DIFFERENTIAL EXPRESSION OF GENES ENCODING SECRETED PROTEINS IN *PENICILLIUM MARNEFFEI*

by

Suzie Rezenom

Submitted in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in the

Biological Sciences

Program

YOUNGSTOWN STATE UNIVERSITY

May, 2013

DIFFERENTIAL EXPRESSION OF GENES ENCODING SECRETED PROTEINS IN *PENICILLIUM MARNEFFEI*

Suzie Rezenom

I hereby release this thesis to the public. I understand that this thesis will be made available from the OhioLINK ETD Center and the Maag Library Circulation Desk for public access. I also authorize the University or other individuals to make copies of this thesis as needed for scholarly research.

Signature:

_____

Suzie Rezenom, Student Date

Approvals:

_____

Dr. Chester R Cooper, *Jr*. Thesis Advisor Date

_____

Dr. Xiangjia Min, Committee Member Date

_____

Dr. Jonathan Caguiat, Committee Member Date

_____

Dr. Brian Depoy, Dean of School of Graduate Studies & Research Date

ABSTRACT

Penicillium marneffei is a dimorphic pathogenic fungus endemic to Southeast Asia. It primarly infects individuals with compromised immunity, particularly those with AIDS. This fungus undergoes a dimorphic switch when shifted from growth at 25ºC to 37ºC. At 25ºC, *P. marneffei* grows as a mycelium, whereas at 37ºC this fungus develops as a single-celled yeast that reproduces by fission. The latter type of morphogenesis is associated with pathogenesis. Presumably, one or more of a subset of 37ºC induced genes that encode secreted proteins may play roles in pathogenesis.

Fungal secreted proteins play an important biological role. Studies have revealed that some fungal secreted proteins are involved in pathogenesis. Therefore, uniquely secreted proteins in *Penicillium marneffei* have been putatively identified by computational analysis. Those secreted protein-encoding genes expressed solely at 37ºC are likely candidates to have roles in pathogenesis. Hence, determining differential expression of these genes and by extension the secreted proteins they encode, may provide an indication in the mechanism of virulence of *P. marneffei* as well as serve as possible drug targets.

Using methods employing a reverse-transcription polymerase chain reaction (RT-PCR) and its related process, quantitative polymerase chain reaction (qPCR), the expression of genes encoding each secreted protein was assessed at 25ºC(mycelia phase) and at 37ºC (yeast phase). The resulting data demonstrated that there are two genes tentatively designated HP1 and HYD, respectively, within

the subsets studied that are induced solely during the yeast phase. Conversely, two other genes, tentatively designated HP2 and HP5 are abundantly expressed in the mycelial phase. Interestingly, HYD is similar to a homolog found in *Aspergillus orzaye*. The gene in the latter fungus has been found to absorb to hydrophobic surfaces. This suggests that HYD may have a role in attachment during pathogenesis of *P. marneffei*. This observation may aid to better understand the mechanism by which phagocytes are parasitized by unopsonized *P. marneffei*.

ACKNOLEDGMENTS

I dedicated this thesis in the loving memory of my father Haile Rezenom and my mother Tereza Branca. I would like to thank you for all the sacrifices you made to make my education possible. I also would like give thanks to my siblings, Yohannes Rezenom, Seare Rezenom and Elsa Rezenom, for their love and encouragement.

I would like to acknowledge the advice and guidance of my advisor Dr. Chet Cooper. I also thank Dr. Lorna Gallagher for sharing her knowledge and assistance that contributed to this study's success. I would also like to thank Dr. Jonathan Caguiat and Dr. Xiangjia Min for their suggestions and assistance. Last but not least I would like to give thanks to my husband Aaron for his support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Penicillium marneffei is a virulent fungus that causes a systemic infection in humans. This fungus can infect both immunocompetent and immunocompromised individuals (Vanittanakom et al., 2006). This disease is common in Southeast Asia, mostly among people possessing a compromised immune system. It was in 1956 when this species was first observed in bamboo rats in Vietnam (Segretain, 1959). While not confirmed, it is suspected that these bamboo rats are the transient sentinels of this disease fungus(Cooper and Haycocks, 2000).

Penicilliosis caused by *P. marneffei* rarely occurs in non-AIDS patients. The first cases of penicilliosis associated with AIDS were discovered in 1988 among individuals infected with HIV (Piehl et al., 1988). Subsequently, penicilliosis in humans has become significant as the number of people infected with the HIV has escalated. The disease agent is now recognized as an AIDS indicator pathogen. The majority of *P. marneffei* cases have been reported in Thailand (Cooper and Haycocks, 2000).

The origin of *P. marneffei* has not been yet identified despite many environmental studies (Cao et al., 2011). Scientists have investigated the possibility that the bamboo rat may be the original source of human penicilliosis infection. Studies analyzed shared characteristics of the genotypes of *P. marneffei* isolates found in humans in different regions. The challenge was to determine whether people were infected from the same unknown source or if the rats acted as

carriers for the infection, dispersing it among humans (Cao et al., 2011). The study was unable to establish if the rats were the source of the infection, and hence the origin of this fungus species remains unknown.

The life cycle of *P. marneffei* can be categorized into three stages: the filamentous vegetative growth at 25°C, asexual reproduction (conidiation) at 25°C, and unicellular yeast-like vegetative growth at 37°C (Fig. 1; Andrianopoulos, 2002). At 25°C, *P. marneffei* conidia germinate to produce a germ tube. The germ tube grows apically into a hypha. Subsequently, the hyphae gives rise to multicellular forms termed conidiophores that bear conidia. The multinucleate aerial cells grow away from the mycelium. Septae can produce secondary stalk cells, called rama. Tips of these cells give rise to metulae and phialides. These are uninucleate produced by budding division. The metulae bud first forms the stalk cells, and then it buds to form phialides from which conidia arise. From this process a chain of spores is attained that in maturation appear to have a pale green color (Andrianopoulos, 2002).

Penicillium marneffei undergoes yeast morphogenesis at 37°C, producing elongated yeast cells (Chan et al., 1990). Morphogenesis takes place 48 hr after germination, along with coupling of cell and nuclear division cycles, giving rise to hyphal compartments. These cells are shorter than the vegetative hyphal cells grown at 25°C. They then undergo arthroconidiation, a process in which the hyphal cells are separated along with the degradation of the cell wall material between the septae which gives uninucleated single cells. These cells are unique to *P. marneffei*

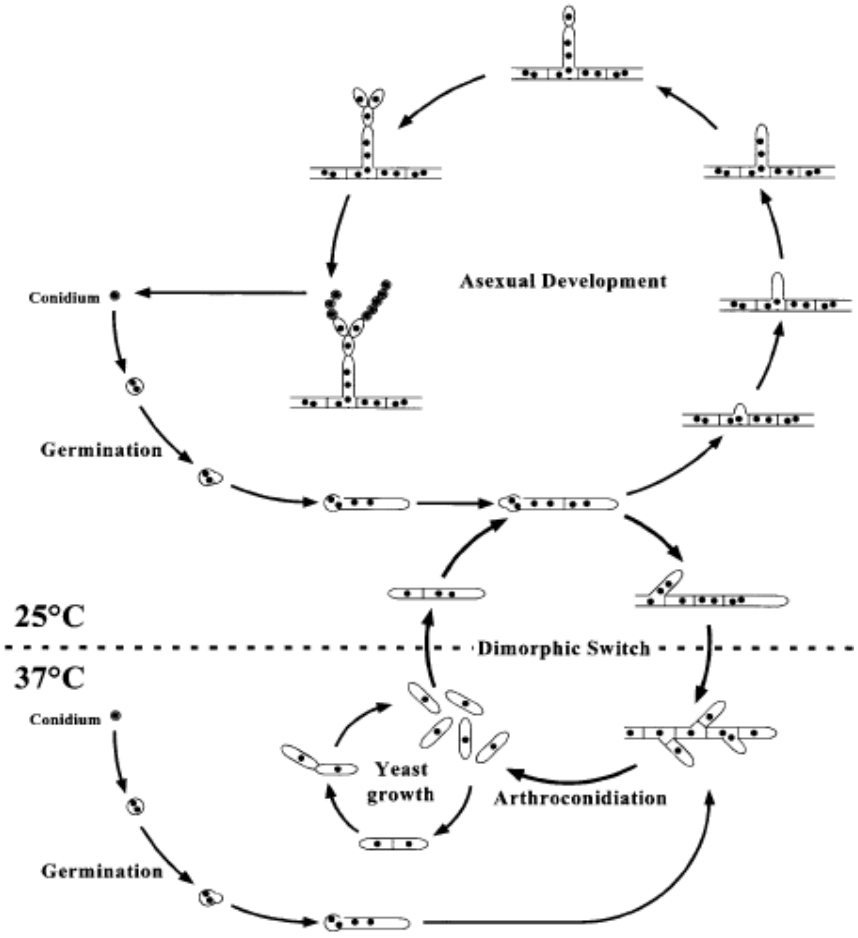

Figure 1: *P. marneffei* life cycle. (Andrianopoulos, 2002)

because they divide by fission unlike the other dimorphic fungi (Andrianopoulos, 2002; Cooper and Vanittanakom, 2008).

Mechanism of pathogenecity

Penicillium marneffei infections can be fatal if not treated. Penicilliosis infection from *P. marneffei* may affect skin, spleen, kidneys, lungs, bone marrow, liver and soft tissues (Supparatpinyo et al., 1994). In individuals already infected with AIDS, anemia, skin lesions, fever and weight are also observed (Ungpakorn, 2000).

Infection from this fungus is believed to be initiated by the inhalation of conidia from the air. These conidia are small enough that they can easily reach the alveoli of the host (Thirach et al., 2008). These conidia are able to identify fibronectin and bind to laminin. This reaction plays a major role in the attachment of conidia to alveolar epithelia prior to phagocytosis (Vanittanakom et al., 2006). After inhalation the conidia are able to transform into yeast cells and proliferate inside the pulmonary alveolar macrophages. The molecular mechanism that the fungus has developed to fight the immune system remains unclear. The primary defenses against this pathogen are the phagocytic cells. The interaction between heat-killed yeast phase *P. marneffei* and leukocytes showed that macrophages bind and phagocytes the disease agent regardless of opsonization. Therefore survival inside the phagocytes is a vital in establishing infection (Vanittanakom et al., 2006).

Penicillium marneffei is able to resist the oxidative stress within these immune cells. The mechanism by which this occurs has also not yet been discovered. However, many other fungi have been known to survive inside phagocytes mostly by inhibition of reactive oxygen metabolites or by neutralization of inhibitory host metabolites. One study demonstrated acid phosphatase activity in *P. marneffei* (Youngchim et al., 1999). Production of this acid phosphatase may help lower the intracellular pH thus helping pathogen survival by inhibiting the phagocyte burst (Thirach et al., 2008).

Secreted Proteins

Secreted proteins are defined as the proteins found in the growth medium when cells are grown in vitro (Chaffin et al., 1998). These proteins travel through cell walls to the extracellular environment. As they pass through they may interact with proteins bound in the cell wall. Studies with fungi showed that secreted proteins are excreted from the cell through hyphae as they colonize a substrate. These proteins can then degrade the organic materials within the environment in order to provide nutrients to the cell (Vinck et al., 2011). In the case of *P. marneffei*, this organic material may consist of the cells from the lungs or kidney, thereby helping it to be pathogenic. These secreted proteins are presumably involved in invasion of the host organism and may also play a role in resistance of the fungi to an immunogenic response by the host.

Secreted proteins are synthesized on ribosomes that have association with endoplasmic reticulum (ER) membranes. The process begins by segregation of the extracellular protein into the lumen of the endoplasmic reticulum. This is ascertained by the information encompassed in the signal sequence, which is an N-terminal domain attached to the protein molecule. Signal sequences from various organisms have certain characteristics in common (Peberdy, 1994). They include 13-30 aminoacid sequences with a basic N-terminal region and a cleavage site. There have been several hypotheses on how a signal sequence operates. One proposal is that as the protein passes through the organelles of the secretory pathway, cleavage of the signal peptide from the nascent protein occurs by an endopeptidase.

Before the secretion of a protein several post-translational modifications occur. These modifications initiate in the ER and continue as the proteins get transported through the Golgi apparatus. There are three modifications that may occur to the protein. Proteolytic cleavage that removes the signal sequence, a folding process involved in the formation of a disulphide bond that helps the tertiary and quaternary forms of the protein, and glycosylation (Peberdy, 1994).

Many secreted proteins in fungi undergo glycosylation. Glycosylation is a process that involves attachment of oligosaccharides to Thr, Asp and Ser residues on the surface of the protein. Glycosylation can be described as N- linked or O-linked. N-linked glycosylation when oligosaccharides attach to Asp residues. Instead, when oligosaccharides attach to serine and threonine the structure is

termed O-linked. Information on the process of glycosylation is limited. However, analyses of this data indicate that N-linked oligosaccharides seem to be essential in giving the protein resistance and stability to environmental influences. Also the O-linked glycosylation is important for secretion (Peberdy, 1994).

