UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Towards developing a new host-vector system for high-level protein expression

BOITUMELO NANA SANDRA PHOMA

A thesis submitted in partial fulfilment of the requirements for the degree of *Magister Scientiae* (MSc) in Microbiology in the Faculty of Natural and Agricultural Sciences, Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria

September 2015

Supervisor: Prof. D. A. Cowan

Co-Supervisor: Dr. E. Berger

Declaration

I, Boitumelo Nana Sandra Phoma, declare that the dissertation, which I hereby submit for the degree of MSc. (Microbiology) at the University of Pretoria, contains my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

_____ _____

Boitumelo Nana Sandra Phoma **Date**

MSc. candidate

Summary

Filamentous fungi are critical for the production of many commercial enzymes and organic compounds. Due to their decomposer lifestyle, they have a natural potential for large scale production of proteins, many of these extracellular. The concept of this project was to exploit the protein synthesis and secretory capability of a large number of indigenous fungi as the basis to generate a new fungal expression platform. Filamentous fungi were obtained from the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI) University of Pretoria, South Africa. Screening of the isolates for extracellular protein expression in malt extract liquid medium shake flasks, led to the identification of a *Clonostachys rosea* 17970 isolate after internal transcribed spacer (ITS) sequencing and morphological studies. A ~35 kDa protein secreted at high levels was identified through LC-MS/MS as a cuticle-degrading serine protease (Cdsp). The optimum pH and temperature for protein production were found to be 9.0 and 32°C respectively, after 96 h of incubation on malt extract liquid media. Designed primers were used for the PCR amplification of an internal fragment of the *cdsp* gene from chromosomal DNA. Sequence data was used for further PCR-amplification of the up and downstream regions through chromosome walking using Inverse and SiteFinding PCR technology. Analysis of the assembled *cdsp* gene fragment (2,923 bp) revealed that the gene harbours a 5' upstream region (1,141 bp), coding region (1,300 bp) and 3' downstream region (509 bp). The coding region contains three introns. Reverse transcriptase-PCR of the coding region revealed a 1,140 bp open reading frame that encodes Cdsp. The putative *cdsp* signal peptide, promoter and terminator regions were identified *in silico* and are predicted to be important elements for the construction of an expression vector. Overall, these results demonstrate that *C. rosea* 17970 could be a good source for extracellular protein production and is a potential candidate for development as a new host-vector system based on its secretory capability and identified regulatory regions via *in silico* analysis.

.

Acknowledgements

- First and foremost, I would like to thank the Lord for continuing blessing me with opportunities of improvement.

- My sincere gratitude goes to my supervisors, **Prof Don Cowan** and **Dr Eldie Berger** for their unwavering support and confidence in me. I am so grateful.

- The **National Research Foundation (NRF)** provided financial support for this research, and their support is greatly appreciated and acknowledged.

- I would like to extend my gratitude to my colleagues (CMEGers) at the Centre for Microbial Ecology and Genomics and members of the Department of Microbiology and Plant Pathology, for a lovely, memorable working environment and resources to accomplish my work.

- To **Helen**, **Alacia**, **Vincent**, **Mark**, **Deon**, **Olivier** and **Yashini**, I appreciate your helpfulness, friendship and the fun company you provided. Ke a leboga. I have only the best wishes for you all.

- Special thanks to my siblings, **Thapelo** and **Tebogo**, for being the people who taught me about fairness, co-operation, kindness and (healthy) competition – often in the hardest ways possible.

This thesis is dedicated to my parents:

Mamalea and Lesego Phoma

Table of Contents

List of abbreviations and symbols

<	Smaller than
>	Greater than
µg	Microgram
µL	Microliter
µM	Micromolar
APS	Ammonium persulphate
BLAST	Basic local alignment search tool
bp	base pair
cm	Centimetre
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleic acid triphosphate
dpi	Days post inoculaion
E. coli	*Escherichia coli*
e.g.	Example
EDTA	Ethylenediamine tetra acetic acid
g	G-force
g	Gram
H_2O	Water
Hr	Hour(s)
kDA	Kilo Dalton
L	Litre
LB	Luria broth
M	Molar
mA	Milliampers
mg	Milligrams
min	Minute(s)
ml	Millilitre
mM	Millimolar
nm	Nanometre
OD	Optical density
PCR	Polymerase Chain Reaction
rpm	Rotations per minute
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	Second (s)
TAE	Tris-acetate EDTA
TEMED	N,N,N-tetramethyl ethylenediamine
Tris	2-amino-2(hydromethyl) propane-1,3-diol
U	Units
UV	Ultra violet
V	Volt
v/v	Volume per volume
w/v	Weight per volume
wpi	Weeks post inoculation

List of tables

List of figures

CHAPTER 1:
LITERATURE REVIEW

"The expression system is not everything, but everything is nothing without a good expression system"

- DYADIC INTERNATIONAL, INC.

1.1 General introduction

The establishment of recombinant DNA technology over thirty years ago (Cohen *et al.*, 1973; Cohen, 2013), coupled with the discovery of restriction endonucleases (Linn & Arber, 1968; Rao *et al.*, 2014), has enabled the transfer, replication and expression of genes of interest, in a foreign organism. Both discoveries introduced techniques that allowed many organisms to be explored as host organisms (Cohen, 2013). While some peptides, proteins and numerous non-proteins (i.e.; organic acids and metabolites) are naturally produced in high quantities, and subsequently isolated from natural sources, many others are either unstable or present in amounts too low for commercial or research purposes (Mahmoud, 2007; Bill *et al.*, 2011). Using genetic engineering, Itakura *et al.* (1977) illustrated that 5 milligrams of recombinant somatostatin, a mammalian hormone, produced from two gallons (~7.57 litres) of *Escherichia coli* was equivalent to the amount extracted from 500,000 sheep brains. After similar successful experiments with other recombinant genes, scientists realized the great potential of biotechnology approaches.

As the building blocks of life, proteins are produced by all living organisms as part of their innate metabolism and to contribute to cell signalling, immune response, and the cell cycle (Demain & Vaishnav, 2009; Rosano & Ceccarelli, 2014). The interest in hosts that can over-produce recombinant proteins is typically attributed to their natural ability to produce and/or secrete high amounts of native protein (Bergquist *et al.*, 2002; Sarrouh *et al.*, 2012). In addition, the choice of an expression system is influenced by the quality and quantity of the recombinant product required. For example, commercial enzymes are often needed at multi-grams per litre and human therapeutic proteins at microgram per litre levels (Gasser & Mattanovich, 2007; Porro *et al.*, 2011). Other factors for consideration to satisfy the needs of both manufacturer and end user include, cell growth, expression level, cost of production, ease of scale-up, Generally Regarded as Safe (GRAS) status and the need for post-translational modifications (i.e.; phosphorylation, N- or O-linked glycosylation, disulphide bridge formation, acylation, and fatty acid acetylation among others) required to ensure

complete biological activity of most eukaryotic proteins (Nevalainen *et al.*, 2005; Sørensen, 2010). Proteins obtained from different biological expression systems continue to contribute to the biopharmaceutical, agricultural, biofuel, food, and beverage industries. Generally, prokaryotic expression systems are often the first choice microorganisms as they are well established and cost effective; whereas eukaryotes are used for specialized applications (Ferrer-Miralles *et al.*, 2009). The host/vector systems can be bacteria (e.g.; Gram negative *Escherichia coli* and Gram positive *Bacillus* spp.), yeasts (e.g.; *Saccharomyces cerevisiae, Pichia pastoris* and *Hansenula polymorpha)*, filamentous fungi (e.g.; species of *Aspergillus*, *Trichoderma*, *Fusarium*, and *Penicillium*), microalgae (*Chlamydomonas reinhardtii)*, insect cell lines (e.g.; *Baculovirus vector*), transgenic plants (e.g.; *Nicotiana tabacum*), protozoa (*Dictyostelium discoideum*), transgenic animals (mice, pigs, sheep, cows, and goats), mammalian cell lines (e.g.; Chinese hamster ovary (CHO), baby hamster kidney (BHK), hybridoma cells and cell-free systems (Nevalainen *et al.*, 2005; Merten, 2006; Terpe, 2006; Gong *et al.*, 2011; Rosano & Ceccarelli, 2014; Specht & Mayfield, 2014). To date, there is no universal expression system for the production of heterologous proteins because all systems have considerable advantages and disadvantages (Sørensen, 2010; Corchero *et al.*, 2013). As a result, scientists are constantly working towards optimizing existing systems and/or developing new systems to overproduce existing and new proteins.

