()	264	10

Motif	Factor	Model	Pos.	Length
TTCCTC	T00702 PU.1	R04413 ()	673	6
TGAGCG	zeste	I00263 (zeste)	658	6
TTCTATTT	Broad-complex_2	J00011 (MA0011)	1814	8
GGTCANNNNNGGTcA	T00719 RAR-alpha1 T01351 T3R-alpha	R03949 ()	1076	15
TGGGGGTTG	Macho-1	J00118 (MA0118)	853	9
CCAATTGG	T02463 GBF1	R09905 ()	1822	8
CACAAAG	T00997 SRY	R03579 ()	921	7
GCTGG	LBP-1	I00191 (LBP-1)	303	5
GCANNNNNNNNGGC	T02841 FACB	R08592 () R08608 ()	285	15
AGGATG	PEA3	I00212 (PEA3)	1103	6
AAAAAAAAAA	Hb	I00248 (Hb)	1872	10
CAGCTG	T00036 AP-4 T00204 E12 T00526 MyoD T00906 XPF-1 T01447 HEN1 T01654 HEN1 T01655 HEN1	R00403 () R01259 () R02076 () R03500 () R04021 () R04244 ()	1675	6
TGACCT	T3R-alpha	I00033 (T3R-alpha)	932	6
TGGGGT	T00011 ADR1	M00048 (F$ADR1_01)	1073	6
CACTTG	T04363 DPBF-1 T04364 DPBF-2	R09889 ()	1425	6
AACAATCAA	T01481 Pbx-1a	M00096 (V$PBX1_01)	1899	9
AAAAA	T00395 Hb	Q00091 (-)	1856	5
GTCATC	H-2RIIBP	I00178 (H-2RIIBP)	1077	6
ACAACA	T00335 GR	R01313 ()	1983	6
TATCGT	T00627 NIT2	R02729 ()	785	6
CCTGG	LBP-1	I00191 (LBP-1)	678	5
CTTGGC	NF1	I00296 (NF1)	1029	6
ACAACA	T00335 GR	R01313 ()	1590	6
AATAAA	TFIID	I00291 (TFIID)	702	6
AAAGT	MNB1A	J00053 (MA0053)	924	5
GCTGG	LBP-1	I00191 (LBP-1)	222	5
TATAAA	T00794 TBP T00796 TBP T00798 TBP T00820 TFIID	R01014 () R02247 () R03172 () R03173 () R03840 () R08480 () R08513 ()	1844	6
GCCNNNNNNNNNNNTGC	T02841 FACB	R08613 ()	1188	17
TTaTTTTTTT	T00395 Hb	R02500 ()	1739	10
TAAACAAA	HNF-5	I00100 (HNF-5)	704	8
TTgGGAAATTT	T00107 C/EBPalpha T01150 AP-3	R02589 ()	612	11
AGAACA	T00333 GR T00337 GR alpha T00696 PR T01661 PR A T01920 GR beta	R00973 ()	1897	6
TATAAAA	T00061 B factor T00197 NC2 T00253 En T00794 TBP T00795 TBP T00797 TBP T00798 TBP T00799 TBP-1 T00817 TFIIA T00818 TFIIB T00820 TFIID T00822 TFIIE T00849 TRF T01038 TFIIF T02168 TFIIF-alpha T02169 TFIIF-beta T02216 TFIIA-alpha/beta precursor (majorT02216 T02217 TFIIA-alpha/beta precursor (minorT02217 T02224 TFIIA-gamma T02242 TFIIF	R00047 () R00770 () R00996 () R00998 () R00999 () R01001 () R01619 () R03156 ()	1844	7
AAAAtAAACA	T04169 FOXJ2 (long isoform)	R09771 () R09772 () R09784 ()	1879	10
GACACA	T00333 GR	R03537 ()	919	6
TGATGG	Gata1	J00035 (MA0035)	1304	6
TGTACA	T00333 GR T00337 GR alpha T00696 PR T01661 PR A T01920 GR beta	R00974 () R01811 ()	1357	6
TTATCA	T00305 GATA-1 T00306 GATA-1 T00308 GATA-2 T01690 elt-2	R02886 () R03737 () R08193 () R08261 () R08934 ()	587	6
AAAGCC	Dof2	J00020 (MA0020)	1481	6
CTCCTGC	T00528 myogenin	Q00111 (-)	948	7
CTCTCT	TFII-I	I00216 (TFII-I)	1714	6
AGCCACT	T00152 CP2	R00444 ()	878	7
ACTTTA	T02695 BBF1	R08285 ()	503	6
TGGAGG	T00011 ADR1	M00048 (F$ADR1_01)	1283	6
CTTGGC	NF1	I00296 (NF1)	1427	6
CTCTCTCTCTCTC	T00301 GAGA factor	R02059 ()	1716	13
TTTGGAAAT	AP-3	I00271 (AP-3)	612	9
GTATATTAT	T00119 CF2-I	R02317 ()	467	9
GATGaYRTGR	T00668 Opaque-2	R02939 ()	1105	10
ATGTGAA	NBF	I00323 (NBF)	1687	7
GGAAGTGAcG	T00671 p53	R04384 ()	1680	10
AGAAA	T00385 HSF1	M00029 (F$HSF_01)	1655	5
GCCATGTG	NF-InsE2 NF-InsE3	I00069 (NF-InsE3) I00070 (NF-InsE2)	102	8
tAAATAAACA	T04169 FOXJ2 (long isoform)	R09749 () R09752 () R09764 ()	700	10
TGCCAGTA	FOXC1	J00032 (MA0032)	815	8
GACCTG	H-2RIIBP	I00178 (H-2RIIBP)	933	6
GCGCCCT	EIIaE-A	I00143 (EIIaE-A)	892	7
GcCGGATGCA	T00111 c-Ets-1	R04060 ()	906	10
YNSNRNSTNGCgTGNNW	T00018 AhR	R02656 ()	852	17