Recent innovation in the area of bioinformatics has been effective in the prediction of which proteins in various organisms will be secreted. This is particularly true with protists (Min et al., 2010). The mechanism of virulence of *P. marneffei* remains unknown. Therefore, investigating such mechanism is of great importance because it will make great contribution in drug discovery and efficient treatment.

Fungal Secretome

Studies on fungal secretomes revealed that secreted proteins are associated with evading the immune system, presumably contributing to pathogenecity. The secretome of the filamentous airborne fungus *Aspergillus fumigatus* was analyzed for the identification of secreted proteins dispensable in virulence (Brakhage, 2005). The survey reported the identification of 64 proteins. Identification of *A. fumigatus* involved growth in minimal medium followed by extraction of cultures supernatants and the separation of proteins by two-dimensional gel electrophoresis.

This investigation contributed to discovering Asp-hemolysin under the assessment of all growth conditions as a secreted protein. Asp-hemolysin is a member of the protein family of aeroglysins which are believed to be involved in

fungal virulence (Wartenberg et al., 2011). The research also demonstrated that some of the secreted proteins were not essential for pathogenesis , but may have some other biological or molecular role in cellular process yet to be elucidated (Wartenberg et al., 2011).

Aspartyl proteinase two (Sap2), a secreted protein, has been indentified in *Candida albicans*. This human opportunistic pathogen secretes Sap2, which is recognized to be a virulence factor (Hube et al., 1997). It was demonstrated to be a virulence factor in an experimental model of canidiasis (De Bernardis et al., 1997). Furthermore, a truncated recombinant protein of the aspartic proteases family was created to develop an anticandidal vaccine. The novel recombinant proved to possess strong immunogenicity and antigenicity (Sandini et al., 2011). Therefore, the secreted Sap2 is proposed to have a potential in fighting vaginal canidiasis, an infection that affects a great number of women worldwide (Aoki et al., 2011).

A previous analysis of the *A. fumigates*, secretome identified several secreted proteins from culture filtrate (Kumar et al., 2011). Homology analysis of the identified proteins using BLAST search was carried out and 15 proteins were recognized to be immunogenic in humans. Such findings are crucial in distinguishing virulence factors associated with fungal disease.

Aspergillus fumigatus also is known to secrete a glutamic protease in acidic protein medium. This protease is involved in preventing Pep deficiency in *A. fumigatus* (Sriranganadane et al., 2011). Therefore, analysis of the fungal secretome

is necessary for their vital biological and industrial roles. Some are involved in developing disease in humans (Ranganathan et al., 2009), some are involved in carbon cycling (Lum and Min, 2011) and some in plant- fungi symbiosis.

RESEARCH GOALS

1. To identify expression of predicted secreted proteins in *Penicillium marneffei*.

2. To observe quantitative expression differences of these proteins between the mold and yeast phase of the *P. marneffei*.

HYPOTHESIS

Penicillium marneffei expresses a distinct subset of 37°C-induced genes that encode secreted proteins having a role in pathogenesis.

MATERIALS AND METHODS

Prediction of secreted proteins

Completely sequenced fungal genomes and new computational analysis tools, have allowed secretome identification and annotation possible. Several publications have reported on such applications. Lum and Min recently predicted several secreted proteins in *P. marneffei* using bioinformatics tools (Lum and Min, 2011). Fungal protein sequences were retrieved from NCBI RefSeq database and were assessed using the following programs: SignalP (version 3.0), TargetP, WolfPsort and Phobius for the prediction of signal peptide and subcellular location (Lum and Min, 2011). The above four predictors were chosen because they have been previously used by the fungal research community and have good evaluations. TMHMM, which is a program for identifying proteins that have transmembrane domains, was used and. PS-Scan was also used for scanning of the endoplasmic reticulum targeting sequences.

Default parameters for fungi were used with each program. The SignalP prediction was based upon entries that were considered to have had a most likely cleavage site by SignalP-NN algorithm and a signal peptide by SignalIP-HMM algorithm were evaluated to be signal peptide positives, using N-terminal 70 amino acids (Soanes et al., 2008). Prediction of membrane proteins used the program TMHMM and the entries that did not have domain membranes located within the N-terminus were considered as true membrane proteins. Sequences that were

predicted to have a signal peptide were then processed by another program FragAnchor in order to identify the glycosylphosphatidyinositol anchors. This implies that those protein sequences that are predicted to have GPI anchor might be attached outside the plasma membrane or might be secreted to be transported to the cell wall (The UniProt Consortium, 2009).

The hypothetical secreted proteins, unique to *P. marneffei*, used in this study were termed as: HP1 (GI: 2921742), HP2 (GI: 7025042), HP3 (GI: 7027740), HP4 (GI: 7027023), HP5 (GI: 7023066), HP6 (GI: 7026181), HYD (GI: 212525965), Carbonic anhydrase (GI: 212536242), Transcription regulator (GI: 212541226) and Thaumatin (GI: 212533347). For the purpose of this study, only the ones in which their function has not been annotated were abbreviated as HP (hypothetical protein).

P. marneffei Growth and Inocula

The *P. marneffei* strain used in this study F4: (CBS 119456),(Pongpom et al., 2005), was isolated originally from a HIV-positive Thai males. This strain was maintained on potato dextrose agar (PDA [Difco brand]; Becton Dickinson, Franklin Lakes, NJ) at room temperature.

For inocula preparation, the fungus was grown in 150 cm^2 cell culture flasks with vented caps containing PDA. Conidia and accompanying hyphae were collected by gently scraping the culture surface after flooding it with 10 ml of sterile water. The suspension was then passed through a screen cap (Bio-Rad, Hercules, CA)

separated with 1-inch glass wool (Corning, Acton, MA: Gifford and Cooper 2008). The screen cap (Bio-Rad) was secured with a 50-ml conical tube. The cell suspension was centrifuged at 150 x g for 30 sec. The concentrations of resulting conidial suspension were then determined using a hemocytometer.

Broth Inoculation

Four 500ml Erlenmeyer flasks containing 50ml Sabouraud dextrose broth (SAB; Becton and Dickson) were inoculated with the conidia harvested at a concentration of 1 x 10^7 conidia/ml. The inoculated SAB broth was cultured in a shaking water bath at 25°C and 37°C. The cells in the flasks were collected after 24 hours by centrifugation at 15,000 rpm at 4°C for 20 minutes. TE (Tris- EDTA) buffer was used for washing the mold and yeast cultures. The pellet weighed 200-400 mg.

RNA Extraction and Purification

The mycelial and yeast cells of *P. marneffei* collected as described above were fractured by bead-beating technique (Chandler et al., 2008). From this supernatant, RNA was isolated and purified using the RNeasy™ Plant Mini kit in accord with the manufacturer's instructions (Qiagen; Valencia, CA). The isolated RNA was DNase treated using a commercially available kit (Invitrogen; Carlsbad, CA). When not in use, the RNA solutions were stored at -80ºC.

cDNA synthesis

cDNA was synthesized using the Iscript kit (Bio-Rad). Each reaction contained 11 µl of RNA, 4 µl of 5X buffer, 2 µl of oligo dTs, 1 µl Superscript reverse

transcriptase, and 2 µl of RNase-free water. The reaction mix was incubated at 42°C for 1 hr. If not used immediately for either RT-PCR or qPCR the cDNA solution was stored at -80 ºC freezer.

Primer Design

Primers were designed with the online life technologies primer design. Parameters such as primer length, melting temperature and % GC content were manually optimized. The primer length selected was 18-24 base pairs and Tm of 55ºC-60ºC. The primers were also checked for specificity.

Reverse-Transcription Polymearse Chain Reaction (RT-PCR)

Template cDNA, primers of interest (forward and reverse), GoTaq mix were thawed and placed on ice. A master mix was prepared using the primers of interest (forward and reverse), cDNA template, GoTaq mix and nuclease free water for a total volume of 20 µl. This was mixed thoroughly and was dispensed in the proper amounts into PCR tubes and placed in the thermo cycle. cDNA was added right before the PCR tubes were placed on the thermo cycler. Thermal profile for RT-PCR was as follows: 95ºC for 3 minutes, followed by 35 repeats of 95ºC for 30 seconds, 62ºC for 30 seconds and 72ºC for 1 minute and final extension for 10 minutes at 72ºC then hold at 4ºC forever.

Quantitative-Polymerase chain reaction (qPCR)

PCR was performed using the iQ SYBR green supermix (Bio-Rad) on the Bio-Rad iQ5 Real-Time PCR System. A polymerase chain reaction (PCR) was carried

out on 96 well PCR plates in a reaction volume of 25µl containing 2.5µl of cDNA. For qPCR reactions, 250ng of cDNA from each sample was used as a template and each reaction was performed in duplicate on each plate, and this was conducted independently three times. In addition to the cDNA the reaction mixtures consisted of the following ingredients; 5 µl forward and 5 µl reverse primer, and 12.5 µl SYBR green mix, thus a total of 25µl was loaded in each tube. Calmodulin primers were also used to amplify a "house keeping" gene (accession no XM_002149230).

A serial dilution of the cDNA was prepared before being added to the tubes containing the reaction mixtures. Thermal profile for qPCR was as follows: 55° C for 10 minutes, 95° C for 5 minutes followed by 40 repeats of 95° C for 10 seconds, 60° C for 30 seconds with real time analysis and 81 repeats of 55° C for 1 minute with melt curve analysis.

Once the run was completed, the program analyzed the data by determining the standard curves with the display of reaction efficient, the coefficient of determination, the slope of the y-intercept and the line. Acceptable results should have a coefficient of determination to be close to 1, and error value of <2 were considered acceptable.

RESULTS

Using the bioinformatics selection procedure described in the methods section, 35 putative secreted proteins were identified. These were determined to be unique to *P. marneffei* as compared to *Penicillium chrysogenum* and *Talaromyces stiputatus*. The function of 23 was unknown, whereas the putative functions of the remaining 12 were identified. Therefore, a total of ten genes; five with unknown functions and five in which their functions had been annotated, were selected for gene expression analysis.

Primer Testing

Before performing any PCR experiments, primer testing was carried out. PCR experiments were performed for each primer pair that was designed, to each of the 10 putative secreted protein of *P. marneffei* (Table 1). All primers produced a product of the expected size (Table 2). Primers that did not show the desired PCR product size were re-designed. The successful primers were subsequently used to perform RT-PCR and qPCR.

Table 1. Primer sequences

Sequence Description	Sequence
PM HP 1 F	ACTTCAATGCCCGGATTCAGA
PM HP 1 R	TAGTAAAAGCAGGGCCAGCA
PM HP 2 F	TCTTCCTCTCGGCTCTAGGA
PM HP 2 R	AACTCATCGTCCGAATCAGC
PM HP 3 F	ACGAGTACCTGGATATTGCTGA
PM HP 3 R	GAGCCTTTCCCTCAAGGTCA
PM HP 4 F	ACGTCGATAAGCGTAAGGTCA
PM HP 4 R	AAGTGTTTGCAAAGCCAGTG
PM HP 5 F	CCTACTCTTACCTGGGGTGA
PM HP 5 R	GTAGACGTACTGGGAGCCAA
PM CARB F	GGATTTTGCTCTCGCTTTTG
PM CARB R	AGCGATTTGAGATGCTTCGT
PM HYDROPH F	GGTCAAGTTTGGCCAGGATA
PM HYDROPH R	GTGGCGCCAACAACTTTACT
PM HP 6 F	GCATTCTTCATACCCTGGTGA
PM HP 6 R	CATATGTGATGGAGACCCAGA
PM THAUMIN F	ATCCTCGTCACAAGGGACAC
PM THAUMIN R	GATTCATGCTTCCCGACATT
PM TRAN REG F	GTCAACCCGTCCTTTGAAGA
PM TRAN REG R	CGTAGAGGTGGTGCAGGAAG

Table 2. Product size

Putative Secreted Protein	Expected Primer Product size
HP1	343bp
HP2	200bp
HP3	320bp
HP4	450bp
HP5	500bp
HP6	500bp
Thaumatin	520bp
Carbonic anhydrase	440bp
Hydrophobic	500bp
Transcription regulator	480bp

Microscopy of *P. marneffei*

Penicillium marneffei cultures prepared in SAB were incubated for 24 hours at 25ºC and 37ºC. Cultures were observed microscopically and morphological differences were noted (Fig. 2). At 25ºC, hyphae appeared thinner and longer than those incubated at 37ºC. The latter were shorter and broader, presumably due to the transition to yeast at 37ºC.