The first recombinant human protein, insulin was produced in *E. coli* by Herbert Boyer (1977) and it became a successful biopharmaceutical product, Humulin™ (licenced by Genentech to Eli Lilly & Co.) in 1982 (Altman, 1982; Genentech, 1982; Porro *et al.*, 2011). Prior to this, insulin could only be extracted from bovine (cow) and porcine (pig) pancreases from slaughterhouses (Johnson, 1983; Walsh, 2005). Since the approval of Humulin™ by the US Food and Drug administration, similar or modified recombinant human insulin formulations are still being used for the treatment of diabetes. Many other recombinant protein-based biopharmaceuticals (e.g.; monoclonal antibodies, interferons, growth hormones, erythropoietin, coagulating factors, and vaccines, etc.) for the treatment of a

wide spectrum of human disorders, have been commercialized in parallel with the optimization of the *E. coli* system, as well as the development of alternative expression systems.

Recombinant protein production is a multi-billion dollar industry, largely comprising biopharmaceutical products and enzymes used for industrial purposes, provided by over 4,200 biotechnology companies worldwide (Adrio & Demain, 2014). Thus far, over 150 biopharmaceuticals have acquired FDA approval and about 200 more are predicted to be available over the next few years (Zhu, 2012; Bill, 2014b). Approximately one third of these biopharmaceuticals are produced in *E. coli*, half in mammalian cell lines and hybridomas, and the rest in yeasts (mostly *S. cerevisiae)* (Mattanovich *et al.*, 2012; Bill, 2014a). The global market for recombinant protein drugs had revenues of nearly $95.2 billion and $151.9 billion in 2008 and 2013, respectively. A leading market research company, BBC Research, predicted the market to reach a value of $179.1 billion by 2018 with a five-year compound annual growth rate (CAGR) of 5.6% from 2013 to 2018 (BBC Research, 2014a).

Similarly, the production of biological catalysts have flourished since their introduction to the commercial market, i.e.; approximately 90% of industrial enzymes are recombinant forms produced in fungi and bacteria (Adrio & Demain, 2014), and the rest are obtained from natural sources. These enzymes (e.g.; proteases, amylases, lipases, and cellulases, peroxidases, etc.) are products of microbial fermentation in large-scale bioreactors, and are used in approximately 150 industrial processes, e.g.; as bulk enzymes in the biofuel, paper and pulp, detergent, food, beverages, animal feed, leather, cosmetics, wastewater treatment and textile industries (Adrio & Demain, 2014). A list of commercial enzymes (updated in April, 2014; used for feed, food and technical applications) published by the Association of Manufacturers and Formulators of Enzyme Products (AMFEP; http://www.amfep.org/content/list-enzymes) is a good indication of the status of certain host strains and the enzymes they provide. The global market for industrial enzymes is predicted to reach US$7.1 billion by 2018 with a five-year CAGR of 8.2%, from 2013 to 2018. The

largest growth is expected to be recorded in the detergent enzyme segment with 11.3% CAGR over the 5-year forecast period (BBC Research, 2014b). The industrial enzyme market is very competitive. The major producers are in Western Europe, USA, and Japan. Whereas, emerging markets are identified as China, Eastern Europe, India, and with Africa lagging far behind (Demain, 2007; Sarrouh *et al.*, 2012; Adrio & Demain, 2014).

1.2 Microbial expression systems

1.2.1 Bacterial expression systems

1.2.1.1 E. coli

E. coli, a Gram-negative ɣ-proteobacterium is the earliest, mostly used prokaryotic expression system for pharmaceutical and industrial protein production. Characteristics that have kept it as a workhorse for over 40 years include, its rapid doubling time, ease of cultivation in affordable growth medium, well-studied physiology, and high level protein production capability (i.e.; up to 50% of total biomass) (Terpe, 2006; Chou, 2007; Sagmeister *et al.*, 2014). Currently, *E. coli* is used to produce about 50% of commercial recombinant proteins and over 70% for research purposes (Bill, 2014a). Additionally, numerous *E. coli* production strains, have acquired approval from the U.S. FDA for the synthesis of gene products that are safe for human use. In fact, the continuous use of this system has led to the accumulation of its whole genome sequence, which greatly simplified gene cloning and expression (Marisch *et al.*, 2013; Sagmeister *et al.*, 2014).

There are drawbacks that must be considered when using this expression system. *E. coli* usually over-expresses eukaryotic proteins as protein aggregates called inclusion bodies, which often protect from proteolytic degradation from host enzymes. Proteins trapped in inclusion bodies are often insoluble, biologically inactive and incorrectly folded; requiring costly and labour intensive solubilisation, *in vitro* protein folding and purification steps, which altogether may result in reduced protein yields (Ferrer-Miralles *et al.*, 2009; Sørensen, 2010). Additionally, eukaryotic codons are commonly inefficiently expressed and important

post-translational modifications, particularly glycosylation present in most eukaryotic proteins cannot be achieved in *E. coli* (Sørensen, 2010; Bachran *et al.*, 2013). However, Valderrama-Rincon *et al.* (2012) have made some progress in engineering glycosylation pathways in *E. coli.* Alternatively, an authentic protein can be obtained using mammalian cell lines. However, these expression systems are complex, expensive to cultivate, require highly-trained personnel and relatively low protein yields are often obtained (Agrawal & Bal, 2012; Kaufman, 1990a; Kaufman, 1990b). For recombinant proteins that do not require post-translational modifications (i.e.; non-glycosylated insulin, growth hormones (somatotropines), interferons (alpha, beta, gamma, or mutein) etc.), *E. coli* is often a good choice (Jeong *et al.*, 2014).

Recombinant protein production in *E. coli* occurs either in the cytoplasm, periplasm or secreted into the culture medium. For the production of most recombinant proteins, extracellular secretion is preferred because downstream processing (i.e.; cell lysis, solubilisation, refolding, purification etc.) is avoided. In addition, the target proteins may be in a soluble, properly folded, and active form (Choi & Lee, 2004). Efforts to guarantee complete extracellular protein secretion have been challenging. *E. coli* may also produce recombinant proteins contaminated with toxic pyrogens resulting from cell lysis. These endotoxins can have an immunogenic effect on humans and can be difficult to eliminate during purification steps (Rai & Padh, 2001; Lagoumintzis *et al.*, 2014).