Motif	Factor	Model	Pos.	Length
AAAAA	T00395 Hb	Q00091 (-)	1869	5
TGTGGT	T01067 AML1	R04441 ()	666	6
AGGTCA	T00619 NGFI-B T00694 PPAR-alpha T00991 PPAR-alpha	R02233 () R03961 ()	528	6
AATTA	Prrx2	J00075 (MA0075)	1692	5
TGTTCT	GR/PR	I00104 (GR/PR)	451	6
GAGGC	T00788 T-Ag	Q00168 (-)	295	5
AAAGC	PBF	J00064 (MA0064)	876	5
TGCACCCG	T00916 ZAP	R08305 ()	912	8
AAGTGA	T00422 IRF1 T00425 IRF-2	R00917 ()	1682	6
GCCATC	YY1	J00095 (MA0095)	130	6
TTaTTTTTTT	T00395 Hb	R02500 ()	1735	10
TGTTCT	T00042 AR T00333 GR T00335 GR T00697 PR B T00698 PR B T01660 PR A	R01101 () R01103 () R01106 () R01107 () R01108 () R01110 () R01117 () R01120 () R01121 () R01122 () R01312 () R01547 () R01549 () R01550 () R02995 () R03890 ()	1289	6
CTTTCT	TFII-I	I00216 (TFII-I)	455	6
AGAACA	T00333 GR T00337 GR alpha T00696 PR T01661 PR A T01920 GR beta	R00973 ()	230	6
AGGAGT	T00011 ADR1	M00048 (F$ADR1_01)	407	6
TGCTCCT	Zmhoxla	I00358 (Zmhoxla)	952	7
GAACAAT	Sox5	J00087 (MA0087)	1898	7
TTaTTTTTTT	T00395 Hb	R02500 ()	1736	10
CCTAATTAGg	T00500 MCM1	R01901 () R01902 ()	1801	10
AAAAAAAAAA	Hb	I00248 (Hb)	1865	10
CTGGGG	MIG1	I00321 (MIG1)	852	6
TTGGAAAT	Oct-1	I00286 (Oct-1)	613	8
CTTGCC	NF1	I00296 (NF1)	1788	6
AAAAAAAAAA	Hb	I00248 (Hb)	1858	10
AGAACA	T00335 GR	Q00076 (-)	1897	6
AAAAAAAAAA	Hb	I00248 (Hb)	1866	10
GAGgCGCGGA	T00759 Sp1	R02103 ()	995	10
AAAAA	T00395 Hb	Q00091 (-)	1851	5
CCACT	CP2	I00132 (CP2)	647	5
AACAAT	T01429 Sox-5	R04109 ()	1899	6
GAGGC	T-Ag	I00215 (T-Ag)	1697	5
AAAGGG	Dof3	J00021 (MA0021)	605	6
TGATA	GATA2	J00036 (MA0036)	445	5
GATATATA	FOXL1	J00033 (MA0033)	1840	8
GACATG	H-2RIIBP	I00178 (H-2RIIBP)	1300	6
CAAAAG	HOXA5	I00184 (HOXA5)	1479	6
AAAAAAAACA	Hb	I00248 (Hb)	1879	10
ATTCCT	T00499 MCBF	R02073 () R02074 ()	1953	6
AAGGAA	T00113 c-Ets-2	R04338 ()	334	6
AGGAGT	Sp1	I00295 (Sp1)	407	6
TAAACAAAcA	T04169 FOXJ2 (long isoform)	R09753 () R09754 () R09758 ()	704	10
GGATG	GATA2	J00036 (MA0036)	278	5
CTTGGC	T00601 NF-1 (-like proteins)	R01192 ()	1029	6
GAGGTCAGGG	T00721 RAR-beta T01334 RXR-beta	R01063 ()	527	10
AAAAAAAAAA	Hb	I00248 (Hb)	1864	10
TTCAAA	T01159 TFIID	R02749 ()	1992	6
GTCCTG	H-2RIIBP	I00178 (H-2RIIBP)	18	6
CAAATAAACA	T04169 FOXJ2 (long isoform)	R09783 ()	700	10
AAAAAAAAAA	Hb	I00248 (Hb)	1847	10
AAAAAAAAAA	Hb	I00248 (Hb)	1860	10
TATATA	TFIID	I00338 (TFIID)	1842	6
GGTTATG	IHF	I00375 (IHF)	520	7
AGAAA	T00386 HSTF	M00028 (I$HSF_01)	1655	5
AAAAA	T00395 Hb	Q00091 (-)	1850	5
CGTGAC	T00870 USF	R00711 ()	375	6
AAAGC	MNB1A	J00053 (MA0053)	876	5
AAATAAA	T00798 TBP	R03841 ()	701	7
GAAACATC	OmpR	I00380 (OmpR)	1656	8
GCCGACGG	MafB	J00117 (MA0117)	1239	8
AGGAGG	Sp1	I00295 (Sp1)	293	6
ACGGAGTA	FOXC1	J00032 (MA0032)	379	8
CTTGGC	T00601 NF-1 (-like proteins)	R01192 ()	1263	6
CTTTGT	T02857 TCF-3	R08594 ()	1701	6
GGAAGT	MAF	I00394 (MAF)	1680	6
TCAGTTCC	V$CAP_01	M00253 (V$CAP_01)	1388	8
AAAAAAAAAA	Hb	I00248 (Hb)	1857	10
AAAAAAAAAA	Hb	I00248 (Hb)	1854	10
TTTGGAAAT	T00107 C/EBPalpha T01150 AP-3	R02583 ()	612	9
GCTGG	LBP-1	I00191 (LBP-1)	1157	5
GCTGT	GT-IIBa	I00160 (GT-IIBa)	664	5
TATATTATA	T00119 CF2-I	R02316 ()	468	9
TGCACCC	T02354 MTF-1	R08299 ()	912	7
CCAAT	CBP/CRF	I00135 (CBP/CRF)	1219	5