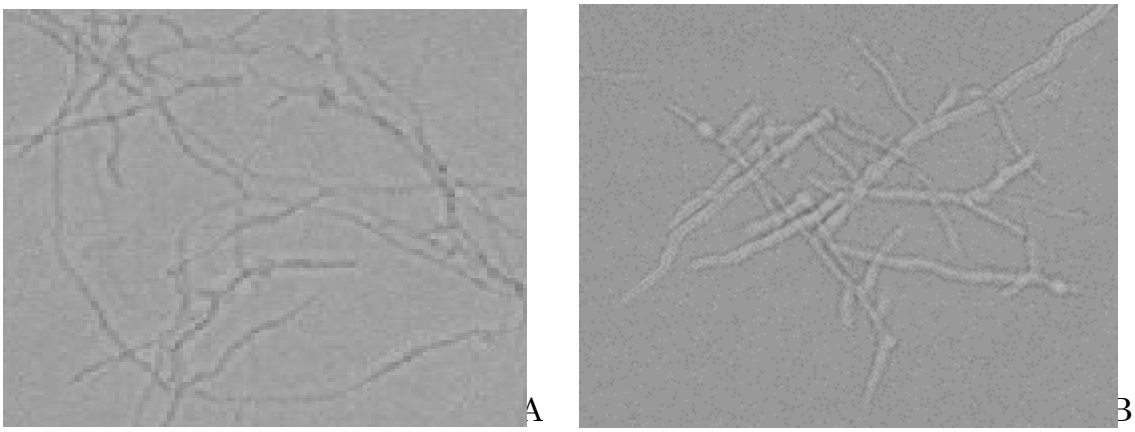

Fig. 2 Microscopic morphology of *Penicillium marneffei* cultures on Sabouraud's dextrose broth at 25°C (A) and at 37°C (B) for one day.

RT-PCR and Differential Gene expression

RT-PCR experiments were carried out using the primers designed with the genes of interest. Calmodulin primers were used to generate a PCR product that would serve as an internal control and to normalize expression levels. Calmodulin was used because it is calcium binding messenger protein expressed in all eukaryotes. Genomic DNA was also used as a positive control. Samples run with water in substitution of the cDNA were termed empty and used as a negative control. RT-PCR was run with all the ten genes with cDNA collected from SAB cultures at 24 hr and 48 hr. Samples were run on a 2% gel for analysis of gene expression.

Figure 3 displays the expression levels of the following genes after 24hr of incubation at 25ºC or 37ºC: carbonic anhydrase, hydrophobic, transcription regulator, HP6 and HP3. The resulting bands formed as expected at the expected respective base pair size. Carbonic anhydrase, transcription regulator and HP6 displayed expression in both mold and yeast forms of *P. marneffei*. Thus, no differential expression is noticed between the mold and yeast phase at 24 hr. In contrast, the hydrophobic surface binding protein (HYD) displayed expression in the yeast and not in the mold at 24hr (Fig3, lanes 5-6).

In addition, expression levels were detected at 24 hr in both the mold and yeast phase for HP5, HP2 and Thaumatin (Fig. 4). HP1 displayed double banding above the desired product size (343bp).

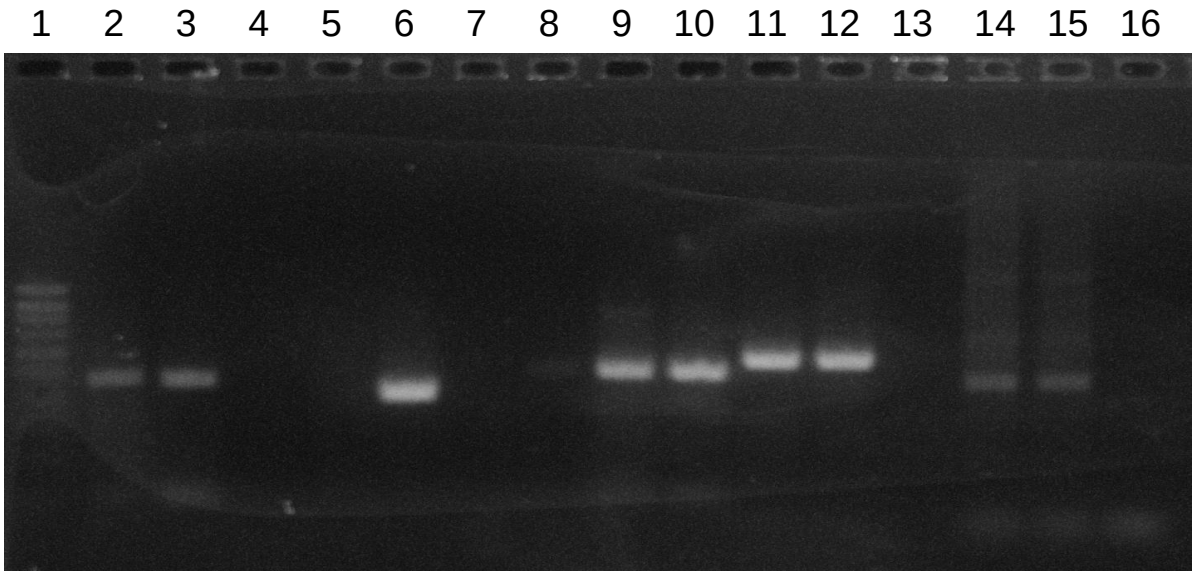

Fig. 3 RT-PCR analysis at 24hr from mold and yeast phase of *Penicillium marneffei*. Lane 1, 100bp ladder; Lanes 2,3 Carbonic anhydrase M, Y; lane 4 empty; lanes 5,6 HYD M,Y; lane 7 empty; lane 8 empty; lanes 9,10 Transcription regulator M, Y; lanes 11,12 HP 6 M,Y; lane 13 empty, lanes 14,15 HP 3 M,Y; lane 16 empty

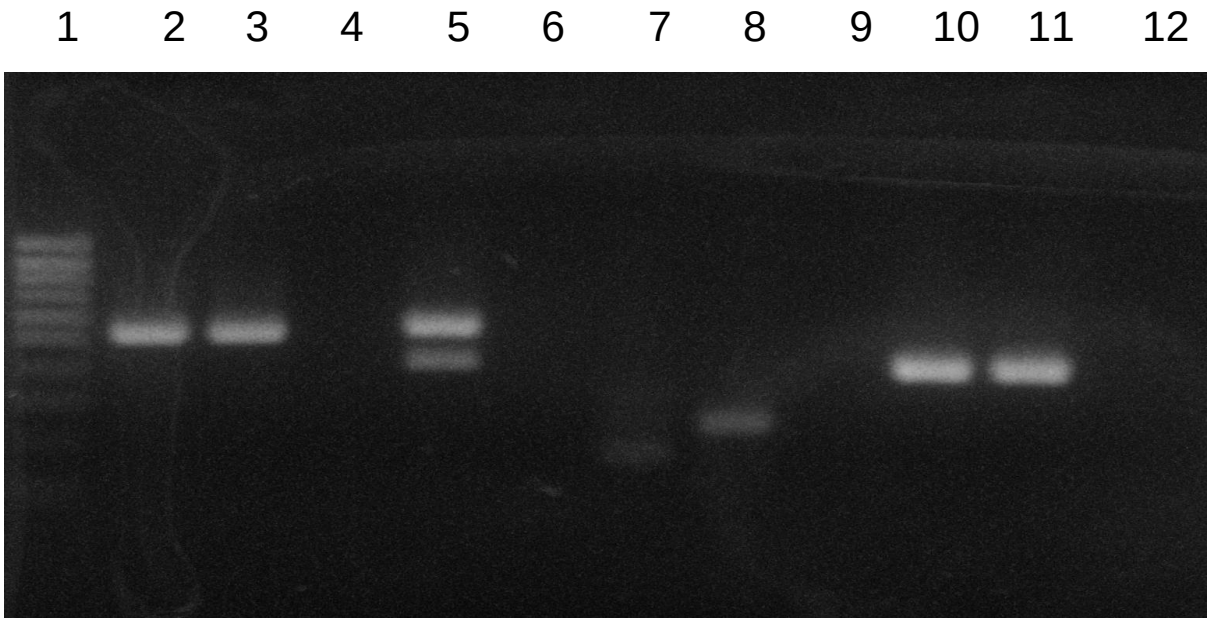

Fig. 4. RT- PCR analysis.
Lane 1, 100bp ladder; lanes 2-3 M, Y HP5; lane 4 empty; lane 5 HP1 yeast; lane 6 empty; lanes 7-8 M,Y HP2; lane 9 empty, lanes 10-11 Thaumatin M, Y; lane 12 empty

In order to eliminate the double band, primers for HP1 were re-designed. HP2 appeared to be differentially expressed as observed by the different band intensities at each temperature (lanes 7-8). Specific primers designed for HP4 failed to work repeatedly. Therefore, no data was derived for this gene.

Figure 5 displays the equal expression of both the carbonic anhydrase and transcription regulator in the mold and yeast form of *Penicillium marneffei* after incubation for 48 hr. HP6 is absent in both mold and yeast phases at 48 hr. Hydrophobic at 48hr displays expression at the yeast form only. This gene also displayed differential expression from samples collected at 24 hours.

In Figure 6, HP5 expression is detected from RNA collected after 48 hr of incubation at 25ºC. However, this gene though is not expressed at 37ºC (lane 7). Conversely, whereas Thaumatin was equally expressed in the mold and yeast forms at 24 hr, it is expressed at greater levels in the mold than the yeast forms at 37ºC (lanes 10 and 11).

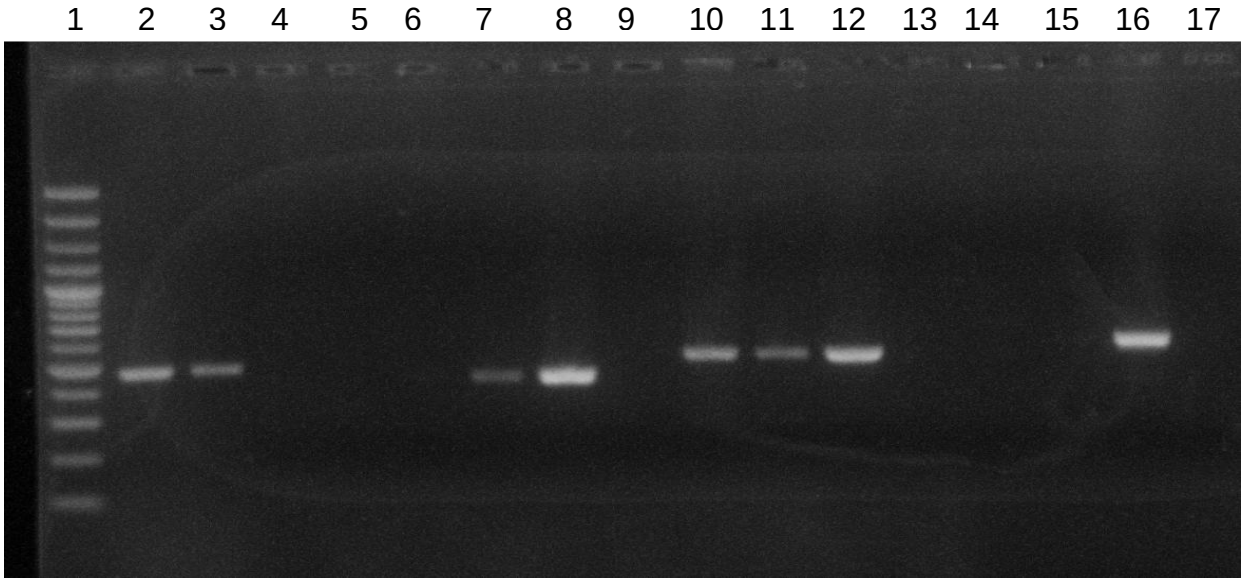

Fig. 5. RT-PCR analysis at 48 hours.
Lane1, 100bp ladder, lanes 2,3 carbonic anhydrase M,Y; lanes 3-4 empty; lanes 6,7 hydrophobic M,Y; lane 8 genomic; lane 9 empty; lanes 10,11 transregulator M,Y; lane 12 genomic; lane 13 empty, lane 13,14 hypothetical protein 6 M,Y; lane 15 genomic; lane 17 empty