Obtaining a soluble and active form of recombinant protein is perhaps the biggest challenge when using *E. coli*. This expression system is not naturally capable to produce target proteins that are not aggregated or intracellular. These challenges are often associated with the expression vector, induction parameters, and growth conditions (i.e.; temperature) used (Sahdev *et al.*, 2008; Voulgaridou *et al.*, 2013). Fusion expression tags (e.g.; glutathione-S-transferase (GST), thioredoxin (TRX), and maltose-binding protein (MBP) etc.) have been used to facilitate protein solubility. Many experiments, have identified MBP as a superior and most effective expression strategy because it also facilitated the formation of disulphide

bonds (Kapust & Waugh, 1999; Voulgaridou *et al.*, 2013). In addition, lowering the temperature of the culture (Semba *et al.*, 2010) or co-expression with molecular folder chaperones (DnaK, DnaJ or GrpE together with either GroES or GroEL co-chaperons) can also facilitate protein solubility and subsequent protein folding (Schumann W. and Fereira L. C., 2004; Yin *et al.*, 2007; Voulgaridou *et al.*, 2013).

1.2.1.2 Alternative bacterial expression systems

Bacterial hosts other than *E. coli* have received attention as developing host-vector expression systems due to their metabolic diversity and/or improved protein secretion capabilities, mainly owing to their ability to thrive in different environments, some of them extreme (Ferrer-miralles & Villaverde, 2013). In fact, the use of *E. coli* cell factories has been decreasing in the past 8 years (Bill, 2014a). Alternative bacterial expression systems include:

(i) *Lactococcus lactis,* a Gram-positive lactic acid bacterium, which has acquired GRAS status and is used for the intracellular or extracellular production of prokaryotic and eukaryotic interleukins (IL-2, IL-6, IL-12, and IL-10 from mice), enzymes (lipase from *Staphylococcus hyicus*, lysozyme from hen egg white, neutral protease from *Bacillus subtilisin* and others), antigens, and membrane proteins, among others (Loir *et al.*; Kuipers *et al.*, 1997; Bill *et al.*, 2011).

(ii) *Bacillus subtilis*, which is also a Gram-positive bacterium, is the most well-characterised Gram-positive bacterium. It has GRAS status, has high product yields (20-25 g/l), and is used for the large-scale fermentation of recombinant amylases, pectate lyase, and more recently, riboflavin (vitamin B2) for the food industry (Yin *et al.*, 2007; Dijl & Hecker, 2013).

(iii) *Streptomyces lividans*, which is an effective secretor of a recombinant cellulase from *Streptomyces. coelicolor* (up to 350 mg/l) and antigens from *Mycobacterium tuberculosis* (Dubé *et al.*, 2008; Anne *et al.*, 2012), (iv) *Corynebacterium glutamicum*, which is a secretor of recombinant protein-glutaminase (Kikuchi *et al.*, 2008).

(iv) *Pseudomonas aeruginosa*, which is an efficient secretor of Penicillin G acylase (Krzeslak *et al.*, 2009; Retallack *et al.*, 2011), among others. Both Gram-positive and Gram-negative systems have protein production limitations that can be avoided by using eukaryotic cells.

1.2.2 Yeast expression systems

1.2.2.1 Saccharomyces cerevisiae

This budding yeast is a versatile, single-cell form of fungi that offers definite advantages over *E. coli* as an effective protein production platform since the early 1980s (Hitzeman *et al.*, 1981; Bill, 2014a). It is one of the simplest forms of fungi belonging to the phylum Ascomycota. This microorganism is also referred to as Brewer's or Baker's yeast due of its use in the preparation of alcoholic beverages and bread, respectively. Its GRAS status has also enabled it to be used for the production of various pharmaceutical proteins, e.g.; following the success of Humulin™, an American biotechnology company, Novo Nordisk received approval to commercialise the first recombinant human insulin (Novolin®) produced in *S. cerevisiae* (Thim *et al.*, 1986; Bill, 2014a). Other recombinant commercial products made in *S. cerevisiae* include; human papilloma virus vaccines, α-acetolactate decarboxylase, Hepatitis B surface antigen, α-galactosidase, urate oxidase, glucagon, hirudin, platelet-derived growth factor, and granulocyte macrophage colony stimulating factor (Demain & Vaishnav, 2009; Ferrer-Miralles *et al.*, 2009; Martinez *et al.*, 2012; AMFEP, 2014). Major advantages of using *S. cerevisiae* for protein production include, (i) a short generation time compared to other eukaryotic hosts, (ii) established fermentation strategies, (iii) capability to grow to a high cell density, (iv) high product yield, (v) available annotated genome sequence, (vi) well-established vector systems and (vii) usage as a model organism to understand eukaryotic cellular mechanisms and metabolic pathways (Goffeau *et al.*, 1996; Cereghino & Cregg, 1999; Demain & Vaishnav, 2009; Favaro *et al.*, 2013).

Unlike *E. coli*, *S. cerevisiae* can secrete proteins devoid of toxic cell wall pyrogens or oncogenes into the culture medium. *S. cerevisiae* can also produce glycosylated recombinant proteins. However, the mode of glycosylation (both N- and O- linked conformations) in yeasts is restricted to the addition of mannose sugars, whereas mammalian proteins only require sialylated O-linked chains (Demain & Vaishnav, 2009). *S. cerevisiae* is known to hyper-glycosylate recombinant proteins by adding up to 200 α-1, 3-linked mannose residues to N-linked glycan structures (Conde *et al.*, 2004). Pharmaceutical proteins produced this way have a shortened *in vivo* half-life and low stability, and may cause an unwanted immunogenic response in patients (Gerngross, 2004). This form of glycosylation also affects folding, function, and translocation of the recombinant protein to the extracellular space (Needham *et al.*, 2011). Therefore, recombinant proteins produced in *S. cerevisiae* are often fused to an α-mating factor (α-MF1) secretion signal peptide to facilitate secretion into the culture medium, e.g.; the secretion of a single-chain antibody fragment (scFv) was higher (16-fold in the best case) when a mutant α-MF1 signal peptide was used (Rakestraw *et al.*, 2009). High-level protein production is often obtained by the use of expression vectors which integrate within the genome by homologous recombination. Even though these integrative vectors are available at lower copy numbers, compared to the autonomously replicating episomal vectors (up to 200 copies per cell), genome integration results in stable cultures that produce the recombinant protein at consistent levels (Partow *et al.*, 2010).

1.2.2.2 Alternative yeast host systems

Similar to *S. cerevisiae, Pichia pastoris* can perform posttranslational modifications, such as protein folding, proteolytic processing, disulphide bond formation and glycosylation. This yeast is fast becoming more popular than *S. cerevisiae* due to its easy cultivation in cheap mineral salt medium, efficient extracellular secretion and ability to grow at relatively higher cell densities (Sørensen, 2010; Gasser *et al.*, 2013). In addition, this yeast produces up to 350 different recombinant proteins for a variety of applications in both food and

pharmaceuticals (Yu *et al.*, 2010; Bill, 2014a). *P. pastoris* produced foreign proteins that undergo less extensive hyper-glycosylation and some strains have been glycoengineered to carry out human-type N-linked glycosylation (Hamilton *et al.*, 2003, 2006). This system is commercially available under the brand name, PichiaPink™ from Life Technologies Corporation (USA) (http://www.lifetechnologies.com). Other upcoming expression hosts include:

(i) *Schizosaccharomyces pombe*, which is referred to as a fission yeast, is capable of producing proteins that require challenging glycosylation arrangements and is currently used to study human G-protein coupled receptors (GPCRs) and olfactory receptors (Ladds *et al.*, 2003; Davey & Ladds, 2011; Emmerstorfer *et al.*, 2014). Efforts to understand its secretory pathway, develop protease-deficient strains and design new vector systems are underway (Idiris *et al.*, 2010; Matsuzawa *et al.*, 2013).