Motif	Factor	Model	Pos.	Length
GGGTCA	Nuclear	J00127 (MF0004)	1075	6
AAAAAAAAAA	Hb	I00248 (Hb)	1862	10
TATATA	TFIID	I00259 (TFIID)	1842	6
AGTCCTcCCC	T00392 H4TF-2	R08215 ()	17	10
AAAGCC	Dof3	J00021 (MA0021)	1481	6
CCTGC	LVc	I00085 (LVc)	773	5
AGCGTG	Arnt-Ahr	J00006 (MA0006)	1134	6
RTCRYNNNNNACG	T00778 TAF	R02579 ()	1080	13
TATACA	TFIID	I00259 (TFIID)	507	6
TGTTCT	PR	I00288 (PR)	451	6
CTTTGA	T02857 TCF-3	R08594 ()	1627	6
AGGAGT	Sp1	I00295 (Sp1)	272	6
AAAAAAAAAA	Hb	I00248 (Hb)	1856	10
CTTGTTTTC	HMG-1	J00044 (MA0044)	1750	9
ATATATA	T01484 Cdx-1	M00101 (V$CDXA_02)	1841	7
CCAAT	CP1	I00066 (CP1)	1219	5
TAAT	Ubx	J00094 (MA0094)	1803	4
YNSNrNSTNGCGT GNNW	T00018 AhR	R02656 ()	1126	17
TGCRCRC	T00495 MBF-I	R02198 ()	90	7
CCAAT	CP2	I00132 (CP2)	1585	5
CTTTCT	TFII-I	I00216 (TFII-I)	1669	6
ACAACA	T00335 GR	R01313 ()	510	6
AAAAA	T00395 Hb	Q00091 (-)	1868	5
TCTCAgTGCA	T00009 Adf-2a	R00055 ()	84	10
TgTGTGTGTGTGT GTGTGTGTGTGTG TGTGTgCAtA	T00900 WT1 I -KTS T01839 WT1 -KTS T01840 WT1 I T01841 WT1-del2 T01842 WT1 I-del2	R02309 ()	1329	36
TAtACAAATA	T04169 FOXJ2 (long isoform)	R09750 ()	704	10
TCTGGC	T00535 NF-1	Q00112 (-)	170	6
AACAAAT	T00997 SRY	R03581 ()	706	7
CTTGGC	T00601 NF-1 (-like proteins)	R01192 ()	1173	6
CTCTCTCTCTcTC	T00301 GAGA factor	R02059 ()	1720	13
CGAGCG	zeste	I00263 (zeste)	1065	6
CTCAGCTGGA	T00036 AP-4	M00176 (V$AP4_Q6)	1673	10
CATCCT	c-ETS	J00098 (MA0098)	1660	6
TTAGGG	T01246 TBF1	R03755 ()	578	6
TCCCCCAAgAA	T01040 Olf-1 T01112 COE1	R03622 ()	1976	11
CAGTTC	T00137 c-Myb	R04184 ()	1389	6
CTCTCTCTCTCTC	T00301 GAGA factor	R02059 ()	1712	13
CAATAG	HOXA5	I00184 (HOXA5)	1220	6
GAACAG	LVa	I00193 (LVa)	231	6
GCGCCCTT	T00246 EIIaE-A	R00327 ()	892	8
ACCATG	YY1	J00095 (MA0095)	1565	6
GCCATT	YY1	J00095 (MA0095)	1469	6
ATTGAA	T00137 c-Myb	R04082 ()	1640	6
AAAAAAAAAA	Hb	I00248 (Hb)	1867	10
TCCNNNNNNNNNN GCA	T02841 FACB	R08611 ()	19	16
TATATA	TFIID	I00338 (TFIID)	416	6
CTGGGA	T01498 IL-6 RE-BP T01499 IL-6 RE-BP T01580 STAT6	R04252 ()	1313	6
RTCRYYNNNNACG	T00056 ABF1 T01379 120-kDa CRE-binding protein	R02377 ()	1523	13
CTCTCTCTCTCTC	T00301 GAGA factor	R02059 ()	1714	13
GGGTGCA	T00515 MTF-1	R08297 ()	563	7
AAAAA	T00395 Hb	Q00091 (-)	1859	5
GCCAAT	T00095 CCAAT-binding factor T00153 alpha-CP1 T01185 CCAAT-binding factor T01203 alpha-CP1	R00511 ()	1218	6
ATTTNNNNATTT	T00360 HiNF-A	R02171 ()	611	12
GGATA	GATA2	J00036 (MA0036)	1828	5
AGAAT	T00386 HSTF	M00028 (I$HSF_01)	1889	5
GTTGGC	T00535 NF-1	Q00112 (-)	1164	6
AGAAC	T00385 HSF1	M00029 (F$HSF_01)	1897	5
CTAATTA	T03458 Crx T03461 Crx	R09049 ()	1802	7
CGGNNNNNNCCG	T02846 UAY	R08868 () R08871 ()	908	12
GCCGCGCT	GCF	I00152 (GCF)	798	8
CCTGC	LVc	I00085 (LVc)	20	5
TTTTCCTTTTTG	ID1	J00120 (MA0120)	481	12
CAAACAATAC	T01429 Sox-5	M00042 (V$SOX5_01)	1966	10
CGATGT	Gata1	J00035 (MA0035)	1253	6
GATGTGAAaT	T01241 INO2	R03839 ()	1686	10
TGATGG	Gata1	J00035 (MA0035)	1127	6
AATAATG	T01484 Cdx-1	M00101 (V$CDXA_02)	1759	7
GCGCCCT	EIIaE-A	I00143 (EIIaE-A)	1186	7
GAGCAAGA	NFAT-1	I00063 (NFAT-1)	1409	8
CTCTCTCTCTCTC	T00301 GAGA factor	R02059 ()	1710	13
GTGACC	T00635 NP-III	R02044 ()	931	6
CAAATAAAtAA	T00395 Hb	R03022 ()	700	11
AAAAAAAAAA	Hb	I00248 (Hb)	1875	10
TTGATATT	DEF	I00073 (DEF)	444	8
TTCCTCT	Elf-1	I00113 (Elf-1)	1954	7

Motif	Factor	Model	Pos.	Length
AAAAAAAAAA	Hb	I00248 (Hb)	1869	10
GCCATG	YY1	J00095 (MA0095)	102	6
TTTTCC	T00550 NFAT-1 T01944 NF-ATp T01945 NF-ATc	R05080 ()	481	6
GGCAGT	MAF	I00394 (MAF)	55	6
GGGAGG	Sp1	I00295 (Sp1)	1315	6
GCTGG	LBP-1	I00191 (LBP-1)	1767	5
SMGGAWGY	T00111 c-Ets-1 T00112 c-Ets-1 T00114 c-Ets-1 54 T00115 c-Ets-1 68 T00684 PEA3 T00685 PEA3 T00686 PEA3	R02153 ()	1102	8

5' Region of PLIP

Motif	Factor	Model	Pos.	Length
TCTTCT	T00335 GR	R01313 ()	104	6
GCGGC	T00788 T-Ag	Q00168 (-)	360	5
MYYMGCCYM	T00752 Sp1 T00753 Sp1 T00754 Sp1 T00755 Sp1	R02245 ()	1191	9
TTYCCAG	T00017 C/EBPbeta	R02170 ()	220	7
GAGGC	T00788 T-Ag	R01244 ()	750	5
CCACC	CAC-binding	I00019 (CAC-binding)	1197	5
CTTCCC	T00113 c-Ets-2	R04343 ()	264	6
GGCAAA	T00033 AP-2alpha T00035 AP-2alphaA	R08502 ()	945	6
CTTTCC	NP-TCII	I00400 (NP-TCII)	218	6
GCTGT	GT-IIBa	I00160 (GT-IIBa)	158	5
TCTCC	T00011 ADR1	R00074 ()	937	5
YAACKG	T00136 c-Myb T00137 c-Myb T00138 c-Myb T00139 c-Myb	R02206 ()	607	6
TTTCT	T00386 HSTF	M00028 (I$HSF_01)	45	5
CTACCGCT	BUF	I00303 (BUF)	543	8
CTTCCT	PEA3	I00212 (PEA3)	149	6
TTTCT	T00386 HSTF	M00028 (I$HSF_01)	557	5
ATTGG	CP2	I00132 (CP2)	36	5
GCCGC	T00788 T-Ag	Q00168 (-)	649	5
GGATG	T00322 GCR1	R03806 ()	1208	5
TGCCC	T00261 ER-alpha T00467 LF-A1 T00759 Sp1	R01171 () R04883 ()	464	5
CCTGG	LBP-1	I00191 (LBP-1)	436	5
ACCTTT	NF-BA1	I00072 (NF-BA1)	554	6
TGACG	T00163 CREB	R02710 ()	524	5
TCTATC	T00305 GATA-1	R04296 ()	510	6
GGTCC	H4TF2	I00180 (H4TF2)	488	5
CCCTCCTC	T00105 C/EBPalpha	R03657 ()	267	8
GATTGG	T00174 CTF T01186 CCAAT-binding factor	R00765 () R00768 ()	1253	6
GAGGG	T00172 CTCF	R02137 ()	444	5
TTTGCC	T00033 AP-2alpha T00035 AP-2alphaA	R08502 ()	293	6
TCGC	T03218 E2F+p107	R08845 ()	974	4
GAATCA	T00321 GCN4	R00832 ()	626	6
TGCGC	T00752 Sp1	R01021 ()	781	5
TGGC	T00537 NF-1	R01681 ()	919	4
CATCC	T00322 GCR1	R03806 ()	246	5
YTATCW	T00267 GATA-1 T00304 GATA-1A T00306 GATA-1 T00307 GATA-2 T00308 GATA-2 T00309 GATA-	R02156 () R02157 () R02158 ()	511	6
CTGCCC	T00753 Sp1	R01540 ()	872	6
TATATA	TFIID	I00338 (TFIID)	835	6
ATTGG	CP2	I00132 (CP2)	1254	5
TGTTCT	GR/PR	I00104 (GR/PR)	868	6
TCCTC	T00302 GAL4	R00492 ()	321	5
MAMAG	T00803 TCF-2alpha T00930 LEF-1 T00999 TCF-1A T01000 TCF-1B T01001 TCF-1C T01002 TCF-1 T01109 TCF-1(P) T01979 TCF-1E T01981 TCF-1F T01982 TCF-1G T02905 LEF-1	R02248 ()	776	5
AAAG	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 () R08442 () R08443 ()	641	4
AGGACA	T00333 GR	R00429 ()	568	6
WNNNAANAWGG	T00278 delta factor T00915 YY1	R04142 ()	1014	11