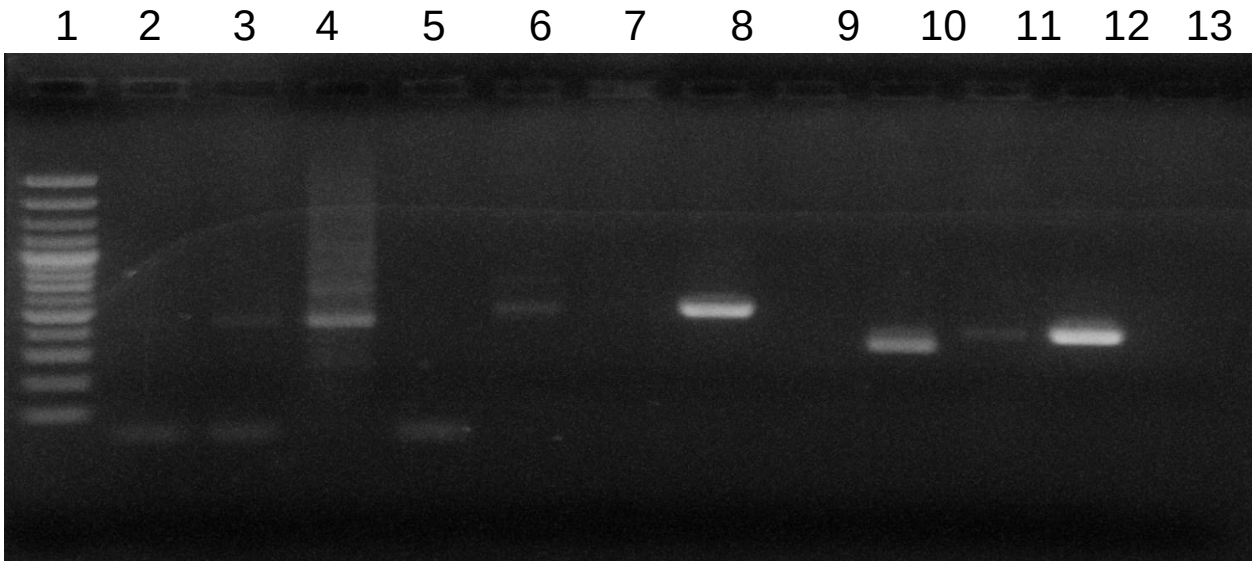

Fig. 6 RT-PCR analysis at 48 hours.
Lane 1 100 bp ladder, lanes 2,3, HP3 M,Y; lane 4 genomic; lane 5 empty; lanes 6, 7 HP5 M,Y; lane 8 genomic; lane 9 empty; lanes 10,11 Thaumatin M,Y; lane 12 genomic, lane 13 empty.

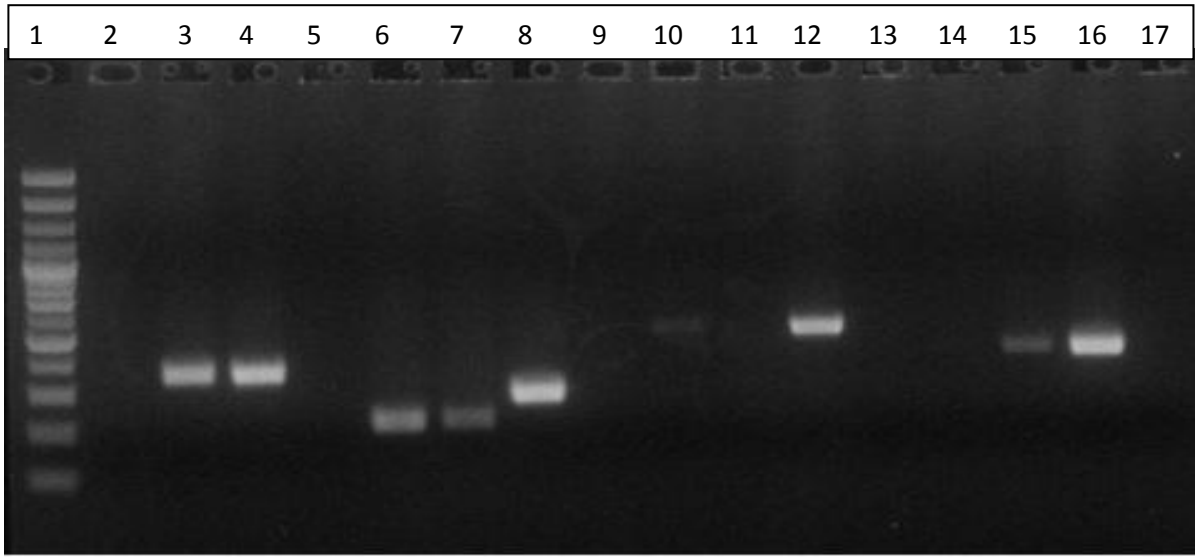

Fig.7. Differentially expressed genes between the mold and yeast phase in *P.marneffei*. Lane 1, 100bp ladder; lane 2-3 HP1 M,Y ;lane 3 genomic; lane 4, empty; lane 5-6 HP2 M,Y; lane 7 genomic; lane 8 empty; lane 9-10 HP5 M,Y; lane 11 genomic, lane 12 empty; lane 13-14 HYD M,Y; lane 15 genomic; lane 16, empty.

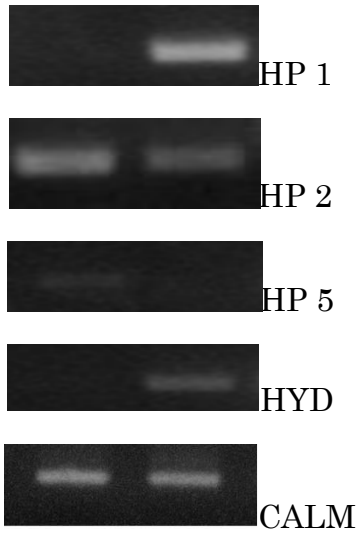

Fig. 8. RT-PCR products showing differential expression between 25ºC and 37ºC. CALM used for normalizing expression and also as an internal control.

qPCR

After performing RT-PCR on the ten genes selected, only four were picked for qPCR analysis. These genes included HP1, HP2, HP5. Unfortunately, the primers designed for HP5 and HYD were not successful for qPCR analysis. Therefore, further qPCR analysis was not performed on the latter genes.

Reverse transcription quantitative PCR (qPCR) was performed on HP1and HP2, using SYBR green I and iQ™ 5 optical system (Bio Rad). The software associated with this system analyzes standard curves, melting curves and normalized gene expression. Each gene was analyzed using 100ng of cDNA from mold and yeast phase at 24, 48 and 72 hours. Calmodulin was used as a control for each gene and for each condition.

Amplification data was analyzed to determine whether the increase in quantification was at the minimum of 3 cycles in range with each other. This analysis was followed by melting curve evaluation. Wells that displayed more than one peak were eliminated from the data. This ensured that the quantification analyzed is only from one gene. Standard curves evaluated were considered to determine efficiency. Standard curves for HP1, HP2 and CALM are shown on Figures 9, 10 and 11 respectively. All possessed R^2 value close to 1.

In contrast HP5 and HYD displayed low efficiency and an R^2 value of less than 0.9(data not shown). Although a standard curve with temperature gradient was created to determine the most suitable annealing temperature, the results

showed the primers to be unsuitable. Therefore HYD and HP5 were eliminated from this study.