(ii) *Hansanuela polymorpha,* which is a thermo-tolerant methylotrophic yeast, has mutant strains that are glycoengineered to effectively produce complex human proteins e.g.; β-1,2-N-acetylglucosaminyltransferase I (GnTI) (Cheon *et al.*, 2012) at high temperatures.

(iii) *Kluyveromyces lactis*, which has currently been shown to over-produce the gene encoding the *Aequorea victoria* (jellyfish) enhanced green fluorescent reporter protein (EGFP) under the regulation of different promoters and/or signal peptides derived from various filamentous fungi (Madhavan & Sukumaran, 2014, 2015). On a commercial scale, *K. lactis* produces lactase (β-galactosidase) from a *Kluyveromyces sp.* and milk clotting enzyme (bovine chymosin) (AMFEP, 2014).

1.2.3 Filamentous fungal expression systems

Filamentous fungi are a diverse group of eukaryotes that are ubiquitous in nature. They are typically saprophytic microorganisms, growing on and deriving nourishment from dead or decaying organic matter. To do so, they naturally produce and secrete hydrolytic enzymes involved in decomposition and nutrient cycling of complex biopolymers from the

environment outside the growing hyphae (Ward, 2012). Many fungi can also be detrimental to the health of plants (as rusts, smuts, wilts, etc.), animals (as mycoses and mycotoxins) and other fungi (as mycoparasites) (Bennett, 1998; Sette *et al.*, 2013). In addition, they are varied in morphology, i.e.; they can exists as microscopic (e.g.; moulds and yeasts) or macroscopic (e.g.; mushrooms and truffles) forms (Sette *et al.*, 2013).

Their natural ability to produce proteins continues to play an important role in the industrial production of products from fungal and non-fungal origin, such as human therapeutics, organic acids, pigments, several food additives, and commercial enzymes (Saunders *et al.*, 1989; Punt *et al.*, 2002). For example, *Trichoderma reesei* can produce homologous hydrolases up to 100 g/l into the culture medium (Durand *et al.*, 1988; Saloheimo & Pakula, 2012). Among the diverse fungal species, some produce polysaccharides, bio-surfactants and additives used in the manufacturing of bread, cheese, alcoholic beverages and foodstuffs (e.g.; mushrooms, single cell protein/biomass (SCP) or lipid-rich biomass) (Ward, 2012). Their ability to thrive on simple and inexpensive substrates (Krull *et al.*, 2013) has made them popular as cell factories in the production and secretion of various extracellular enzymes (Nevalainen *et al.*, 2005; Sharma *et al.*, 2009; Ward, 2012). Their use has resulted in several species gaining GRAS status by the U.S. FDA (Conesa *et al.*, 2001). A large number of fungal enzymes are commercially available and their applications extend beyond their traditional uses in biopharmaceutical and food processing industries. Therefore, the development of new expression systems is a continuing desire for many biotechnology companies and research labs (Fowler & Berka, 1991; Punt *et al.*, 2002). Currently, the production of microbial enzymes is a rapidly growing sector of the fermentation industry (El-Enshasy *et al.*, 2007) and filamentous fungi are the sources for over 40% of the available industrial enzymes (Archer *et al.*, 1997; Lubertozzi & Keasling, 2009; Adrio & Demain, 2014). Most of these are hydrolysing enzymes used as aids in the bioprocessing of animal feed, detergent, textile, paper and pulp, biofuels, food and other industries (Fleißner & Dersch, 2010; Visser *et al.*, 2011; Adrio & Demain, 2014).

Extracellular protein production of proteins has several advantages over intracellular secretion. The secreted protein of interest into culture media usually remains induced longer and this is a preferred mode of expression if the protein produced is toxic to the host cell (Palomares *et al.*, 2004). In addition, working with extracellular proteins lessens the burden of difficult downstream processing steps, since post harvesting, and cell disruption steps are avoided. The protein produced is often active and properly folded, because protein folding requires oxidative conditions which occur outside the cytoplasm (Meyer & Schmidhalter, 2012).

The most commonly used fungi include commercial strains such as *Aspergillus niger*, *A. oryzae, Trichoderma reesei* and *Myceliophthora thermophila* (previously known as *Chrysosporium lucknowense* C1) (Visser *et al.*, 2011). These commercially successful host expression systems have all been whole-genome sequenced. This milestone has led to the use of genomics, proteomics, synthetic biology and metabolomics techniques to understand fungal physiology and to enhance protein productivity (Nevalainen & Peterson, 2014). For instance, the information obtained from the different fields of study could improve areas of bioprocessing, i.e.; discovery of novel enzymes, strain improvement strategies, understanding of secretory and metabolic pathways (Ward, 2012).

Aspergillus oryzae (also called, koji mold) is used for the production of fermented foods, e.g.; sake (rice wine), miso (soybean paste) and shoyou (soy sauce) in Asia (Machida, 2008). Its native enzyme, Taka-diastase was the first microbial enzyme to be patented (U.S. Patent 525,823) in 1894 (Ward, 2012). It is used for the production of recombinant enzymes, e.g.; glucose oxidase from other *A. niger spp.*, laccase from *Myceliophthora sp.* and *Polyporus sp.*, and endo-1,3(4)-β-glucanase from *Thermoascus sp*. (AMFEP, 2014). *Aspergillus. niger* is mainly used for the recombinant production of glucose oxidase from *Penicillium sp.*, lipase from *Candida sp.* and phospholipase from porcine pancreas (AMFEP, 2014). Both Aspergillus strains are capable of producing correctly folded recombinant human proteins, e.g.; lactoferrin, single-chain variable region fragment (scFv) and lysozyme

(Gomi *et al.*, 2000) by *A.* oryzae, and interferon-α-2 (MacRae *et al.*, 1993), mucus proteinase inhibitor (Mikosch *et al.*, 1996) and granulocyte macrophage colony stimulating factor (G-CSF; Davies, 1994) by *A. niger*.

Trichoderma reesei is used for the hyper-secretion of cellulases and other plant cell wall degrading enzymes (20-100 g/l; Seiboth *et al.*, 2011). These enzymes are important for the development of biofuels (Bouws *et al.*, 2008; Wilson, 2009; Schuster & Schmoll, 2010). In fact, it is a model organism for the regulation, biochemistry and expression of (hemi-) cellulose-degrading pathways and enzymes (Schuster & Schmoll, 2010; Jørgensen *et al.*, 2014). In addition, *T. reesei* is able to glycosylate in a manner similar to that in mammalian cells (Salovouri *et al.*, 1987). It is used to produce commercial enzymes, e.g.; xylanase from *Actinomadura sp.* and *Aspergillus sp.*, amylases, phospholipase B, catalase, α-glucanase, pectin lyase, pectin methylesterase or pectinesterase, phospholipase A2, phytase and polygalacturonase or pectinase, from *Aspergillus sp.*, cellulase from *Staphylotrichum sp.*, laccase from *Thielavia sp.* and phytase from *Buttiauxella sp.* (AMFEP, 2014).