Motif	Factor	Model	Pos.	Length
AGGCT	T00302 GAL4	R00495 ()	634	5
CACCTG	T00140 c-Myc T00878 USF2 T04353 CAN	R09477 () R09876 ()	1198	6
TTATCT	GATA-1	I00109 (GATA-1)	240	6
GNGYGCA	T00515 MTF-1	R02204 ()	770	7
TTCCCAG	IL-6.RE-BP	I00218 (IL-6.RE-BP)	220	7
TTATCT	GATA-3	I00108 (GATA-3)	116	6
GACAG	T01214 NF-E	R00558 ()	956	5
CTTT	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 R08442 () R08443	851	4
TCCA	T00537 NF-1	R01681 ()	536	4
TCGC	T03218 E2F+p107	R08845 ()	143	4
TTCCCAG	H-APF-1	I00175 (H-APF-1)	220	7
TGGC	T00537 NF-1	R01681 ()	1214	4
TCGCCA	T00275 F-ACT1	R02154 ()	722	6
CCGA	T00714 RAF	R00256 ()	209	4
AAATG	T00915 YY1	R00603 ()	680	5
CCACACCC	T00807 TEF-2	R02251 ()	345	8
GGGACA	T00333 GR	R03539 ()	1129	6
TCTATC	T00305 GATA-1	R04296 ()	602	6
CATYAS	T04362 IPF1	R04228 ()	1156	6
TGGA	T00537 NF-1	R01681 ()	968	4
TTATCT	GATA-3	I00108 (GATA-3)	240	6
RCWTCCKS	T00111 c-Ets-1 T00112 c-Ets-1 T00114 c-Ets-1 54	R02153 ()	1080	8
TCGC	T03218 E2F+p107	R08845 ()	286	4
TATATA	TFIID	I00338 (TFIID)	835	6
MAMAG	T00803 TCF-2alpha T00930 LEF-1 T00999 TCF-1A T01000 TCF-1B T01001 TCF-1C T01002 TCF-1	R02248 ()	947	5
ATTGG	CP1	I00066 (CP1)	1254	5
CTKTK	T00803 TCF-2alpha T00930 LEF-1 T00999 TCF-1A T01000 TCF-1B T01001 TCF-1C T01002 TCF-1 T01109 TCF-1(P) T01979 TCF-1E T01981 TCF-1F T01982 TCF-1G T02905 LEF-1	R02248 ()	292	5
CCCC	T00302 GAL4	R00496 ()	1270	4
ATCTCA	T00918 Zeste T02100 Zeste	R04948 ()	242	6
GCTGT	GT-IIBa	I00160 (GT-IIBa)	866	5
TATAT	T00765 SRF (504 AA)	R09535 ()	74	5
GATAG	T01214 NF-E	R00554 ()	1119	5
GCGA	T03218 E2F+p107	R08845 ()	707	4
CCCNCNNNCT	T00913 Yi	R03154 ()	503	10
WNNATGAC	T01599 LCR-F1	R05088 ()	496	8
TTATC	T00302 GAL4 T00794 TBP	R00494 () R03182 ()	116	5
CATGTTGC	DTF-1	I00235 (DTF-1)	1074	8
AGATAG	T00305 GATA-1	R04294 ()	1118	6
CGTCA	T00163 CREB	R02710 ()	671	5
CCTGC	T00478 LVc	R01644 ()	173	5
MYYMGCCYM	T00752 Sp1 T00753 Sp1 T00754 Sp1 T00755 Sp1 T00757 Sp1 T00758 Sp1 T00759 Sp1 T01228 Sp1	R02245 ()	275	9
CACTGGAT	T00150 CP1 T00306 GATA-1 T00641 POU2F1	R04200 ()	965	8
GCCA	T00537 NF-1	R01681 ()	724	4
CACCTG	Tal-1	I00412 (Tal-1)	1198	6
TGAC	T00029 AP-1	R00368 ()	524	4
CTGTC	T01214 NF-E	R00558 ()	1201	5
CCWTNTTNNNW	T00278 delta factor T00915 YY1	R04142 ()	1231	11
GTTTGC	EBP-45	I00136 (EBP-45)	1031	6
AAGACC	RC2	I00329 (RC2)	135	6
TGGC	T00537 NF-1	R01681 ()	327	4
CCACC	T00076 CAC-binding protein	R04295 ()	725	5
ATTWNNATK	T00630 POU3F2 T01524 POU3F2 T01873 POU3F2	R04327 ()	472	9
GATGACC	T00724 RC2	R00260 ()	498	7
TATAT	T00765 SRF (504 AA)	R09535 ()	835	5
TGAC	T00029 AP-1	R00368 ()	500	4