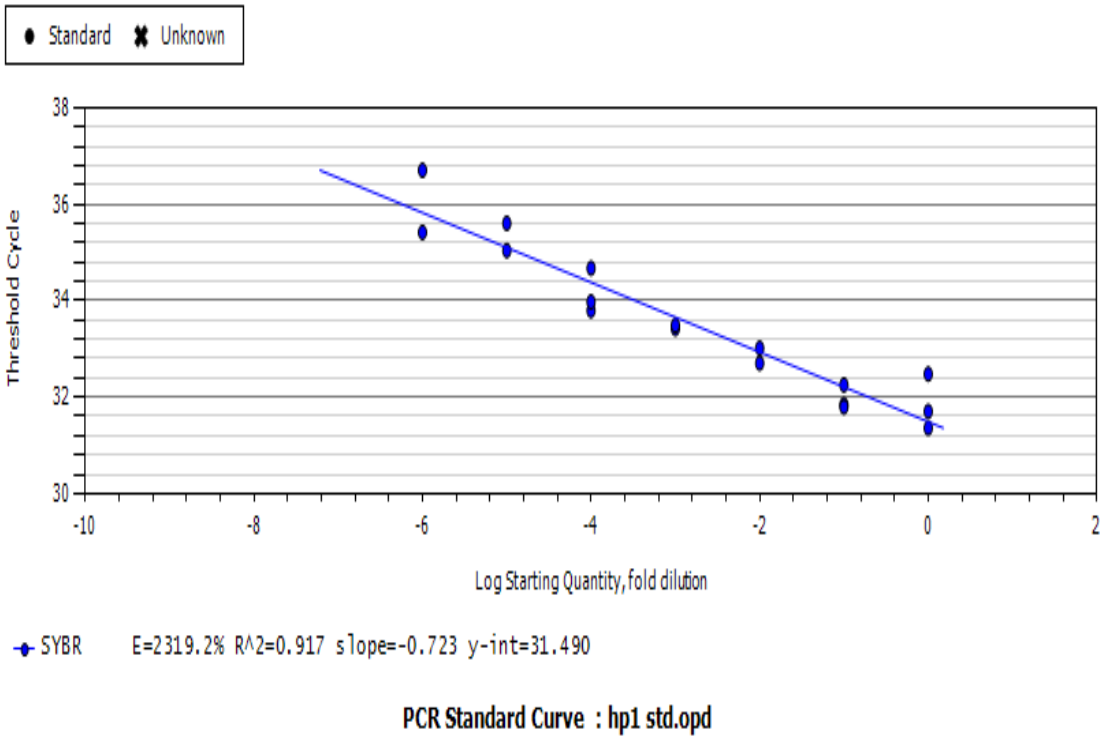

Fig. 9 Standard curve for HP1 showing R^2 at 0.917

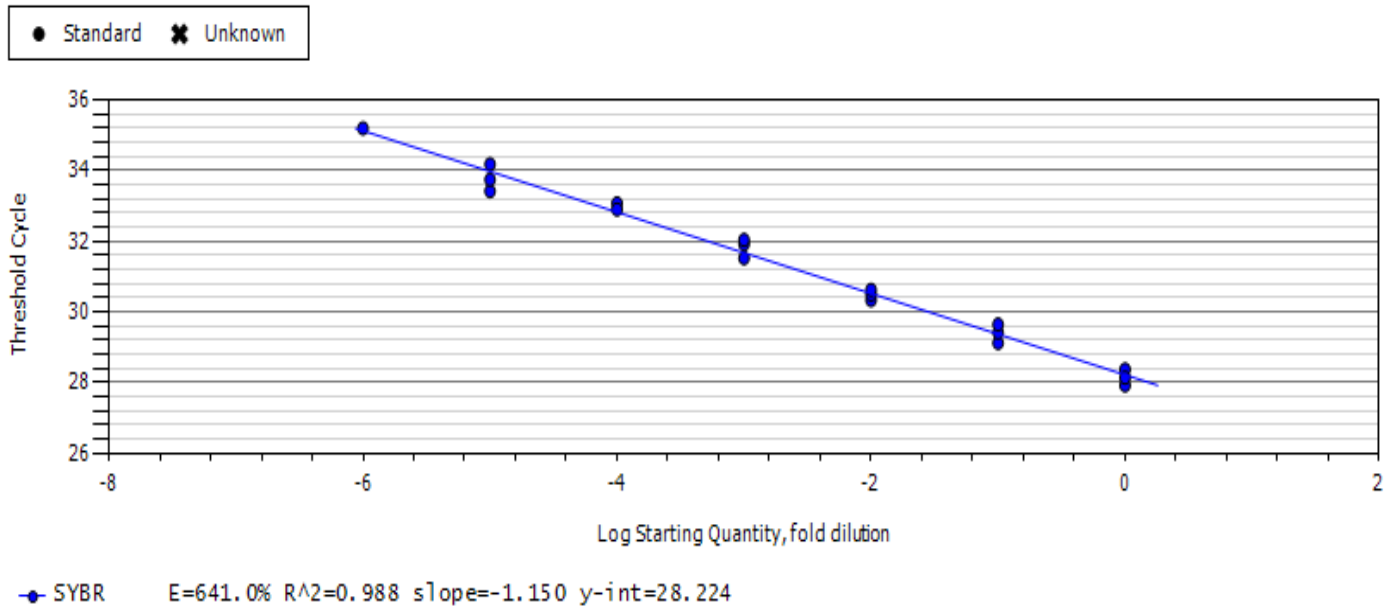

Fig. 10 Standard curve for HP2 showing R^2 AT 0.988

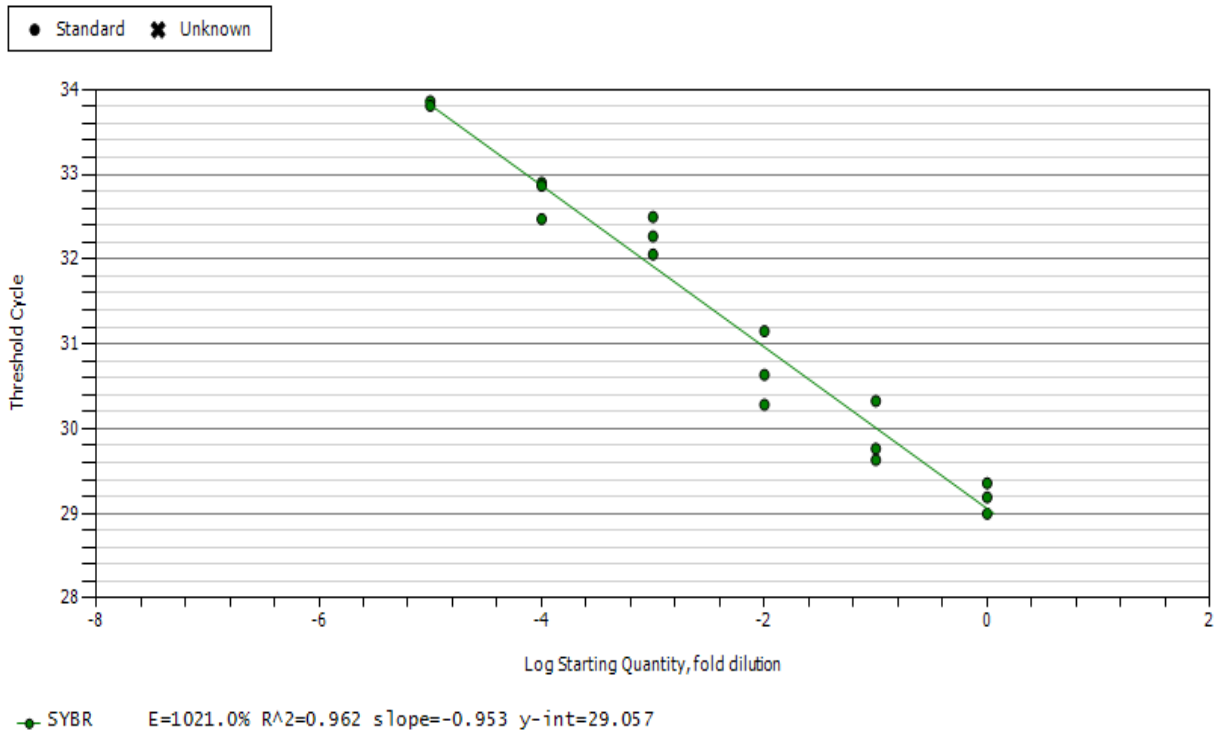

Fig. 11 Standard curve for CALM showing R^2 at 0.962

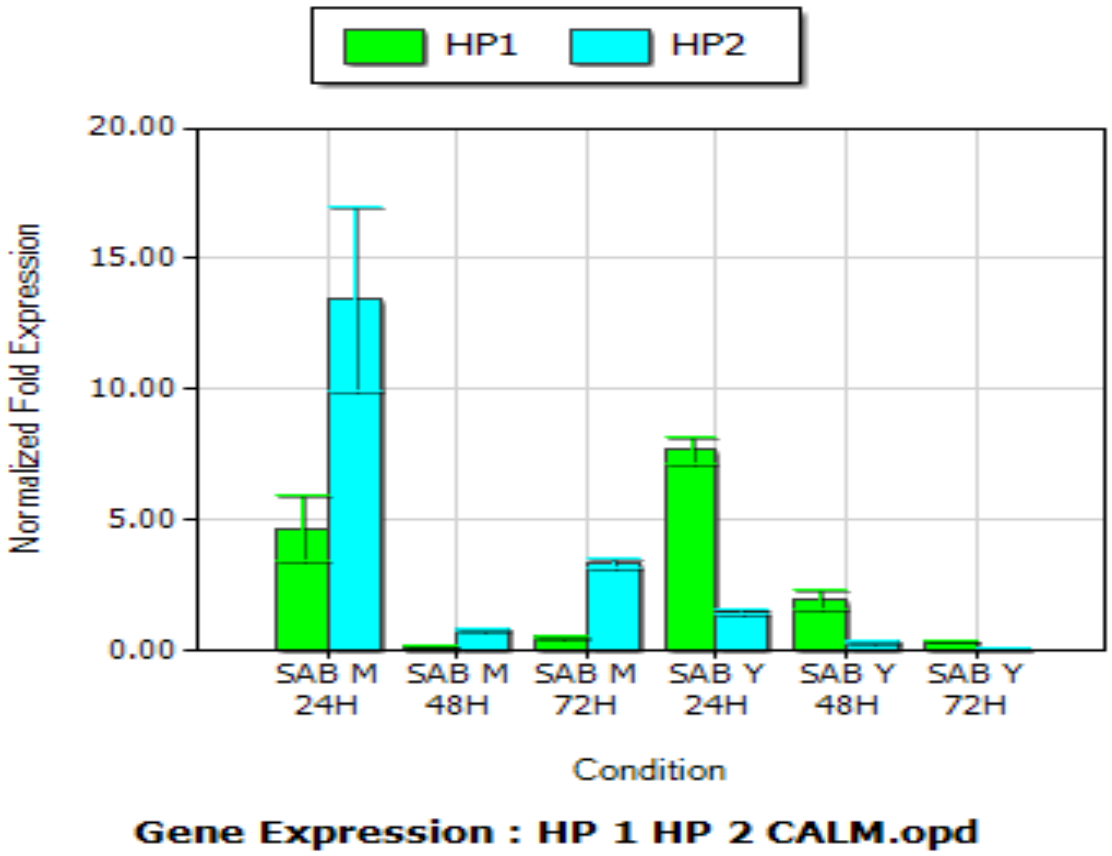

Fig 12. HP1 and HP2 normalized gene expression.

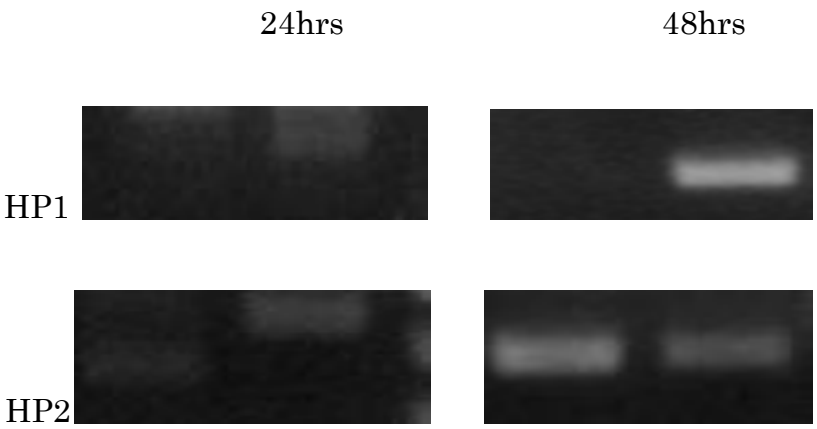

Fig 13. RT-PCR analysis for HP1 and HP2.

HP 1 and HP 2

Normalized fold expression levels were run for HP1 and HP2. Because the RT-PCR data detected intensity level of expression between the 24 hours and 48 hours, it was of interest to perform quantitative analysis. At 24 hr, it appears that there is some level of expression for HP1 mold but not in 48 hr or 72 hr (Fig. 12). This result is consistent with the RT-PCR data (Fig. 13). The level of expression for the yeast appears to be highest at 24 hr and decreasing at 72 hr.

HP 2 level of expression is more at the mold form rather than yeast form.

DISCUSSION

There are three dimorphic fungi that are pathogenic in the Unites States: *Histoplasma capsulatum, Coccidioides immitis and Blastomyces dermatitidis*. They share the same morphological characteristics with *P. marneffei*, which is endemic only to parts of Southeast Asia. Although, there is suggested information on how these dimorphic fungi infect humans, the exact mechanism of virulence is yet to be determined. Therefore, enhancing our understanding of the mechanism of dimorphism would clarify the modes of infection.

In addition, fungal secreted proteins have been reported to evade the immune system. Thereby, this study developed an approach to predict and characterize uniquely secreted proteins in *P. marneffei*'s secretome. Unique secreted proteins were predicted by a computational analysis as described in the methods and subjected for gene expression analysis.

Expression for the genes encoding secreted proteins was investigated in different developmental phases of *P. marneffei.* This study assessed ten predicted secreted proteins. Four of these showed differential gene expression between the yeast and mycelial phases of *P. marneffei* (Fig. 8), whereas four showed equal expression in both growth conditions. Valid data could not be determined for HP4 and HP3 because the PCR primers designed to these genes failed to amplify any DNA products at their appropriate annealing temperature.

HP1, HP2, HYD and HP5 displayed differential expression. Therefore these were selected for qPCR analysis. However, qPCR data is reported only for HP1 and HP2 due to a problem with primer annealing temperatures for HYD and HP5.

HP1 displayed differential expression at all growth phases according to RT-PCR. At 24 hr the mold phase displayed low transcript levels. However, at 48 hr it was absent. They yeast phase was expressed abundantly at both growth conditions. Upon further analysis with qPCR, there seem to be expression at 24 hr for the mold phase, but expression significantly dropped at 48 hr and remained constant for 72 hr. In contrast, transcripts levels for HP2 were observed to be twice as much at 24 hours in the yeast phase when compared to the mold at same time. Expression after 24 hr decreased, but remained at detectable levels. The large amount of transcript found in the yeast at 24hr suggests that HP1 may play a role in this pathogenic form but the protein does not seem to be necessary for the mycelial form at that given time. Furthermore, the HP1 gene investigated in this research is identical to the Mp1p gene recently characterized (Cao et al., 1999).

Mp1p is a cell wall mannoprotein, detected in culture supernatants of *P. marneffei* and also in sera of penicilliosis patients (Cao et al., 1999). It has also been demonstrated that anti-Mp1p antibody is specific to this dimorphic fungus because it failed to react with lysates of *Histoplasma capsulatum*, *Candida albicans*, or *Cryptococcus neoformans*.

A strategy employing an antibody screening approach, allowed isolation of a gene termed MPLP6(Pongpom and Vanittanakom, 2011). This gene encodes for a yeast specific antigenic mannoprotein in *P. marneffei*. This novel gene exhibits high similarity with Mp1p according to nucleotide and aminoacid sequence analysis. Results showed that there is high similarity of MPLP6 to the two halves of the Mp1p. The first half of Mp1p has 46% homology to a whole sequence of MPLP6, and the remaining part of the Mp1p gene has a 44% homology to the whole sequence of MPLP6 (Pongpom and Vanittanakom, 2011). This novel gene also shares similar expression characteristics with HP1. However, MPLP6 lacks a GPI anchoring sequence (Frieman et al., 2002) suggesting that it will not be localized in the cell wall. In contrast, HP1 (Mp1p) is located at the cell wall. HP1 and MPLP6, with the exception of the GPI anchoring sequence, both share certain structural features such as; a signal peptide-needed for signaling the transport of the protein to the cell surface, and a serine/threonine region (Frieman et al., 2002).

Studies have shown that polymorphism of HP1 (Mp1p) exists among clinical isolates indicating that HP1 (Mp1p) may be trivial. In contrast, sequence variation was not found among ten isolates in MPLP6, suggesting this gene being highly

conserved in *P . marneffei*. Therefore, more investigation is necessary to elucidate function for HP1 and MPLP6.

The expression in HP2 was higher in the mycelia phase at all conditions when compared to the yeast phase (Fig. 7). qPCR analysis of this gene showed significant early mold phase expression up to three times more than the yeast form at 24 hr. This finding may indicate that it is necessary for germination. Assessing conidia based expression of HP2 would aid in elucidating this suggestion. At 72 hr, transcript levels for HP2 were lower but still significant. It is observed that the large amount of transcript level at 24 hr in the mold phase followed by a drop at 48 hr, indicates it might be involved in the onset of conidiation. In contrast, expression in the yeast starts at very low levels at 24 hr and gradually lowers at 48 hr and is absent by 72 hr. It should be noted that this is a reverse trend in the HP1 gene. Therefore, this gene may not be called a yeast specific gene. This may somehow insinuate that HP2 is not as necessary in the later phases of the pathogenic form as might be HP1 (Fig. 8 and Fig. 12).

The expression of the HP5 gene was investigated in the mold and yeast phase at 24 hr and 48 hr. HP5 displayed differential expression between mold and yeast form of *P. marneffei* at 48hr. As shown in Fig. 8, the HP5 is absent in the yeast form (2 days old), suggesting that the function of HP5 is not necessary in this form. In contrast, HP5 at 24 hr displayed significant expression equally. There is no literature on this gene to explain any purported functions. Therefore, the switching

on and off characteristics noted at different time points in this study, may used as a concept in starting to understand the mechanisms of virulence.