Myceliophthora thermophila can hyper-produce a wide variety of extracellular homologous cellulases (mixture of cellobiohydrolases and endo-1,4-β-glucanases) up to 100 g/l, on par with *T. reesei* (Visser *et al.*, 2011). It has been used to produce heterologous fungal enzymes (xylanase, amylase, cellulase, endo-polygalacturonase, oxidase and phytase); bacterial enzymes (xylosidase, and cellulase) and a human protein (immunoglobulin, IgG1 against tumour necrosis factor (TNF-α) at g/l level. The protein was biologically active and not hyper-glycosylated (Visser *et al.*, 2011).

1.2.4 Overview of commonly used host expression systems

Table 1.1 shows different features that are considered important when selecting an expression system for the production of a heterologous protein.

Table 1.1. Comparison of the different expression systems used for the production of heterologous proteins (Saunders *et al.*, 1989; Yin *et al.*, 2007; Ferrer-Miralles *et al.*, 2009; Ferrer-miralles & Villaverde, 2013)

Characteristics	*E. coli*	*Bacillus sp.*	S. cerevisiae	*P. pastoris*	*Aspergillus sp.*	T. reesei	Insect cell lines	Mammalian cell lines	Plant cell culture
Duration of fermentation	2-3 days	2-3 days	4-5 days	4-5 days	5-7 days	5-7 days	5-7 days	15-20 days	> 30 days
Complexity of growth media	Minimum	Minimum	Minimum	Minimum	Minimum to complex	Minimum to complex	Complex	Complex	Complex
Cost of growth media	Low to high	Low to high	Low to high	Low to high	Low to high	Low to high	Low to high	High	High
Recombinant protein expression level	Low to high	Low to high	Low to high	Low to high	Low to high	Low to high	Low to high	Low to moderate	High
Destination of product	Extracellular/ intracellular	Extracellular/ intracellular	Extracellular/ intracellular	Extracellular	Extracellular	Extracellular	Extracellular	Extracellular	Extracellular
Ease of molecular cloning	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent	Good	Sufficient	Good
Up scaling	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent	Sufficient	Poor	Poor
Production cost	Low	Low	Low	Low to moderate	Moderate	Moderate	High	Highest	Moderate
Stage of development	Production	Production	Production	Production	Production/ development	Production/ development	Production/ development	Production	Research/ development
Post-translational Modifications									
Protein folding	Refolding usually required	Refolding usually required	Refolding might be required	Refolding might be required	Refolding might be required	Refolding might be required	Proper refolding	Proper refolding	Proper refolding
N-linked glycosylation	None	No	High mannose	Less mannose	Yes*	Yes*	Yes*	Yes	Yes*
O-linked glycosylation	No	No	Yes*	Yes*	Yes*	Yes*	Yes	Yes	Yes
Phosphorylation	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Acetylation	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Acylation	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
γ-carboxylation	No	No	No	No	No	No	No	Yes	Yes
Disulphide bridge	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Differs from mammalian;

1.3 Requirements for the development of a new host-vector system

The general features of a successful fungal gene expression system are similar to those developed for both yeasts and bacteria. The three basic components are: (i) a suitable host strain, (ii) an expression cassette (which comprises of regulatory elements originating from the host or other organism, fused to DNA/cDNA sequences that encode the protein of interest), and (iii) a suitable transformation method to introduce the expression vector inside the host (Berka & Barnett, 1989; Valencia & Chambergo, 2013a).

For the scope of this study, five major requirements for the development of a new-host vector system will be considered, i.e.; (i) identification of a new fungal isolate, (ii) identification of optimum growth conditions, (iii) discovery of regulatory gene regions for a prospective vector system, (iv) methods to improve the host, and (v) identification of a suitable transformation method.

1.3.1 Screening of fungal culture collections and identification of a high protein producing isolate

The fungal kingdom is estimated to contain at least 1.5 million species (Hawksworth, 2012) divided into four phyla (divisions): Ascomycota, Zygomycota, Basidiomycota, and Chytridiomycota. Filamentous (mould) fungi are mainly found in Ascomycota and Zygomycota. Most mushrooms belong to Basidiomycota; however, some are classified as ascomycetes, e.g.; truffles and morels. Fungi belonging to the phylum Chytridiomycota (also called chytrids) are known as the oldest and the only phylum that produces motile spores (zoospores) (Barr, 1978; Letcher & Powell, 2014).

The initial step towards designing a new fungal expression system involves the identification of a putative candidate with a natural ability to produce high levels of protein extracellularly. On average, 1,200 new fungal species are described yearly and most of these are preserved in microbial culture collections (Sette *et al.*, 2013). Culture collections may vary in size and

functions, i.e.; some are private, public or maintained by a research institution affiliated either to a university, national laboratory or non-profit organisation. These culture collections are great resources for the discovery of novel organisms and the compounds they produce (Smith, 2012). In this study, filamentous fungi were obtained from the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI) University of Pretoria, South Africa. The CMW culture collection houses more than 40,000 isolates which are obtained locally and globally from plant material, insects or soil for research purposes mainly related to tree health (http://www.fabinet.up.ac.za). For the scope of this study, only fungi indigenous to South Africa were screened.

Many screening programs have been developed to exclusively target a particular fungal extracellular hydrolytic enzyme, e.g.; lipase, cellulase, amylase, protease, chitinase, etc. (Hankin & Anagnostakis, 1975; Saritha *et al.*, 2012). This involves the cultivation of fungi on basal media supplemented with a carbon or nitrogen source (substrate) that can be degraded. After several days post inoculation, the plates are flooded with a dye that will detect a hydrolysis zone representing enzyme activity following destaining and washing steps. Potential isolates could then be selected on the diameter size of the clear zones surrounding the colonies. Quantitative assays can then be performed on positive isolates. For example, screening for lignocellulolytic fungi involves the cultivation on basal media containing carboxymethyl cellulose (CMC) as sole carbon source, and Congo red for detection of yellow clear zones against a red background of the dye (Teather & Wood, 1982; Khokhar *et al.*, 2013). The major drawback in using qualitative assays, is the lack of standardized methods for the detection of a particular enzyme (Pointing, 1999; Yuan *et al.*, 2011).

For this study, screening for an efficient protein producer is carried out in shake flasks containing malt extract liquid medium. Thereafter, proteomics techniques are used to select potential isolates. Extracellular proteins from culture supernatants are precipitated and separated on SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) gels. Promising fungal candidates are then chosen on the basis of observed dark-blue Coomassie-

stained bands indicating highly expressed proteins. Protein bands are excised and sent for liquid chromatography-tandem mass spectroscopy (LC/MS-MS) peptide sequencing. It is often necessary to separate proteins of interest with 2-D gel electrophoresis to obtain better separation of proteins prior to sequencing, due to separation of proteins in two dimensions (>200 protein spots) by size and isoelectric point (Jun *et al.*, 2011). However, due to the small number of bands observed in the culture supernatants and only highly concentrated proteins being targeted in this study; 1-D gel electrophoresis was sufficient for separating the proteins by size only.

In this study, the isolate *Clonostachys rosea* was identified as a potential fungal candidate for protein production. *C. rosea* (Link: Fr.) Schroers, Samuel, Seifert, & Gams; formerly known as *Gliocladium roseum,* is the asexual stage and often preferred name of the telemorph *Bionectria ochroleuca* (Schw.) Schroers & Samuels (Schroers *et al.*, 1999). *C. rosea* belongs to the order Hypocreales and the family *Bionectriaceae*. *C. rosea* is a common facultative soil saprophyte that is found in a broad range of habitats in tropical, temperate, sub-arctic and desert regions, particularly soils with pH values of neutral to alkaline (Sutton *et al.*, 1997; Schroers, 2001). Although it is naturally found in soil, it also colonizes plant debris (e.g.; leaves, roots, stems, and seeds) (Sutton *et al.*, 1997). It has been reported as a potential biocontrol agent against fungal plant pathogens and parasitic nematodes, which infect plants and animals (Lahlali & Peng, 2014).