Motif	Factor	Model	Pos.	Length
GCGA	T03218 E2F+p107	R08845 ()	528	4
WGATAR	T00267 GATA-1 T00304 GATA-1A T00307 GATA-2 T00308 GATA-2 T00309 GATA-2 T00310 GATA-3 T00311 GATA-3 T00312 GATA-3 T00313 GATA-3 T00314 GATA-3 T00567 NF-E1b T02312 GATA-1B	R02156 () R02157 () R02158 ()	1118	6
TGGCTCA	T00923 Zta	Q00195 (-)	919	7
GTCA	T00029 AP-1	R00368 ()	677	4
GCAGGACC	T00843 Ttk 69K	R02061 ()	614	8
GTCA	T00029 AP-1	R00368 ()	672	4
MTTNCNNMA	T00017 C/EBPbeta T00459 C/EBPbeta T00581 C/EBPbeta T00583 C/EBPdelta	R02216 ()	254	9
GCAGG	T00478 LVc	R01644 ()	614	5
GATAA	T00302 GAL4 T00794 TBP	R00494 () R03182 ()	884	5
STRATG	T04362 IPF1	R04228 ()	745	6
GCCCA	T00788 T-Ag	R01372 ()	1195	5
GCCTC	T00788 T-Ag	R01244 ()	279	5
AAAG	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 () R08442 () R08443 ()	991	4
TCGC	T03218 E2F+p107	R08845 ()	315	4
GAGGA	T00302 GAL4	R00492 ()	567	5
MAMAG	T00803 TCF-2alpha T00930 LEF-1 T00999 TCF-1A T01000 TCF-1B T01001 TCF-1C T01002 TCF-1 T01109 TCF-1(P) T01979 TCF-1E T01981 TCF-1F T01982 TCF-1G T02905 LEF-1	R02248 ()	990	5
CTKTK	T00803 TCF-2alpha T00930 LEF-1 T00999 TCF-1A T01000 TCF-1B T01001 TCF-1C T01002 TCF-1	R02248 ()	867	5
TCCTGC	Ttk	I00261 (Ttk)	172	6
GCGA	T03218 E2F+p107	R08845 ()	1040	4
TGGCA	T00599 NF-1/L	R01322 ()	1214	5
TGAC	T00029 AP-1	R00368 ()	911	4
TTCCTT	T00113 c-Ets-2	R04339 ()	451	6
GCCTC	T00788 T-Ag	R01244 ()	20	5
GCCTC	T00788 T-Ag	R01244 ()	1011	5
TTTCT	T00386 HSTF	M00028 (I$HSF_01)	56	5
ATTGG	CP1	I00066 (CP1)	36	5
CCGGCCAG	LBP-1	I00280 (LBP-1)	364	8
GCCCA	T00788 T-Ag	R01372 ()	1035	5
TGTTCT	T00042 AR T00333 GR T00335 GR T00697 PR B T00698 PR B T01660 PR A	R01101 () R01120 R01122 () R01312 () R01550	868	6
ACAACA	T00335 GR	R01313 ()	571	6
GAGGA	T00302 GAL4	R00492 ()	684	5
CATCC	T00322 GCR1	R03806 ()	1081	5
CTATC	T01214 NF-E	R00554 ()	511	5
TGTCCC	T00333 GR	R03539 ()	1178	6
CTGTC	T01214 NF-E	R00558 ()	159	5
CCCTAA	T01246 TBF1	R03755 ()	656	6
CTKGNTNKNGC	T00152 CP2	R04171 ()	967	11
CAAATG	twi	I00266 (twi)	679	6
TATATTATA	T00119 CF2-I	R02316 ()	74	9
TGCGC	T00752 Sp1	R01021 ()	353	5
ATTGG	CBP/CRF	I00135 (CBP/CRF)	1254	5
GGTGG	CAC-binding	I00019 (CAC-binding)	917	5
TGTGC	T00754 Sp1	R08166 ()	704	5
GTCCTC	H-2RIIBP	I00178 (H-2RIIBP)	320	6
TATA	T00061 B factor T00286 factor i T00557 NF-E2	R00689 () R01129 () R04293 ()	835	4
CCGA	T00714 RAF	R00256 ()	620	4
TATCTC	T00627 NIT2	R02731 ()	241	6
GCTGT	GT-IIBa	I00160 (GT-IIBa)	1056	5
TTTCT	T00386 HSTF	M00028 (I$HSF_01)	473	5
CCWTNTTNNNW	T00278 delta factor T00915 YY1	R04142 ()	32	11
CATGTTGC	T00201 DTF-1	R01874 ()	1074	8

Motif	Factor	Model	Pos.	Length
CGATA	T04321 BEAF-32A T04322 BEAF-32B	R09849 ()	883	5
CCTGC	LVc	I00085 (LVc)	173	5
CCCTC	T00172 CTCF	R02137 ()	875	5
TGCCC	T00261 ER-alpha T00467 LF-A1 T00759 Sp1	R01171 () R04883	873	5
CCTTTG	TCF-1	I00029 (TCF-1)	850	6
TTATAT	T00302 GAL4	R00497 ()	73	6
CCACACCCT	CAC-BF	I00137 (CAC-BF)	345	9
TAAC	T00137 c-Myb	R04341 ()	886	4
GCCA	T00537 NF-1	R01681 ()	367	4
GAGGAC	H-2RIIBP	I00178 (H-2RIIBP)	567	6
AGGACA	T3R-alpha	I00033 (T3R-alpha)	568	6
GGTGG	T00076 CAC-binding protein	R04295 ()	917	5
TGTTCT	T00337 GR alpha T00696 PR T01661 PR A T01920 GR beta	R00973 ()	868	6
GCCA	T00537 NF-1	R01681 ()	356	4
GATGAC	TF68	I00025 (TF68)	498	6
TCGCGG	T03218 E2F+p107	R08844 ()	1204	6
ACAAAG	T02857 TCF-3	R08594 ()	989	6
GCAGGACC	T00843 Ttk_69K	M00009 (I$TTK69_01)	614	8
CCGCCC	T00270 ETF T01171 CP1	R01375 () R03039	1193	6
RTRTTGCA	T00212 E2BP	R02140 ()	1075	8
CCACC	T00076 CAC-binding protein	R04295 ()	805	5
CMGTTR	T00136 c-Myb T00137 c-Myb T00138 c-Myb T00139 c-Myb	R02206 ()	1174	6
CAAAGG	TCF-1	I00029 (TCF-1)	640	6
CTTGTG	T01649 HES-1	R04468 ()	91	6
CCGA	T00714 RAF	R00256 ()	1251	4
TAAC	T00137 c-Myb	R04341 ()	121	4
GCCGC	T00788 T-Ag	Q00168 (-)	408	5
GCAAT	T00104 C/EBPalpha T00105 C/EBPalpha T00107 C/EBPalpha T00108 C/EBPalpha	R02132 ()	460	5
CYSATTGGYY	T00084 CBF (2) T00150 CP1 T00537 NF-1	R05057 ()	1251	10
TCGG	T00714 RAF	R00256 ()	762	4
TATCGT	T00627 NIT2	R02729 ()	1062	6
GCACA	T00754 Sp1	R08166 ()	987	5
TGGC	T00537 NF-1	R01681 ()	1009	4
TCGC	T03218 E2F+p107	R08845 ()	722	4
CTGTTC	LVa	I00193 (LVa)	867	6
CCCC	T00302 GAL4	R00496 ()	61	4
GCCTC	T-Ag	I00215 (T-Ag)	759	5
TGACTCC	T01140 AP-1	R02582 ()	84	7
TCNNNNNNNACG	T00056 ABF1	R01923 ()	178	12
CATATG	T04353 CAN T00081 alpha-CBF T00100 CUTL1 T00151 CP2 T00153 alpha-CP1 T00156 alpha-CP2a T00388 H1TF2 T00426 alpha-IRP T00538 NF-1 T00556 NF-E T00764 SRF T00832 TGGCA-binding protein T01202 CP1 T01214 NF-E T01995 Clox T02042	R09877 () R00039 () R00510 () R00532 () R00561 () R00562 () R00571 () R00572 () R00660	385	6
ATTGG	Cutl1 T02043 CDP2 alpha-CP2b	() R01543R02848	1254	5
CCAGG	LBP-1	I00191 (LBP-1)	223	5
TGACTC	T00321 GCN4	R00829 ()	84	6
CCTCCTCCTC	MalT	I00377 (MalT)	268	10
GCCTC	T-Ag	I00215 (T-Ag)	1011	5
CACAAG	T01649 HES-1	R04468 ()	630	6
GAGGAA	PU.1	I00047 (PU.1) R00231 () R00232 () R00335 ()	684	6
ATTGG	T00082 CBF (1) T00083 CBF (2) T00087 CBF-A T00088 CBF-B T00092 CCAAT-binding factor T00099 CDF T00108 C/EBPalpha T00170 CRF T00613 NF-Y T00614 NF-Y'	R00668 () R00669 () R01081 () R01445 () R01446 ()	1254	5
CCAAT	CP2	I00132 (CP2)	395	5
TCGG	T00714 RAF	R00256 ()	561	4
AGCCT	T00302 GAL4	R00495 ()	758	5
CTTTCC	NP-TCII	I00400 (NP-TCII)	341	6
TGACCCC	T01331 RXR-alpha T01332 RXR-beta	R04810 ()	500	7