RT-PCR analysis for HYD shows that this gene is induced only at 37ºC. Expression at 24 hr (Fig. 4) and 48 hr (Fig. 8) displayed yeast specific expression. The mold phase is completely absent, this can be indicating the HYD is playing a role in the virulent form and may not be required for germination.

A study showed the isolation of a secreted hydrophobic surface binding protein isolated from cultures of *Aspergillus orzaye* (Ohtaki et al., 2006). Looking at the chemistry of this hydrophobic surface protein, it comprises the amino acid sequence of a hydrophilic protein. Nevertheless, it was found to adsorb polybutylene succinate co-adipate surfaces (PBSA) in the presence of $CaCl_2$ or NaCl. Once this protein adsorbed into PBSA surface, it allowed PBSA degradation via cut like 1protein (CutL1) (Ohtaki et al., 2006). This novel protein produced in *Aspergillus orzaye* is proposed to use different type of proteins for recruiting lytic enzymes to hydrophobic surfaces in order to promote degradation (Ohtaki et al., 2006).

Numerous studies of the pathogenic *Candida albicans* revealed that hydrophobic cells at room temperature are more virulent than hydrophilic cells. Involvement of the yeast cell surface hydrophobicity in mediating adherence of *C. albicans* to human tissues has also been reported (Hazen, 1989) . Perhaps the intracellular parasite *P. marneffei* is escaping opsonization by secreting a hydrophobic surface binding protein, to recruit lytic enzymes, which might degrade the antibodies that would otherwise make it susceptible to phagocytosis. Altogether,

the literature about the hydrophobic surface binding protein and the expression specific to the yeast phase suggests that it may play an important role in pathogenesis.

Reverse transcription polymerase chain reaction for Thaumatin, displayed equal and abundant expression between the yeast and mycelia forms at 24 hr (Fig. 4). Samples that were collected at 48 hr were subjected to RT-PCR, displayed differential expression. The mycelia transcript was present but the yeast phase was at very low transcript levels. Further quantitative analysis was not performed for this protein. However, using modern bioinformatics tools a computational analysis on Thaumatin was performed.

Thaumatin belongs to the family of thaumatins known as natural sweeteners. Thaumatins are pathogenesis related proteins, can be induced by several agents ranging from ethylene to pathogens. The FASTA sequence of Thaumatin present in *P. marneffei* was obtained from NCBI protein database. The FASTA sequence was then used as input in basic local alignment to identify similar protein sequences to *P. marneffei*'s Thaumatin. The protein showed as uncharacterized protein in programs such as Uniprot, Expassy and MMDB. Signal 4p allowed predicting the presence of a signal peptide and cleavage site. This protein, showed close sequence similarity using multiple sequence alignment and neighbor-joining tree with the following species: *Neurospora crassa*, *Zea mays*, *Orza satica japonica*, *Coprinorpris Cinerea* and *Magnaporte Orzaye* (data not shown).

None of the species to which Thaumatin shares homology cause disease in humans. Hence, this suggests that Thaumatin may not be involved in virulence.

According to RT-PCR analysis, transcriptional regulator and carbonic anhydrase genes displayed equal expression between the given conditions. Transcripts levels of the mold and yeast phase were both present as seen in Fig. 4. According to NCBI database, transcriptional regulator is a provisional protein and it has not yet been classified to any domain super family. Therefore, without any additional information it is not possible to give suggested functions. Recent studies revealed that the zinc binding histidines and other residues of carbonic anhydrase are essential for enzymatic activity (Marchler-Bauer et al., 2011). So far, there is not much known about this proteins function in other fungi.

CONCLUSION

The pathogenic fungi *P. marneffei* is the only *Penicillium* species that displays a temperature dependent dimorphism. Pathogenecity of this fungus is associated with dimorphism. Presumably, genes that encode secreted proteins induced at 37ºC may play an important role in pathogenesis.

The present study verified our hypothesis that *P. marneffei* expresses a distinct subset of 37ºC genes that encode secreted proteins.

Such findings in the expression of secreted proteins during different growing phases would contribute to propose gene functions. RT-PCR was performed on all selected secreted proteins to display qualitative expression. Differential expression

was displayed throughout the genes at different time phases. Further analysis using qPCR gave insight to the relative amount of expression per gene.

Finally, proposed functions of the various genes studied were suggested based on the function of similar genes in other species. The HYD gene appears to be a yeast specific gene at all growth condition. Therefore, this yet undiscovered gene deserves more investigation since it may potentially be involved in virulence and may also be a target for drug discovery.

References:

Andrianopoulos, A. (2002). Control of morphogenesis in the human fungal pathogen< i> Penicillium marneffei</i>. International Journal of Medical Microbiology *292*, 331–347.

Aoki, W., Kitahara, N., Miura, N., Morisaka, H., Yamamoto, Y., Kuroda, K., and Ueda, M. (2011). Comprehensive characterization of secreted aspartic proteases encoded by a virulence gene family in Candida albicans. J Biochem *150*, 431–438.

De Bernardis, F., Boccanera, M., Adriani, D., Spreghini, E., Santoni, G., and Cassone, A. (1997). Protective role of antimannan and anti-aspartyl proteinase antibodies in an experimental model of Candida albicans vaginitis in rats. Infection and Immunity *65*, 3399–3405.

Brakhage, A.A. (2005). Systemic fungal infections caused by Aspergillus species: epidemiology, infection process and virulence determinants. Current Drug Targets *6*, 875.

Cao, C., Liang, L., Wang, W., Luo, H., Huang, S., Liu, D., Xu, J., Henk, D.A., and Fisher, M.C. (2011). Common Reservoirs for Penicillium marneffei Infection in Humans and Rodents, China. Emerging Infectious Diseases *17*, 209.

Cao, L., Chan, K.-M., Chen, D., Vanittanakom, N., Lee, C., Chan, C.-M., Sirisanthana, T., Tsang, D.N.C., and Yuen, K.-Y. (1999). Detection of Cell Wall Mannoprotein Mp1p in Culture Supernatants of Penicillium marneffei and in Sera of Penicilliosis Patients. J. Clin. Microbiol. *37*, 981–986.

Chaffin, W.L., López-Ribot, J.L., Casanova, M., Gozalbo, D., and Martínez, J.P. (1998). Cell Wall and Secreted Proteins ofCandida albicans: Identification, Function, and Expression. Microbiol. Mol. Biol. Rev. *62*, 130–180.

Chan, Y., Path, M.R.C., and Chow, T.C. (1990). Ultrastructural observations on Penicillium marneffei in natural human infection. Ultrastructural Pathology *14*, 439–452.

Chandler, J.M., Treece, E.R., Trenary, H.R., Brenneman, J.L., Flickner, T.J., Frommelt, J.L., Oo, Z.M., Patterson, M.M., Rundle, W.T., Valle, O.V., et al. (2008). Protein profiling of the dimorphic, pathogenic fungus, Penicillium marneffei. Proteome Science *6*, 17.

Cooper, C.R., and Vanittanakom, N. (2008). Insights into the pathogenicity of Penicillium marneffei. Future Microbiology *3*, 43–55.

COOPER JR, C.R., and Haycocks, N.G. (2000). Penicillium marneffei: An Insurgent Species Among the Penicillia1. Journal of Eukaryotic Microbiology *47*, 24–28.

Frieman, M.B., McCaffery, J.M., and Cormack, B.P. (2002). Modular domain structure in the Candida glabrata adhesin Epa1p, a β1,6 glucan-cross-linked cell wall protein. Molecular Microbiology *46*, 479–492.

Hazen, K.C. (1989). Participation of yeast cell surface hydrophobicity in adherence of Candida albicans to human epithelial cells. Infect Immun *57*, 1894–1900.

Hube, B., Sanglard, D., Odds, F.C., Hess, D., Monod, M., Schäfer, W., Brown, A.J., and Gow, N.A. (1997). Disruption of each of the secreted aspartyl proteinase genes SAP1, SAP2, and SAP3 of Candida albicans attenuates virulence. Infection and Immunity *65*, 3529–3538.

Kumar, A., Ahmed, R., Singh, P.K., and Shukla, P.K. (2011). Identification of virulence factors and diagnostic markers using immunosecretome of Aspergillus fumigatus. Journal of Proteomics *74*, 1104–1112.

Lum, G., and Min, X.J. (2011). FunSecKB: the Fungal Secretome KnowledgeBase. Database *2011*, bar001–bar001.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., and Gonzales, N.R. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research *39*, D225–D229.

Ohtaki, S., Maeda, H., Takahashi, T., Yamagata, Y., Hasegawa, F., Gomi, K., Nakajima, T., and Abe, K. (2006). Novel hydrophobic surface binding protein, HsbA, produced by Aspergillus oryzae. Appl. Environ. Microbiol. *72*, 2407–2413.

Peberdy, J.F. (1994). Protein secretion in filamentous fungi — trying to understand a highly productive black box. Trends in Biotechnology *12*, 50–57.

Piehl, M.R., Kaplan, R.L., and Haber, M.H. (1988). Disseminated penicilliosis in a patient with acquired immunodeficiency syndrome. Arch. Pathol. Lab. Med. *112*, 1262–1264.

Pongpom, M., and Vanittanakom, N. (2011). Characterization of an MPLP6, a gene coding for a yeast phase specific, antigenic mannoprotein in Penicillium marneffei. Medical Mycology *49*, 32–39.

Pongpom, P., Cooper, C.R., and Vanittanakom, N. (2005). Isolation and characterization of a catalase-peroxidase gene from the pathogenic fungus, Penicillium marneffei. Medical Mycology *43*, 403–411.

Ranganathan, S., Garg, G., and others (2009). Secretome: clues into pathogen infection and clinical applications. Genome Medicine *1*, 113.

Sandini, S., La Valle, R., Deaglio, S., Malavasi, F., Cassone, A., and De Bernardis, F. (2011). A highly immunogenic recombinant and truncated protein of the secreted aspartic proteases family (rSap2t) of Candida albicans as a mucosal anticandidal vaccine. FEMS Immunology & Medical Microbiology *62*, 215–224.

Segretain, G. (1959). Penicillium marneffei n. sp., agent d'une mycose du systeme reticulo-endothelial. Mycopathologia *11*, 327–353.

Soanes, D.M., Alam, I., Cornell, M., Wong, H.M., Hedeler, C., Paton, N.W., Rattray, M., Hubbard, S.J., Oliver, S.G., and Talbot, N.J. (2008). Comparative Genome Analysis of Filamentous Fungi Reveals Gene Family Expansions Associated with Fungal Pathogenesis. PLoS ONE *3*, e2300.

Sriranganadane, D., Reichard, U., Salamin, K., Fratti, M., Jousson, O., Waridel, P., Quadroni, M., Neuhaus, J.-M., and Monod, M. (2011). Secreted glutamic protease rescues aspartic protease Pep deficiency in Aspergillus fumigatus during growth in acidic protein medium. Microbiology *157*, 1541–1550.

Supparatpinyo, K., Khamwan, C., Baosoung, V., Nelson, K.E., and Sirisanthana, T. (1994). Disseminated Penicillium marneffei infection in southeast Asia. Lancet *344*, 110.

The UniProt Consortium (2009). The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Research *38*, D142–D148.

Thirach, S., Cooper, C.R., and Vanittanakom, N. (2008). Molecular analysis of the Penicillium marneffei glyceraldehyde-3-phosphate dehydrogenase-encoding gene (gpdA) and differential expression of gpdA and the isocitrate lyase-encoding gene (acuD) upon internalization by murine macrophages. J Med Microbiol *57*, 1322–1328.

Ungpakorn, R. (2000). Cutaneous manifestations of Penicillium marneffei infection. Current Opinion in Infectious Diseases *13*, 129–134.

Vanittanakom, N., Cooper, C.R., Fisher, M.C., and Sirisanthana, T. (2006). Penicillium marneffei Infection and Recent Advances in the Epidemiology and Molecular Biology Aspects. Clinical Microbiology Reviews *19*, 95–110.

Vinck, A., De Bekker, C., Ossin, A., Ohm, R.A., De Vries, R.P., and Wösten, H.A.B. (2011). Heterogenic expression of genes encoding secreted proteins at the periphery of Aspergillus niger colonies. Environmental Microbiology *13*, 216–225.

Wartenberg, D., Lapp, K., Jacobsen, I.D., Dahse, H.-M., Kniemeyer, O., Heinekamp, T., and Brakhage, A.A. (2011). Secretome analysis of Aspergillus fumigatus reveals Asp-hemolysin as a major secreted protein. International Journal of Medical Microbiology *301*, 602–611.

Youngchim, S., Vanittanakom, N., and Hamilton, A.J. (1999). Analysis of the enzymatic activity of mycelial and yeast phases of Penicillium marneffei. Medical Mycology *37*, 445–450.