It has also been shown to suppress disease symptoms of clubroot caused by *Plasmodiophora brassicae* in canola roots (Lahlali & Peng, 2014), leaf and head rot caused by *Pythium tracheiphilum* in Chinese cabbage (Møller *et al.*, 2003) and grey mould/Botrytis blight caused by *Botrytis cinerea* in several crops e.g.; strawberry, rose and eucalyptus (Sutton *et al.*, 1997; Cota *et al.*, 2009). To suppress infections by other fungi (mycoparasites), this fungus releases cell-wall degrading extracellular hydrolytic enzymes such as chitinases, glucanases, and volatile secondary metabolites to carry out mycoparasitism (Roberti *et al.*, 2008; Chatterton & Punja, 2009; Kosawang *et al.*, 2014). As a nematophagous fungus, *C. rosea* can trap and

suppress infections caused by plant or animal associated nematode diseases by releasing cuticle-degrading proteases, i.e.; serine proteases (Yang *et al.*, 2007; Liang *et al.*, 2011; Ahmed *et al.*, 2014).

The ability of *C. rosea* strains to produce extracellular homologous hydrolytic enzymes has contributed to their involvement in the development of alternative biological control agents (BCAs) (Lahlali & Peng, 2014). For example, *Clonostachys rosea* f. *catenulata* (*Gliocladium catenulatum*) strain J1446 is commercially available as Prestop WP (Verdera Oy, Finland) and Prestop Mix (Verdera Oy, Finland) as a biological control agent against phytopathogenic fungi (McQuilken *et al.*, 2001; Chatterton *et al.*, 2008; Lahlali & Peng, 2014). Another biocontrol agent based on *C. rosea* (strain IK726) is currently under development (Jensen *et al.*, 2007; Kosawang *et al.*, 2014).

1.3.2 Selection of culture conditions for optimum protein production

Filamentous fungi occupy a wide range of habitats in nature and they require specific growth conditions for product (protein) formation. A wide range of growth media are available to cultivate filamentous fungi in the laboratory. Fungi require water, molecular oxygen, organic carbon, hydrogen, a source of nitrogen and inorganic elements (e.g.; magnesium, iron, zinc, copper and molecular nitrogen, among others) for growth (Papagianni, 2004). In industry, filamentous fungi are grown in submerged (liquid) cultures to produce desired bioproducts. This is contrary to how they thrive in terrestrial ecosystems where they live off of solid nutrients (Demir *et al.*, 2011; Barrios-González, 2012). For instance, the production of a heterologous protein, chymosin was increased by 500-fold when *A. oryzae* was grown on wheat bran in comparison to submerged fermentation (Elinbaum *et al.*, 2002). Hence, it would be expected that most commercial strains would be grown under solid-state fermentation (SSF); however the opposite is usually used in industry (Demir *et al.*, 2011). Submerged fermentation (SmF) is still preferred because it is easier to control physical parameters (pH, temperature, moisture,

nutrients, and aeration), carry-out scale-up experiments in bioreactors and purify the produced protein (Sarrouh *et al.*, 2012).

The composition of the culture medium for submerged fermentation is perhaps the most fundamental step, because it influences protein production and growth morphology (Ahamed & Vermette, 2009). Therefore, once a suitable fungal isolate is identified, optimum growth conditions should be selected. Most studies test numerous types of complex media containing different concentrations of carbon sources (D-glucose, maltose, sucrose, or D-fructose), nitrogen sources (ammonia, nitrate, amino acids, proteins, yeast extract, or whey powder), phosphates, and trace elements (Ahamed & Vermette, 2009; Posch *et al.*, 2013a). These experiments are often carried out by varying one parameter at a time according to the isolate and protein produced. In addition, the ideal physical parameters can also be selected to optimize protein formation (El-Enshasy *et al.*, 2007). The resulting protein production could then be compared by proteomic analysis and/or quantitative protein assays.

1.3.3 The effect of growth morphology on protein production

Filamentous fungi are composed of thread-like branching structures (termed hyphae) and referred collectively as mycelia. In many species, the hyphae are separated by cross-walls (septa). Alternatively, filamentous fungi may exist in nature as spores (conidia) (Ward, 2012). A typical lifecycle for an industrial competent strain involves asexual hyphae and conidia that germinate forming more vegetative hyphae. However, in rare occurrences, the production of sexual spores that undergo meiosis may also occur (Nevalainen & Peterson, 2014). Protein secretion has been shown to occur mainly at the tips of growing hyphae due to the presence of organelles (endoplasmic reticulum (ER) and Golgi apparatus) that are involved in protein secretion (Punt *et al.*, 1994; Wang *et al.*, 2005). However, a study has shown that the secretion of enhanced GFP-fused α-amylase (AmyB-EGFP) in *A. oryzae* was detected at septa, in addition to hyphal tips (Hayakawa *et al.*, 2011; Shoji *et al.*, 2014).

When submerged, filamentous fungi can grow either as pelleted or dispersed (further divided into freely dispersed hyphae and/or clumps) forms of mycelium (Grimm *et al.*, 2005). Pellets, filaments and clumps may vary in shape, size and appearance (either smooth or 'hairy') (Gibbs *et al.*, 2000; Porcel *et al.*, 2005; Peciulyte *et al.*, 2014). The favoured growth morphology greatly depends on the type of protein required. For instance, *A. niger* pellets are preferred for the fermentation of secondary metabolites (e.g.; citric acid) and freely dispersed filaments are recommended for enzyme production (e.g.; pectin) (Papagianni, 2004). Other factors which influence growth morphology include the type of initial culture inoculum (e.g.; spores, pellets, dispersed mycelium), agitation speed, pH, and the composition of the growth medium (Porcel *et al.*, 2005; Ahamed & Vermette, 2009; Krull *et al.*, 2013; Li *et al.*, 2013). The preferred production morphology can affect mass transfer, available oxygen concentration, nutrient transport and uniform mixing of the culture in bioreactors (Grimm *et al.*, 2005; Wucherpfennig *et al.*, 2010; Posch *et al.*, 2013b); which altogether influence protein productivity.

Freely dispersed filaments have been reported to largely increase the viscosity of the culture medium. In addition, the growth form may be more susceptible to the shear force in large bioreactors The increase of viscosity can affect aeration and mixing properties of the culture, thus reducing protein production (Gibbs *et al.*, 2000; Žnidaršič & Pavko, 2001; Papagianni, 2004; Visser *et al.*, 2011). In contrast, pelleted morphology results in lower viscosities and require substantially less power for mixing and aeration of the culture. However, pellets may restrict oxygen and nutrient intake by fungal cells (Grimm *et al.*, 2005). Efforts to manipulate and control the type of growth morphology to improve protein production have been reported. The viscosity of filaments by the *M. thermophila* parental strain (NG7C-19) was successfully reduced (up to 50-fold) when using a mutant strain (HC) produced by physical mutagenesis. In addition, the use of HC resulted in a 3-fold increase than NG7C-19 (Visser *et al.*, 2011). Other ways to modify morphology include, changing the culture pH (Wucherpfennig *et al.*, 2010), customizing the initial inoculum concentration (Papagianni, 2004), altering the growth medium by adding microparticles (e.g.; talc, aluminium and titanate) (Driouch *et al.*, 2010,

2012), and regulation of genes associated with pellet and filamentous morphology (Dai *et al.*, 2004).