Motif	Factor	Model	Pos.	Length
GAGGA	T00302 GAL4	R00492 ()	1092	5
GTCA	T00029 AP-1	R00368 ()	951	4
GCCA	T00537 NF-1	R01681 ()	645	4
GATTTC	T00391 H4TF-1	R00680 ()	253	6
CANCTGY	T00204 E12 T00207 E47	R02139 ()	1198	7
TTTCT	T00386 HSTF	M00028	102	5
CTKTK	T00803 TCF-2alpha T00930 LEF-1 T00999 TCF-1A T01000 TCF-1B T01001 TCF-1C T01002 TCF-1 T01109 TCF-1(P) T01979 TCF-1E T01981 TCF-1F	R02248 ()	851	5
CACTCC	T00918 Zeste T02100 Zeste	R04950 ()	1133	6
AGATAG	T00306 GATA-1	R00566 ()	1118	6
TCCTTC	Ttk	I00261 (Ttk)	452	6
TATAT	T00765 SRF (504 AA)	R09535 ()	79	5
TATCTC	T00627 NIT2	Q00133 (-)	241	6
CCACACCC	TEF2	I00027 (TEF2)	345	8
AGAGGACA	T00697 PR B	R01555 ()	566	8
TGAC	T00029 AP-1	R00368 ()	84	4
TTACTC	T00321 GCN4	R00647 ()	1158	6
TTGGCT	T00537 NF-1	R01681 ()	593	6
GAGGAC	H-2RIIBP	I00178 (H-2RIIBP)	1092	6
TGACCC	T00721 RAR-beta	Q00153 (-)	500	6
GAGGC	T00788 T-Ag	R01244 ()	585	5
CTCTCT	TFII-I	I00216 (TFII-I)	507	6
CCTGC	LVc	I00085 (LVc)	351	5
TCTATC	T00305 GATA-1	R04296 ()	1060	6
CTTTCC	T00539 NF-1	R00802 ()	341	6
TGCRCRC	T00495 MBF-I	R02198 ()	781	7
TGCAC	T00752 Sp1	R01021 ()	986	5
GTTCT	T00385 HSF1	M00029 (F$HSF_01)	869	5
CTATC	T01214 NF-E	R00554 ()	1189	5
CTCCTTTCT	ILF	I00093 (ILF)	41	9
CTKGNTNKNGC	T00152 CP2	R04171 ()	191	11
TTGGCT	T00537 NF-1	R01681 ()	37	6
YTATCW	T00267 GATA-1 T00304 GATA-1A T00306 GATA-1 T00307 GATA-2 T00308 GATA-2 T00309 GATA-2 T00310 GATA-3 T00311 GATA-3 T00312 GATA-E1b T02312 GATA-1B	R02156 () R02157 () R02158 ()	603	6
TATCG	T04321 BEAF-32A T04322 BEAF-32B	R09849 ()	1062	5
GTCA	T00029 AP-1	R00368 ()	161	4
TCGTCA	T00321 GCN4	R00644 ()	670	6
GGGG	T00302 GAL4	R00496 ()	1128	4
GCCGC	T00788 T-Ag	Q00168 (-)	753	5
CTATC	T01214 NF-E	R00554 ()	603	5
CCAAT	CBP/CRF	I00135 (CBP/CRF)	395	5
TTCCTT	T00113 c-Ets-2	R04339 ()	150	6
CATYAS	T04362 IPF1	R04228 ()	258	6
TGCRCNC	T00515 MTF-1	R02204 ()	781	7
GTTGCATC	T00608 NF-W1 T00609 NF-W2	R02220 () R02221	1077	8
TCCTTC	Ttk	I00261 (Ttk)	147	6
GCCTC	T-Ag	I00215 (T-Ag)	20	5
CCCC	T00302 GAL4	R00496 ()	1249	4
GGTCC	T00392 H4TF-2	R00681 ()	488	5
CTTTGA	T02857 TCF-3	R08594 ()	851	6
TGCCTGG	T00528 myogenin	R00018 ()	434	7
GCGCA	T00752 Sp1	R01021 ()	772	5
ACCACA	T02256 AML1a	M00271 (V$AML1_01)	422	6
CTGTTC	T00475 LVa	R01135 ()	867	6
TATCTA	T00627 NIT2	M00142 (F$NIT2_01)	117	6
TTTCCACA	T00410 IgPE-1	R00849 ()	342	8
CATTCY	T01085 abaA	R03779 ()	305	6
TTATC	T01274 ABF2	R03826 ()	240	5
CTGATG	HOXA5	I00184 (HOXA5)	745	6

Motif	Factor	Model	Pos.	Length
GCGA	T03218 E2F+p107	R08845 ()	1238	4
TGACCC	RAR-beta	I00402 (RAR-beta)	500	6
GGTCTA	RC2	I00329 (RC2)	600	6
TCGC	T03218 E2F+p107	R08845 ()	406	4
GGACC	T00392 H4TF-2	R00681 ()	617	5
AAAG	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 () R08442 () R08443 ()	539	4
TCGG	T00714 RAF	R00256 ()	1126	4
CTGTC	T01214 NF-E	R00558 ()	1057	5
GCAAT	T00104 C/EBPalpha T00105 C/EBPalpha T00107 C/EBPalpha T00108 C/EBPalpha	R02132 ()	900	5
TTATCT	T00267 GATA-1	Q00060 (-)	116	6
CATAAAAC	HOXD10	I00179 (HOXD10)	731	8
AAAG	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 () R08442 () R08443 ()	948	4
TGTTCT	PR	I00288 (PR)	868	6
YTATCW	T00267 GATA-1 T00304 GATA-1A T00306 GATA-1 T00307 GATA-2 T00308 GATA-2 T00309 GATA-2 T00310 GATA-3 T00311 GATA-3 T00312 GATA-3 T00313 GATA-3 T00314 GATA-3 T00567 NF-E1b T02312 GATA-1B	R02156 () R02157 () R02158 ()	116	6
CTTT	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 () R08442 () R08443 ()	44	4
GTTGCATC	T00608 NF-W1 T00609 NF-W2	R02220 () R02221 ()	996	8
TGGC	T00537 NF-1	R01681 ()	389	4
TGTTCT	T00335 GR	Q00076 (-)	868	6
TGTGTTT	T00360 HiNF-A	R00658 ()	9	7
GCCGC	T00788 T-Ag	Q00168 (-)	897	5
ATCACA	T00333 GR	R01814 ()	628	6
CTTT	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 () R08442 () R08443 ()	341	4
CCAAT	T00081 alpha-CBF T00100 CUTL1 T00150 CP1 T00151 CP2 T00153 alpha-CP1 T00156 alpha-CP2a T00388 H1TF2 T00426 alpha-IRP T00538 NF-1 T00556 NF-E T00764 SRF T00832 TGGCA-binding protein T01202 CP1 T01214 NF-E T01995 Clox T02042 Cutl1 T02043 CDP2 alpha-CP2b	R00039 () R00510 () R00532 () R00561 () R00562 () R00564 () R00571 () R00572 () R00660 () R01543 () R02848	395	5
CACGAG	T01649 HES-1	R04470 ()	767	6
AAAG	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 R08442 R08443	1048	4
CTTTCC	T00539 NF-1	R00802 ()	218	6
TTATC	T00302 GAL4 T00794 TBP	R00494 () R03182 ()	240	5
GATTTC	T00391 H4TF-1	R00680 ()	970	6
CATCAG	HOXA5	I00184 (HOXA5)	258	6
CCCACC	T00759 Sp1	R08207 ()	804	6
CCACC	CAC-binding	I00019 (CAC-binding)	725	5
GCCTC	T-Ag	I00215 (T-Ag)	279	5
TTATCT	T00267 GATA-1	Q00060 (-)	240	6
TCGC	T03218 E2F+p107	R08845 ()	198	4
ATTGGCT	T00536 NF-1	R01443 ()	1254	7
MAMAG	T00803 TCF-2alpha T00930 LEF-1 T00999 TCF-1A T01000 TCF-1B T01001 TCF-1C T01002 TCF-1	R02248 ()	538	5
AAGTCA	T00321 GCN4	R00831 ()	949	6
CGTNNNNNNNGA	T00056 ABF1	R01923 ()	201	12
TATCG	T04321 BEAF-32A T04322 BEAF-32B	R09849 ()	494	5
AAGATAG	R2	I00041 (R2)	1117	7
TTATCT	GATA-1	I00109 (GATA-1)	116	6
TTCTTGCG	SEF1	I00354 (SEF1)	1233	8
TGGC	T00537 NF-1	R01681 ()	438	4
GGACC	H4TF2	I00180 (H4TF2)	617	5
TCTTTCC	GT-IIA	I00172 (GT-IIA)	217	7
TATA	T00061 B factor T00286 factor i T00557 NF-E2	R00689 () R01129 () R04293 ()	74	4