APPENDIX

Hypothetical protein 1

atgaagttcttaccctctctcatcgtcctcggtctttctacccaggctcttgcgagctcttacatcgattatgttactaaggacggtcatggtctta
ctgtttacaaggtcatagattacattaacactactacctcagacttcaatgcccggattcagagttacgagggcggtgatctccgaaacattctcga
tggcggtaataaagttactagaataatgaagataggggctatgatggtcaagcagaaaacgacggaaccccttactgatgctcagttccctagcttc
ctcttacatatggggaaagagggtggttttgaagatacgttgaaaaaggcgatggatactcttatcgttaagaagtcgcttattcttgcttctggca
agggttctaagcttgggttagtgctttacagtcagcagatggcttccatagtcctcagtgccgagatcattgggaagattcccgcagaaaaggtcca
agattttagatcacactggcatgagaaggtcatccagatcatcgggagagctgttgatacctttgatacctctatccatcccgcgccagtaccctct
atgtcttctattaccttagacgctagaggtgctgtagtaacaaataattcgcctgctggccctgcttttactagtgctgctagtgctaactatactt
caagaatgggagagagtagcaggagccagaaagatgggtccgagaacgctcgagaatgtagctaatacaatgattcctgccagtaccagaatgacaa
atgttccagtatctaatgtaattgtgatgatggagtctgaagttacgggcctaagtctcaatgctcccgccactagaggctttgtaagagctaccgc
cattatttcaatgatcttattacttctatccccgatactaccagttcgtcttttgctgccttggggaatgactacgagtcaagggatcgaaggagaa
tgtataccccttttctcccaccaaaacttctatgcaacttttttccgctacctatgagaatagcaattcttcagcgaataagaacgaagaccgttccca
agattcgaacccaatcacgaaatgtcagtcgtctaccgaaggtatcaggagtcacggctctagtaacccttctaagggcgtcttttccaggttctaa
aatctagtgtgaccgtactcttccagtaggtctagtagccctctcgacaactatggaaactatggagataggtagggcgcggtcatgggagatacag
aagataatggaatctgcgatctccacgacatcattgtttattaagcggacgaccgggacgaaaatgatcacgacgatcacgattgat

Hypothetical protein 2

Atgcgctgtatgaacatcttcctctcggctctaggactttgtgtcctatacgtgagcgctctacctacaccagattccggttccgacgctgtttcag
cagcagcggtggcttctaccagcacaactccctcgcaggatggttcttcagcaccggcagaagctgattcggacgatgagtttagctaccagtggcg
aaagtgtacgcgacatacttgtagaaggagagccgagatcctgaaacacaggatatgcactcgcgagatggatgtggtctaaggccaaggctgcgac
aaagtcgtcgtcgccaccgaagatggtcgtgttgagggagcgtcctaccaagaagtcgtggccgtcttcgactaagcctgctactcaaatcgatggt
caccgctttcac

Hypothetical protein 3

atgaagttcaatattgccgccatatttggcgcctttgcggccgtctctatggcattcccaattgatggcaaggctcagggaaaggcccagcgtgatg
ttgacgcagcacgcgagtaccttgaaggtctgtcagcggtcaaagcgcgtgacctcgacggcaaggctcagggcaaagctcagggcaaggcgcaacg
tgatcttgaagaagacgaagagtacgagtacctggatattgctgagcgtgacctcgatggtaaggcccaaggaaaagcgcagggcaaggcccagcgc
gatctagagggcaaagcccaaggaaaagcacaacgcgatttggatggcaaagctcagggcaaggctcaacgtgatctagatggaaaggcgcagggta
aggctcagggtaaagcccagcgcgatttggaaggaaaggctcagggcaaggctcaacgtgatctagatggaaaggctcaaggaaaggcccagcgcga
tctagaaggtaaagcccagggaaaagcacaacgcgatttggatggcaaagctcagggcaaggctcagggtaaagcccaacgtgatcttgagggcaaa
gctcaaggaaaggctcagggtaaagcccagcgcgatttggaaggaaaggctcagggcaaggcccaacgtgaccttgagggaaaggctcaaggaaagg
cccagcgcgatctagagggtaaagcccagggaaaagcacaacgcgacttggaaggaaaggctcagggtaaggctcagggtaaagcccaacgtgatct
tgagggcaaagctcaaggaaaggctcagcgcgacttagatggcaaagcccagggcaaagcacaacgcgaccttgatggtaaagctcagtatacttca
agttataacggcggtataaaccgcggaaacgccggcagagataccgtaagggttaactaccgttccgagtccctttccgggtcgcactacaactgcg
tcgtgcgctcatggaacttccagacagtcgccagtttcgcgcactggagctgccgttccgagtcccgtttcgagtcccgttccgcgttgcactagaa
cttcttctgcttctcatgctcatggacctataacgactcgcactggagctaccattccgggttccttttcgcgtcccgttccgggtcgcgctagatc
tcccgtttcgggttccttttcgtgttgcgctaaacctaccgtttcgagtcccgttccgagttgcactagatctacctttccgcgtcccattccgagt
cccatttcgggtcgcgctaaaccttcctttccgagtcccgttccgagttgcactagatctacctttccgagttcctttccgggtcgcgctagatctt
ccatttcgggtccctttttcgtgttgcgctaaacctaccgtttcgagtcccgttccgagtcccatttcgggttgcactagaactcccgtttcgagttc
ctttccgagtcccatttcgggtcgcgctaaaccttcctttccgagtcccgttccgggttgcactggaactccctttccgagttcctttccgggtcgc
gctagatctcccatttcgggtccctttttcgtgttgcgctgaaccttcctttccgagtcccattccgagtcccatttcgggttgcactagaactcccg
tttcgagttcctttccgagtcgcgctgaatctaccgtttcgggtcccgtttcgtgttgcgctggaactaccatttcgagtcat

Hypothetical protein 4

atgttgggcaaatacgtcttggcattgggtcttgtcgcaagcaccgtgcttgcacagcctgtgattcccaacaaggtcgacgtcgacaagcgtgtgg
atgttgataaggttgatgtcgacaagcgtgtgattgccaataaggttgatgtcgacaagcgtgtggatgttgataaggttgatgtcgacaagcgtgt
gatggccaataaggttgatgtcgacaagcgtgtggatgttgataaggtcgacgtcgacaagcgtgtggatgttgataaggtcgatgtcgacaagcgt
gtggatgttgataaggtcgacgtcgacaagcgtgtggatgttgataaggtcgacgtcgacaagcgtgtggatgttgataaggtcgacgtcgacaagc
gtgtggatgttgataaggtcgacgtcgacaagcgtgtggatgttgataaggtcgacgtcgacaagcgtgtggatgttgataaggtcgacgtcgataa
gcgtaaggtcaacgtcgattatacaacccgtttatgcagaaccgtaacccagaacagcgttcgtggcacgaacgtgtcggacactaagggttgttcc
agctgcagctgttcgcacacctacaactattccaactacagctgttcgcacactaacggttattccaactacagctgttcgcacacctacaactatt
ccaactacagctgttcgcacactaccggttattccaactacagctgttcgcacacctacaactattccagctgcagctgttcgcacacctacaacta
ttccagctacagctgttcgcacacctacaactattccagctgcagctgttcgcacacctacaactattccagctgcagctgttcgcacacctacaac
tattccagctgcagctgttcgcacacctacaactattccagctgcagctgttcgcacacctacaactattccagctgcagctgttcgcacacctaca
actattccagctgcagctattcgcattccagttgcagctaat

Hypothetical protein 5

atggttaaatacgcacttgctctttcactggtcgctggagctgttgccgctctgccggcagctcctggagtgaacagccaggctcctcctagcctac
ccaacgctaacccagcggctataatgtctgtcccacaaagcaatgttaagagctaccagaccggtggtcttgttcgttacaacgtcactgaaagttc
cgggctctccaagcgttggggctgcagcgcctcgcctactcttacctggggtgatgccgataatggcggtccaggtatcactatcgacaacgattcc
aatgactggcgtggtttctacttctaccacaacagctgtgacagcattccctggaagtacatttggattgctggtaaaaccacccagttcgtctccg
ttcctacaggctgggcgggtcgtgtccagcgaggcgtcgatgctagcatgctcaatggtcagccacagttgctcggttcttggttagagattagctg
ggacgccgccaatggtaataccggctgggctgacgtctcgctcattcggggtaatgatggtggtattcttgtctggaatgccaatggtgactggaag
ggattcacacagtgggttctggatggcgcacctgaaggtgcctatgacatgaagaacgacggacaatgggttatcaagtacaccgagaacaatgacg
gctcgatcaacaccattccccgcgattgggacatccaaaaaattggctcccagtacgtctacgtcgacgatgcccatggcagccctgtcatttcaac
ccccaaccagcgcttctccaccttctggcccgacggccgcgcttataccaatttatgcgtgaacgagaaagtgaccagcgacctcgacaacggcgag
acggccgtcgaggacctcacttgtcggtccgaggaggatcggatgggttgcgattgggtcgccgatattacagacagggtgtttcgttacaattctc
gatggtctggccaccagaacaagcaatgttgcagtgactttcaaggcccgagaggttcgcaaccccgacgtcgcggagcggatgagaatggacccca
ctacggctattaccgccaggtccatagtgatagctgttgctaaggttactgaccgcaccaaagatgaagatggtgttgtcgacactgtcgtaaggga
ccttcatgtaaacctaacgaccattttggtgggtcaagcagaggcaaggatgtccgacccgcccagcacaggtcgctccgcagctacgatcgtacga
gttaccagtcggtgtcaacgagccaagaaccaatctctaatcgaccctgcggcggttaccattatggccgacccgactgcagagcgagtaagcccca
ttactaccaccataagaacagaccttacggttaccactgaccttccctaagtgtgtcacccaagacctaccgcgtggacttccacggatactgtact
tcttgctgcctgttacccaatagttcatgtggctcttgttactgccgagctagttgtggtaaggggcgctaaccctgtaggttttttaaccgagggt
catgcagatgcagctgctacgggtaccgtcgggacagtaaagttgggggttggtcgcgaagaggtggaagaccgggctgccggcgcgaat

Hypothetical protein 6

atggttcagtcaaaacttacgttactggcacttgctagtacggcgcttgctggtgtggttcgacgagaagaccccgaagt
tgacgacttaatcccggaggatctctcctatctcctgccactagactgggttaccccaacacttgattacttcgaggaaa
cagatccattacctacgccctttgtacaagcattctcagactatttacaagagccacactcaacttacctctcaactggt
ttacctctttatcttccaccgcccacagaggatgcgcttccaccattcattacatcttttcctggggatggtgagtttgc
ccacactaaagattcacccaagcccactacatttattccaacgatccccccatccagcacctctaggtgcttcctatgctg
gaccgccgcctactttcagtgagccatcgagaccagaagaatcgcagctggccaccccaccacctgcatttaaagcctcc
tccagccatccacctatcccgtcctcgcaagcccagacgccatatgcatcgccctcaaagcgcccatcgaccggtgggca
gccggcaccttctatcccagcttctccaaaacctatggggccctgggatggtcttgatcctgtgaaaccaagcggttacc
tggaaactccatcgtggagctcacacataccaccaccatccaaatctacaccaggtccagaaggatcagaaaaaaatcac
ggcattacccccctccgctttgtccgaggcttacctgaggccatctttgccatttaactctatcccagatgagctctctga
gcctcatggcaactcccctggtagttctcaccaggctccaccaccatacgaatctgagcaggaaacacattctccgccac
caaattttgcttctggtagttcggctagacccgatagtagccccccctgctacatctgatcaacttcctccactacacgga
tcctaccaggggcctccgccatttaattctggcctcgacggcccatctggggagagcagtcctcctgcggagtctcataa
aaccaatcctcaacctgaatcgtaccagggccccttctccaactgtcaccgtccctgaggagccatctcagccgaaaggta
gcccacccgctgaaactcaccaaattccacctcccctcacgaagccttatcaagagtcatttcctccgccatttgtctct
gcagcgggagaacaacccaagcctgagagcagctcgacccagtcttatcaacttcctccactctcatctctgcattcttc
ataccctggtgatgttgatattcctcctgaacctactaattctcacgaggagccatcgcctcactatggtgccgctcccc
ttgcatcctacacttgctcatcactcgatgaagctcccgcagacgctcattctccgcaccccggggaaaccagtccgccg
gccattatccctccatttgagaatgataaggaacctcatgatcatgaactagctccttcacatccagtttcagcgtctga
gaaaccaagagttacacatactaacgtatatgtacctcatttctcgaagccggaacctgcgggcaccgaagccccatctg
tctctcatgccgccattccaacaaccgtaccatcaaattgcgctgatgtaaattccaaaccggccattccttcaccgaca
agtggcaaacctcaactacagggaaccaccggggaacatcacctttctgggtctccatcacatatgcatctcgcggaggc
ttccggtgtagttgaacaaccacctgcaacagaagcggctccagaaccttccaaaatgggcgtgccaaatgctgatgagt
cggcttttcattattcaagttttgccccaacctccgctgttccaatcactgcattgttagaaactttcggtgctttggcg
tttttctttctata