1.3.4 Designing vector systems for high level protein expression

Briefly, the central dogma of molecular biology specifies that protein production is initiated by transcription followed by the translation of mRNA which leads to the assembly of amino acids into polypeptides. It is important to control the efficiency of both transcription and translation at the DNA level in order to produce high amounts of heterologous proteins. Hence, to express any protein of interest, it is essential to construct an expression system in which the DNA sequence of the protein is inserted. The basic features of a filamentous fungal expression vector system (Berka & Barnett, 1989 and Valencia & Chambergo, 2013) include:

(i) A strong promoter and upstream regulatory regions.

(ii) DNA sequences encoding a signal peptide.

(iii) DNA sequences encoding the target gene.

(iv) Downstream sequences for transcription terminator and polyadenylation signals.

(v) A transformation selectable marker for screening positive transformants

All components are usually assembled in a vector backbone/plasmid (e.g.; pUC19, pBluescriptKS(-), etc.) which can replicate in an appropriate *E. coli* strain for easy preparation of the DNA (Richey *et al.*, 1989; Lv *et al.*, 2012; Schoberle *et al.*, 2013). The order in which each sequence element is arranged on the expression vector system is shown in Figure 1.1.

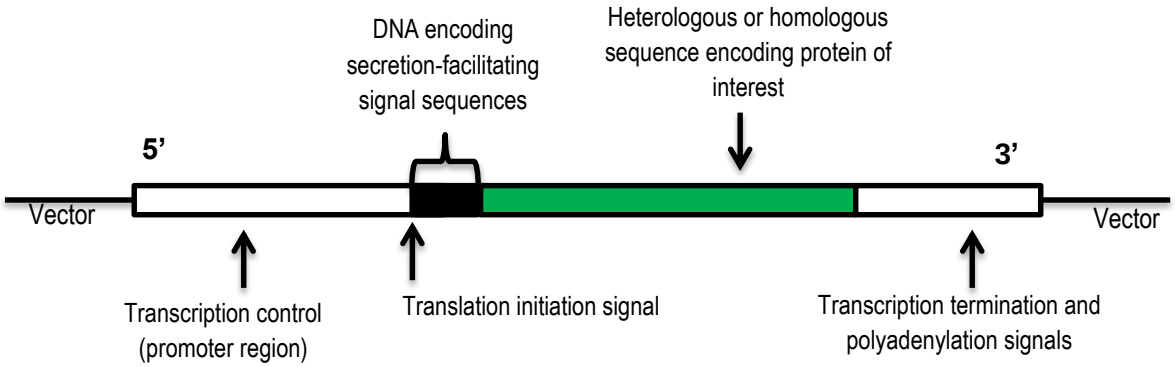


Figure 1.1. Schematic representation of a typical filamentous fungus expression cassette adapted from Berka & Barnett (1989) and Valencia & Chambergo (2013)

1.3.4.1 Promoters and the role of regulatory elements

The efficient production of a homologous or heterologous protein relies on the operation of strong transcription control elements (e.g.; promoter and transcription factors). In many experiments, the promoter region of a highly expressed fungal gene has been used to transcribe and drive the expression of a gene of interest (Andrie *et al.*, 2005; Li *et al.*, 2012; Madhavan & Sukumaran, 2014). Well-studied filamentous fungal promoters may either be constitutive or inducible.

Constitutive promoters are derived from genes associated with metabolic pathways, i.e.; glycolytic genes or housekeeping genes, which are continuously transcribed without the use of inducers. These promoters may also be active in growth media that contains high amounts of glucose (Li *et al.*, 2012). The *gpdA* gene encodes an enzyme, glyceraldehyde-3-phosphate dehydrogenase (GPDA) that is involved in glycolysis in *A. nidulans* (Punt *et al.*, 1990). This gene is constitutively expressed and its promoter (P*gpdA*) has been used to initiate the transcription of numerous heterologous genes (e.g.; EGFP, *E. coli* β-galactosidase (*lacZ*) and β-glucuronidase (*uiaA*), *T. reesei* β-1,4-xylanase 2 *(xyn2) and* β-1,4-endoglucanase I *(egl)*) using *A. nidulans* as a host (Punt *et al.*, 1991; Rose & Van Zyl, 2002; Meyer, Wu, *et al.*, 2011). Alternatively, P*gpdA* has been successfully used to produce recombinant proteins in other hosts, e.g.; human interleukin-6, hen egg-white lysozyme (HEWL) or *Trametes versicolor* laccase in *A. niger* (Archer *et al.*, 1990; Bohlin *et al.*, 2006; Fleißner & Dersch, 2010).

Currently, the commonly used promoters are inducible. They allow the control and manipulation of protein expression in the presence of inducers. These promoters are preferred over constitutive promoters for the production of toxic or unstable proteins (Radzio *et al.*, 1997). Inducible promoters include the *A. niger* α-glucoamylase encoding gene (P*glaA*) (Brunt, 1986), *A. nidulans* alcohol dehydrogenase encoding gene (P*alcA*/P*alcR*) (Gwynne *et al.*, 1989), *M. thermophila* cellobiohydrolase I encoding gene (P*chb1*) (Visser *et al.*, 2011) and the popular *T. reesei* cellobiohydrolase I encoding gene (P*chb1*) (Harkki *et al.*, 1991). Similar to

P*gpdA*, these promoters can function in other fungal species of the same genus or different genus. For example, P*chb1* from *T. reesei* can be functional in *T. harzianum* and in the yeast *K. lactis* when induced with lactose (Margolles-Clark *et al.*, 1996; Madhavan & Sukumaran, 2014).

The expression of genes by these promoters requires the presence of particular transcription factors (Ballance, 1986; Berka & Barnett, 1989; Sibthorp *et al.*, 2013). For example, P*chb1* is regulated by several conserved transcription factor motifs within the DNA sequence. These cis-acting elements are short DNA sequences that bind to consensus regions associated with carbon source utilization, e.g.; 5'-SYGGRG-3' (Cre1 binding site for carbon catabolite repression; Strauss *et al.*, 1995) , 5'-AGGCA-3' (ACEI binding site for cellulose and xylan utilization; Aro *et al.*, 2003) and 5'-GGCTAATAA-3' (ACEII binding site for cellulose and xylan utilization, Aro *et al.*, 2001), among others, are involved in the regulation of P*chb1* (Su *et al.*, 2011).

The identification and isolation of promoters involves the use of different molecular techniques. The 5' untranslated region of a highly-expressed gene may be isolated by screening DNA/cDNA libraries using oligonucleotide probes (Vanhanen *et al.*, 1989; Chen *et al.*, 2010), chromosome walking (Tonooka & Fujishima, 2009; Leoni *et al.*, 2011), or transcriptional profiling of genes induced during different growth conditions using quantitative real-time reverse-transcription polymerase chain reaction (RT-qPCR) (Li *et al.*, 2012) or sequencing of the whole transcriptome (Sibthorp *et al.*, 2013). Reverse genetics of strongly expressed protein(s) using proteomic and PCR-based techniques (Lim *et al.*, 2001; Nevalainen & Peterson, 2014) have also been successful. This method is often preferred because it is understood that the presence of a highly expressed protein in culture supernatants could be produced from a gene possessing a strong promoter.