Motif	Factor	Model	Pos.	Length
TCTTCT	T00335 GR	R01313 ()	47	6
CTGTTCT	T00336 GR	R01659 ()	867	7
CTTT	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 () R08442 () R08443 ()	292	4
CCTGG	LBP-1	I00191 (LBP-1)	485	5
CCACC	T00076 CAC-binding protein	R04295 ()	1197	5
CTTGCC	NF1	I00296 (NF1)	858	6
GCNCTNNAG	T00856 TTF-1 T00857 TTF-1 T00858 TTF-1 T00859 TTF-1 T02098 TTF-1	R04645 ()	579	9
CATACCG	IHF	I00375 (IHF)	1025	7
CTTT	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 () R08442 () R08443 ()	556	4
GGGCA	T00261 ER-alpha T00467 LF-A1 T00759 Sp1	R01171 () R04883 ()	944	5
TCCA	T00537 NF-1	R01681 ()	1086	4
CCACC	CAC-binding	I00019 (CAC-binding)	805	5
TTATCT	T00305 GATA-1 T01302 GATA-2	R03884 ()	240	6
TGACC	T00261 ER-alpha	R01194 ()	500	5
GACAG	T01214 NF-E	R00558 ()	1106	5
CCGCCC	Sp1	I00295 (Sp1)	1193	6
GCAGG	LVc	I00085 (LVc)	614	5
CCCNCNNNCT	T00913 Yi	R03154 ()	267	10
GCTCCTGC	T00843 Ttk 69K	R02057 ()	170	8
CTATCC	T00305 GATA-1	R08167 ()	1189	6
CCAAT	CP1	I00066 (CP1)	395	5
CGCTCCC	T00301 GAGA factor	R02065 ()	800	7
TCCA	T00537 NF-1	R01681 ()	1147	4
TGACC	ER	I00276 (ER)	500	5
TGGGC	T00788 T-Ag	R01372 ()	943	5
GCTGACA	NF-S	I00202 (NF-S)	909	7
TGCCC	T00261 ER-alpha T00467 LF-A1 T00759 Sp1	R01171 () R04883 ()	860	5
WWWCCACA	T00039 AP-3 (2)	R02122 ()	342	8
YTTCCT	T00114 c-Ets-1 54 T00115 c-Ets-1 68 T00116 c-Ets-2 58-64 T00684 PEA3 T00685 PEA3 T00686 PEA3	R02232 ()	149	6
CTTGGC	T00601 NF-1 (-like proteins)	R01192 ()	592	6
ATTGG	CBP/CRF	I00135 (CBP/CRF)	36	5
TCCA	T00537 NF-1	R01681 ()	963	4
STRATG	T04362 IPF1	R04228 ()	930	6
CCAAT	T00082 CBF (1) T00083 CBF (2) T00087 CBF-A T00088 CBF-B T00092 CCAAT-binding factor T00099 CDF T00108 C/EBPalpha T00170 CRF T00174 CTF T00613 NF-Y T00614 NF-Y'	R00231 () R00232 () R00335 () R00668 () R00669 () R00761 () R01081 () R01445 () R01446 ()	395	5
CGCGAAA	T00096 CCBF T00775 SWI4 T01013 SWI6	R03735 ()	527	7
ACAGC	GT-IIBa	I00160 (GT-IIBa)	957	5
TTATA	T00182 DBF4 T00270 ETF T00530 NC1 T00794 TBP T00798 TBP T00817 TFIIA T00818 TFIIB T00820 TFIID T00835 TMF T00862 UBP-1 T02216 TFIIA-alpha/beta precursor (majorT02216 T02217 TFIIA-alpha/beta precursor (minorT02217 T02224 TFIIA-gamma	R00046 () R00705 () R00706 ()	78	5
GAGGAA	T00702 PU.1	R04413 ()	684	6
CTACCGGT	BUF	I00303 (BUF)	517	8
TGCAAAG	T02878 TCF-4E	R08645 ()	638	7
AGCCT	T00302 GAL4	R00495 ()	278	5
GCAGGA	Ttk	I00261 (Ttk)	614	6
AGGACA	T00333 GR	R00429 ()	954	6
TGACTCC	AP-1	I00270 (AP-1)	84	7
GGCTCAGCGCG	T02855 CDC5	M00361 (P$CDC5_01)	920	11
ATTGG	T00081 alpha-CBF T00100 CUTL1 T00151 CP2 T00153 alpha-CP1 T00156 alpha-CP2a T00388 H1TF2 T00426 alpha-IRP T00538 NF-1 T00556 NF-E T00764 SRF T00832 TGGCA-binding protein T01202 CP1 T01214 NF-E T01995 Clox T02042 Cutl1 T02043 CDP2 alpha-CP2b	R00039 () R00510 () R00532 () R00561 () R00562 () R00571 () R00572 () R00660 () R01543 () R02848 ()	36	5