Hydrophobic surface binding protein

atgaagctcttaccctcccttatactccttagtctttctgcccaagcacttgcaagcccttatatcaatcactatactactaaggaccagcgtgatc
tctctacttataaatctgttatccagaatatcctcacgactgtggtcaagtttggccaggatattgtgctaaatacctcacatcttaataatgactt
tgataatattattgatgctatagcccaagggactaagaaactcgaacctgaggctccccttaactactttcaagctggtgtccttcttatccctatt
ataccagtaaacaaagcgatccatatggctcttcaaagtcttatcaataggagggatattcttgttgctgctgtaaaattctccttaagaaggtaca
cacagtcctacactgagatcattgacgacatcactgaagctattaatgtttatagtaaagttgttggcgccacaagccataatacttgtacttcgag
aatgggagggaatatgaggaatcagaaagacgggttcgtgaacgttcgggaatatagttagtgatatgatgattcctggtcgcactagagagatgaa
tatttagacaataggtcttataggagtgctgacaccagttcaaaccggtcctataacacgatttatggagtgtagaattattactgaaactattata
ataactacgatatcgggttccctgattctttgagcttggactccgaggggaattgatgaaagttcgaccacaggaagaatagggataatatggtcat
ttgtttcgctaggtataccgagaagtttcagaatagttatcctccctataagaacaacgacgacattttaagaggaattcttccatgtgtgtcagga
tgtgactctagtaactgctgtagtgacttcgataattacaaatatcatttcaacaaccgcggtgttcggtattatgaac

Carbonic Anhydrase

atggattttgctctcgcttttgtgtctcttatcacggcagcttctgcctcctgtatttatggcacatcgctcatgcctcgtgccgctgaaggcgtag
tggacatactgtccttcaactatactgccactggcggccctctgaactggcatttattaaatacaacagctaataatgcctgcgccacgggcaaaaa
ccagtctcccgtggatatagtgatggaaggtatcacttacgctattcctggatctgtgaaactggacataccctgtgcaggtggagttgagctggag
aacctgaacaacgttgtggaggtagtggcaaataatggtactcttactactcccgaatctatctacaagcttgcccagttccatttttcatactccaa
gcgagcaccgtgtcaacgaggagtactttccaatggaagcccactttgtcttcgagaacgaagcatctcaaatcgctgttgctgccttcctcttcca
gctctctgagtctggcgcttccaatccgctttcgactcggttttgcccacctcgacgacatcactacacctggtaccttcacaaagaccggaccc
ttggacttcactactgtcaaccaacatttttcaaatcacggtatcttccagtattcgggttctcttacgactccaccctgcagtgaagggctttcct
ggtacatcagtacagagccaatgcctctcaatgtccaaacctataataaggtgaagaaggtggtgaaattcaacgctcgttatacccagaataccct
cggtcagaacaacttgttggaacttgcagccactccgtcgggatatacctaaaacgagagcgaaaacacagagaatagtgccgtcgaagacggagga
cataaataccgtgtagcgagtacggagcacggcgacttccgcatcacctgtatgacaggaagttgatatgacggtgaccgccgggagacttgaccgt
aaataatttatgttgtcgattattacggacgcggtgcccgtttttggtcagagggcacctatatcactaccttccatagtgaatgcgataaggacct
agacactttgacctgtatgggacacgtccacctcaactcgacctcttggacttgttgcaacacctccatcaccgtttattaccatgagaatgatgag
ggcttagatagatgttcgaacgggtcaaggtaaaagtatgaggttcgctcgtggcacagttgctcctcatgaaaggttaccttcgggtgaaacagaa
gctcttgcttcgtagagtttagcgacaacgacggaaggagaaggtcgagagactcagaccgcgaaggttaggcgaaaagctgagccaaaaacgggtg
gagctgctgtagtgatgtggaccatggaagtgtttctggcctgggaacctgaagtgatgacagttggttgtaaaaagtttagtgccatagaaggtca
taagcccaagagaatgctgaggtgggacgtcacttcccgaaaggaccatgtagtcatgtctcggttacggagagttacaggtttggatattattcca
cttcttccaccactttaagttgcgagcaatatgggtcttatgggagccagtcttgttgaacaaccttgaacgtcggtgaggcagccctat

Transcription Regulator

atgggtcactgggcaattgctgctgctttggcaaccctttgccagccggtagtctggcattgacattgagcagcaccgcatcatcagccacgacta
cagccgccaccgcgtccaccattgtcacctgtaccacaggaattcccaacttgatcgtcaacccgtcctttgaagatggcttcagtagctgggggtt
tgcctccggtacctccggatccgtcgtcagtggacatgcgtcggacggcagctattatctcgcggccactgacatacaaacgcactccggcatcatg
atctaccagtatgtcaccggtttcaaggcaggcatgacatacacagcatcgttggactggtccctgtcggcgctttcaagtggcatttcacagagct
gccacctttattggtatatggacagttatggcagcacaattggcagtaccagtgccaccctcacgactaacagcatggactggacccattttgctac
cgtctatacaccttccaccgatggaactcatttgttggagatctatgcattctgccaagctagcgtcgcatcgttcacaggtgctgcaatcaagctt
gataacattaaattaagtaacggcgttgcgactacttcctgcaccacctctacggtgaccccgacggcgatctctgtgtcttcgaccagttccagcg
ctgcagtaccgcctaccgtcactgctagttcctctgctgtgccatctatgatttcttcttcggcggcagtctcgtcttctgccgttccacccgtcat
tagttctagttcggcagtaccatccatgatttcttcttcggcggcagtctcgtctcccgccattccatccgccattagttctagctctgccgttcga
atcccgagcagcatcccctcctcttccgccgccgtctcttctctgcccagcacctcacaatctgtgatccctccggcccaggcaaccagcggcagcc
cgactagcagtatgcccaccacacgaagatgcttaaagtcgaccactttaaccagtgtccctgtcgcaagccagactactacctccggggacgacag
ccaacaattcaccacctcaacagtcttcagcacactcaccgccaccatcaccgcttgcccatctactgtatccgactgccccgcaaactcccgcact
acgtatgttactactgagaccgtggtgctgtccaccaccatctgccccgttactgcggtggcacacacgacctctgtccctggctctggctctagct
ctggctcaggatcaggctcaggatcaggatcgggttcaggcccgagctcaggttccacctctggctcaggttctggctctggctctggctctggctc
tggctctggctcaggttctcttgcaacagccgtcgtcagttcctcgaccactaaacccactggcccgagcgcgacgagtagtgccactggtaccttg
tatactggtgccggctcaacagaacatgtcagtcttccaggggtaatttccttggtggcgatgcttctgtatagcatctactatacccagtgacccg
ttaacgacgacgaaaccgttgggaaaacggtcggccatcagaccgtaactgtaactcgtcgtggcgtagtagtcggtgctgatgtcggcggtggcgc
aggtggtaacagtggacatggtgtccttaagggttgaactagcagttgggcaggaaacttctaccgaagtcatcgacccccaaacggaggccatgga
ggcctaggcagcagtcacctgtacgcagcctgccgtcgataatagagcgccggtgactgtatgtttgcgtgaggccgtagtactagatggtcataca
gtggccaaagttccgtccgtactgtatgtgtcgtagcaacctgaccagggacagccgcgaaagttcaccgtaaagtgtctcgacggtggaaataacc
atatacctgtcaataccgtcgtgttaaccgtcatggtcacggtgggagtgctgattgtcgtacctgacctgggtaaaacgatggcagatatgtggaa
ggtggctaccttgagtaaacaacctctagatacgtaagacggttcgatcgcagcgtagcaagtgtccacgacgttagttcgaactattgtaatttaa
ttcattgccgcaacgctgatgaaggacgtggtggagatgccactggggctgccgctagagacacagaagctggtcaaggtcgcgacgtcatggcgga
tggcagtgacgatcaaggagacgacacggtagatactaaagaagaagccgccgtcagagcagaagacggcaaggtgggcagtaatcaagatcaagcc
gtcatggtaggtactaaagaagaagccgccgtcagagcagagggcggtaaggtaggcggtaatcaagatcgagacggcaagcttagggctcgtcgta
ggggaggagaaggcggcggcagagaagagacgggtcgtggagtgttagacactagggaggccgggtccgttggtcgccgtcgggctgatcgtcatac
gggtggtgtgcttctacgaatttcagctggtgaaattggtcacagggacagcgttcggtctgatgatggaggcccctgctgtcggttgttaagtggt
ggagttgtcagaagtcgtgtgagtggcggtggtagtggcgaacgggtagatgacataggctgacggggcgtttgagggcgtgatgcatacaatgatg
actctggcaccacgacaggtggtggtagacggggcaatgacgccaccgtgtgtgctggagacagggaccgagaccgagatcgagaccgagtcctagt
ccgagtcctagtcctagcccaagtccgggctcgagtccaaggtggagaccgagtccaagaccgagaccgagaccgagaccgagaccgagaccgagtc
caagagaacgttgtcggcagcagtcaaggagctggtgatttgggtgaccgggctcgcgctgctcatcacggtgaccatggaacatatgaccacggcc
gagttgtcttgtacagtcagaaggtccccattaaaggaaccaccgctacgaagacatatcgtagatgat

Thaumatin

atgaagaaatctagaccgacaaaactttcaagtcttctggcgattctgacttctctcgtccacgaagtcaccgccatccatcacctgcgactatcac
ccgtaatcctcgtcacaagggacacatccactcgcgcctttagaataacaaaccactgccccgtcgacatatacccagcaatccaaacacaagccgg
caccggcccagcattcggcggctatcatgccctgcccggaaacacaaccagcttcgacgtcagcgcaggttggcaggggcggatatgggcgcgcacc
aactgcactttcaatgaaaatggaatgagtacgaatgggatgggggcagatgcttgtttgacgggggattgtggtgggttgttggattgtcagggga
ctgggaaaccagctacacttgcagaattcacaatgtcgggaagcatgaatctatcttattatgacatttcgctagttgatggatataatctcccatt
ggggataatttcattgcatatcgaatcaagcgaccccgacctgaaaaacattccatcaaatacaacaaatcctgtctgcatcggaagtcccaatttc
ttacaagccacaaatcagcttgaaaatggtacgacagtaccttccaactccaccattggcagtatcacctattcaacgccactagagatgacactga
gcacacacgcagtttcacaatggtgtccctggcctcttcagatacaaccccccccgaaacctggggcgggcgtatatccatatcccgatgatggaat
tgcgcgtccagattttgatccttgtttgtctctgtgctcgagatatggaaagtcgaattattgttgcaccggtaaacacggtacggcggacaaatgt
agcccgaatatgtatagtaaagctgcaaagaaggtgtgtccggatgcgtatagttatgcgtatgatgatacgacatcaactttcactgtcaaacagg
gcgctgggttcgaagttgtcttttgtccgtctggtcggtcgacgaatattcaaaagtcgttgacgacttatacttctttagatctggctgttttgaa
agttcagaagaccgctaagactgaagagagcaggtgcttcagtggcggtaggtagtggacgctgatagtgggcattaggagcagtgttccctgtgta
ggtgagcgcggaaatcttattgtttggtgacggggcagctgtatatgggtcgttaggtttgtgttcggccgtggccgggtcgtaagccgccgatagt
acgggacgggcctttgtgttggtcgaagctgcagtcgcgtccaaccgtccccgcctatacccgcgcgtggttgacgtgaaagttacttttaccttac
tcatgcttaccctacccccgtctacgaacaaactgccccctaacaccacccaacaacctaacagtcccctgacccttggtcgatgtgaacgtctta
agtgttacagcccttcgtacttagatagaataatactgtaaagcgatcaactacctatattagagggtaacccctattaaagtaacgtatagcttag
ttcgctggggctggacttttgtaaggtagtttatgttgtttaggacagacgtagccttcagggttaaagaatgttcggtgtttagtcgaactttta
ccatgctgtcatggaaggttgaggtggtaaccgtcatagtggataagttgcggtgatctctactgtgactcgtgtgtgcgtcaaagtgttaccacag
ggaccggagaagtctatgttgggggggctttggaccccgcccgcatataggtatagggctactaccttaacgcgcaggtctaaaactaggaacaaa
cagagacacgagctctatacctttcagcttaataacaacgtggccatttgtgccatgccgcctgtttacatcgggcttatacatatcatttcgacgt
ttcttccacacaggcctacgcatatcaatacgcatactactatgctgtagttgaaagtgacagtttgtcccgcgacccaagcttcaacagaaaacag
gcagaccagccagctgcttataagttttcagcaactgctgaat