Motif	Factor	Model	Pos.	Length
CTTT	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 () R08442 () R08443 ()	55	4
GAGGC	T-Ag	I00215 (T-Ag)	585	5
GCGA	T03218 E2F+p107	R08845 ()	826	4
GCCTC	T00788 T-Ag	R01244 ()	759	5
GATAA	T01274 ABF2	R03826 ()	884	5
MTTNCNNMA	T00017 C/EBPbeta T00459 C/EBPbeta T00581 C/EBPbeta T00583 C/EBPdelta	R02216 ()	341	9
ATTGGCT	T00150 CP1 T00174 CTF T00536 NF-1	R01443 ()	36	7
ATTGG	T00082 CBF (1) T00083 CBF (2) T00087 CBF-A T00088 CBF-B T00092 CCAAT-binding factor T00099 CDF T00108 C/EBPalpha T00170 CRF T00613 NF-Y T00614 NF-Y'	R00231 () R00232 () R00335 () R00668 () R00669 () R01081 () R01445 () R01446 ()	36	5
CCCC	T00302 GAL4	R00496 ()	503	4
CCCTC	T00172 CTCF	R02137 ()	505	5
CCTGC	T00478 LVc	R01644 ()	351	5
CACCC	T00074 gammaCAC1 T00075 gammaCAC2 T00077 CACCC-binding factor	R00559 () R01474 ()	348	5
CGCGAAA	E4	I00126 (E4)	527	7
TCCA	T00537 NF-1	R01681 ()	248	4
ACGATA	T00627 NIT2	R02729 ()	882	6
TATTT	T00915 YY1	R00603 ()	471	5
GAGGC	T-Ag	I00215 (T-Ag)	750	5
AGGACA	T3R-alpha	I00033 (T3R-alpha)	954	6
CCGA	T00714 RAF	R00256 ()	373	4
TCCTC	T00302 GAL4	R00492 ()	1083	5
CCCTC	T00172 CTCF	R02137 ()	267	5
GCCCA	T00788 T-Ag	R01372 ()	465	5
TCAGTCTT	V$CAP_01	M00253 (V$CAP_01)	110	8
TGCCC	T00261 ER-alpha T00467 LF-A1 T00759 Sp1	R01171 () R04883 ()	1034	5
TCCA	T00537 NF-1	R01681 ()	344	4
MAMAG	T00803 TCF-2alpha T00930 LEF-1 T00999 TCF-1A T01000 TCF-1B T01001 TCF-1C T01002 TCF-1 T01109 TCF-1(P) T01979 TCF-1E T01981 TCF-1F T01982 TCF-1G T02905 LEF-1	R02248 ()	640	5
TTTCCACA	AP-3	I00150 (AP-3)	342	8
TTTTGTT	T02878 TCF-4E	R08644 ()	13	7
CACCTG	E12	I00081 (E12)	1198	6
ANATGG	T00865 YY1	R02133 ()	386	6
TTTGCC	T00033 AP-2alpha T00035 AP-2alphaA	R08502 ()	1032	6
TGACTC	GCN4	I00312 (GCN4)	84	6
TTATC	T01274 ABF2	R03826 ()	116	5
ATGACC	RC2	I00329 (RC2)	499	6
TCTCC	T00011 ADR1	R00074 ()	58	5
GCCCC	T00788 T-Ag	R01372 ()	1269	5
TCCTC	T00302 GAL4	R00492 ()	270	5
TTATCT	T00305 GATA-1 T01302 GATA-2	R03884 ()	116	6
CATTA	T00915 YY1	R00603 ()	1156	5
TGTGAT	T00333 GR	R01814 ()	93	6
TATCTA	T00627 NIT2	R02728 ()	117	6
CWWWCCAC	T00104 C/EBPalpha T00105 C/EBPalpha T00107 C/EBPalpha T00108 C/EBPalpha	R02132 ()	341	8
CAGAG	T00333 GR	R01813 ()	1108	5
CTTT	T01059 MNB1a T02690 Dof2 T02691 Dof3 T02692 PBF	R08440 () R08441 () R08442 () R08443 ()	218	4
TTATA	T00182 DBF4 T00270 ETF T00530 NC1 T00794 TBP T00798 TBP T00817 TFIIA T00818 TFIIB T00820 TFIID T00835 TMF T00862 UBP-1 T02216 TFIIA-alpha/beta precursor (majorT02216 T02217 TFIIA-alpha/beta precursor (minorT02217 T02224 TFIIA-gamma	R00046 () R00705 () R00706 ()	73	5
CCGA	T00714 RAF	R00256 ()	664	4

Motif	Factor	Model	Pos.	Length
TATA	T00061 B factor T00286 factor i T00557 NF-E2	R00689 () R01129 () R04293 ()	79	4
YNSNRNSTNGCGTGNNW	T00018 AhR	R02656 ()	191	17
CATAAA	Cad	I00232 (Cad)	731	6
YTATCW	T00267 GATA-1 T00304 GATA-1A T00306 GATA-1 T00307 GATA-2 T00308 GATA-2 T00309 GATA-2 T00310 GATA-3 T00311 GATA-3 T00312 GATA-3 T00313 GATA-3 T00314 GATA-3 T00567 NF-E1b T02312 GATA-1B	R02156 () R02157 () R02158 ()	240	6
CTATC	T01214 NF-E	R00554 ()	493	5
GCTGACR	T00605 NF-S	R02219 ()	909	7

Appendix 5: Extended Abstract For The 'Protein Expression Europe' Conference – September 2007

Development of *Ophiostoma floccosum* and *Ophiostoma piliferum* as hosts for Recombinant Protein Expression.

L. M. Robson., R. L. Farrell
University of Waikato, Hamilton, New Zealand

Ophiostoma floccosum and *Ophiostoma piliferum* are dimorphic ascomycete fungi found throughout the world colonising timber. Both species are important economically as they are known to cause discoloration of wood thus reducing its aesthetic value and subsequently price. Albino variants of the two species are used as biological control agents to prevent sapstaining and have been used commercially for the past 15 years to reduce pitch/wood extractives in paper manufacturing. Extracellular xylanases and lipases were characterised from these species and have been demonstrated to have post translational modifications. Due to the capability of these *Ophiostoma* species to be fermented and their biotechnology applications, they are particularly suitable as hosts capable of excreting extracellular recombinant proteins. Our research is aimed at developing the two species as hosts for protein expression for both biotechnological application and to better understand protein expression in these organisms. Recombinant strains of both species have been produced using various transformation methods. To date, vectors used in these transformations have contained heterologous promoters and termination sequences. To optimise the efficiency of these systems we have sought to identify promoters in both species that may be applied to a vector system. Using expressed tag sequence analysis (EST's), abundantly transcribed proteins and proteins of interest have been identified. Promoters and various transcriptional elements such as secretion signals have been elucidated from selected proteins using chromosome walking methods. Current efforts are focussed on incorporating these host transcriptional elements into select vectors. The paper will focus on describing the transformation system used and vectors, and description of both quantity and quality of recombinant proteins produced by the *Ophiostoma* species.

Why this paper should be considered.

The use of *Ophiostoma floccosum* and *Ophiostoma piliferum* as host expression systems has not previously been described. Little is known about protein production and biochemistry in these two species despite their economic significance, on a global scale, as members of a sapstaining group of fungi. It is only in the last two years that there has been a drive to further understand protein transcription in this genus with over 16,000 EST sequences being deposted onto NCBI compared to no deposits from this genus two years ago. The development of these unique eukaryotic hosts for protein expression contributes greatly to the knowledge of protein production in these fungi and has wider applications, both in their current biotechnological applications and as sapstain causing species.