Once isolated, the nucleotide sequence data could then be analysed *in silico* based on a Basic Local Alignment Search tool (BLAST) search (Altschul *et al.*, 1990; Johnson *et al.*, 2008) for

promoters with shared homology. Similarly, putative transcription factor binding sites can be predicted by using different web-based bioinformatics tools based on statistical algorithms (Sinha & Tompa, 2003; Bodén & Bailey, 2008; Buske *et al.*, 2010; Coradetti *et al.*, 2012). Thereafter, the isolated putative regulatory elements could be characterised by using various experimental techniques which can analyse and measure their functional activity.

Two strategies that are well-published include, northern blot analysis of mRNA levels of the putative regulatory sequences using oligonucleotide probes and the use of reporter gene assays (Visser *et al.*, 2011). Briefly, these assays can measure the regulatory activity of numerous lengths of the putative promoter region fused to a reporter gene of interest (e.g.; β-galactosidase (GUS), β-glucuronidase, GFP and luciferase) in a plasmid followed by detection tests (i.e.; fluorescence absorbance and/or luminescence) to measure the expression of the reporter gene used (Wood, 1995; Zorzano *et al.*, 2012). Therefore, the identification and use of strong promoters and their associated regulatory motifs could contribute towards the development of a new expression system, understanding gene regulation and improving fungal genome annotation (Coradetti *et al.*, 2012; Nevalainen & Peterson, 2014).

1.3.4.2 Signal peptides

Proteins produced extracellularly are trafficked and facilitated through the secretory pathway and out of the cell by a secretion signal peptide (also referred to as signal sequence, leader sequence or targeting signal) (von Heijne, 1990; Kapp *et al.*, 2009). Most eukaryotic signal peptides are located at the N-terminus of the preprotein and are usually cleaved off by a signal peptidase. They vary in length from 15 to 50 amino acid residues. Signal peptides are made up of a positively charged N-terminal (n-region), a hydrophobic core (h-region) consisting of 7 to 16 amino acid residues and a C-terminal (c-region) which consists of 4-6 relatively neutral and polar residues (Kapp *et al.*, 2009). Signal peptide sequences are not the same for all organisms, however the order of the n-, h- and c- regions and function is often conserved (Hegde & Bernstein, 2006; Zorzano *et al.*, 2012).

The N-terminal signal sequence amino acid residues are identified from an open reading frame (ORF) encoding an extracellular protein using web-based bioinformatics programmes, e.g.; SignalIP (http://www.cbs.dtu.dk/services/signalP) which predicts amino acid residues based on a statistical model that identifies a signal peptidase sequence within the ORF which can differentiate a putative signal peptide from the rest of the protein (Petersen *et al.*, 2011).

Majority of experiments have used N-terminal signal sequences of extracellular hydrolases to facilitate the secretion of homologous and/or heterologous genes in filamentous fungi. For example, the signal sequence of the cellobiohydrolase I (chb1) protein from *T. reesei* is often incorporated in expression vectors to facilitate the secretion of the gene product into the culture medium (Schuster & Schmoll, 2010).

1.3.4.3 Target gene sequence

The heterologous DNA sequence due to be cloned and expressed into a filamentous fungal host or vector may be a complete gene, i.e.; containing the gene regulatory, exonic and/or intronic sequences originating from the organism from which it was isolated. Alternatively, it can be in the form of cDNA sequence derived by reverse transcription of mRNA. This form of the heterologous gene will be devoid of regulatory sequences and introns (Curran & Begeja, 2011).

Although filamentous fungi and several yeasts are able to recognize foreign transcriptional elements and effectively remove exons from heterologous genes, many commercial proteins are produced from the insertion of a cDNA sequence into an appropriate expression system (Curran & Begeja, 2011; Finkelstein, 2013). Table 1.2 shows some of the different genes that have been expressed in filamentous fungi.

Table 1.2: Examples of heterologous genes expressed in *T. reesei* (Valencia & Chambergo, 2013b)

Host	Vector	Promoter used	Heterologous gene	Product yield	Reference
T. reesei	pGFP pGFP-HFBI	P*chb1*	*GFP* (green fluorescent protein) from *Aequorea victoria*	52.2 mg/l	Mustalahti *et al.* (2011)
T. reesei	pWEF31 pWEF32	P*chb1*	*DsRed2* (red fluorescent protein) from *Discosoma sp.*	active	Lv *et al.* (2012)
T. reesei	pXBthg-pbgl1	P*chb1*	*bgl1* (β-glucosidase I) from *Penicillium decumbens*	101 IU/mg	Ma *et al*. (2011)
T. reesei	pSK-Lip	P*chb1*	*Lip* (lipase) from *A. niger*	310 mg/l	Qin *et al.* (2012)
T. reesei	PTrCHB-EPO	P*chb1*	*HuEPO* (erythropoietin) from *Homo sapiens*	46-97 mg/l	Zhong *et al.* (2011)

1.3.4.4 Terminators and polyadenylation signals

Whereas strong promoter elements aim to mediate and increase the rate of gene transcription; the terminator region (located at the 3' untranslated region of a gene) is capable of facilitating the end of transcription during which the cleavage of 3'-mRNA and poly(A) addition occurs. In addition, transcription terminators can affect the stability, localization, translational efficiency and half-life of the heterologous mRNA, which may altogether increase or decrease protein productivity (Berka & Barnett, 1989; Radzio *et al.*, 1997; Su *et al.*, 2011; Ito *et al.*, 2013). Similar to promoters, the terminators used in expression vectors are often derived from overexpressed native fungal genes, including *chb1* from *T. reesei* and *trpC* (encodes a tryptophan biosynthesis gene) from *A. nidulans*. As a result, the terminators T*chb1* (from *chb1*) and T*trpC* (from *trpC*) from *T. reesei* and *A. nidulans* respectively, are often placed downstream of the gene of interest in an expression cassette (Berka & Barnett, 1989; Saunders *et al.*, 1989; Fleißner & Dersch, 2010).

While most published work is focused on the isolation, characterisation and engineering of filamentous fungal promoters, less work has been carried out to understand the effect of fungal terminators, mRNA processing elements, and polyadenylation sites on heterologous protein production (Su *et al.*, 2011). However, studies on yeast terminators have been increasing over the years. Most of these studies involve the identification of conserved sequence motifs which are related to yeast mRNA processing sites using bioinformatics tools (Graber *et al.*, 1999,

2002; Curran *et al.*, 2015) and characterisation of 'strong' native terminators using reporter gene fusions in order to engineer synthetic terminators (Curran *et al.*, 2013, 2015; Ito *et al.*, 2013; Yamanishi *et al.*, 2013). Therefore, it is often understood that similar observations and conclusions may be relevant for filamentous fungi.

1.4 Strategies for strain improvement

The production level of proteins in naturally occurring strains is usually too low for commercial use. High-yielding industrial strains can be produced by modifying culture conditions or introducing changes in their DNA sequence by mutation, genetic recombination or genetic engineering strategies (Punt *et al.*, 2002; Nguyen *et al.*, 2015). Classical mutagenesis and subsequent strain selection programs have successfully produced mutants with increased protein production (Rowlands, 1984; Saunders *et al.*, 1989). These strategies involve the application of random mutagenesis, during which a microorganism is exposed to one or a combination of physical mutagens (UV irradiation) and/or chemical mutagens (ethyl methanesulphonate (EMS), ethidium bromide (EtBr), and N-methyl- N'-nitro-N-nitrosoguanidine (NTG)), followed by a selection of surviving strain variants according to a specific phenotypic characteristic, i.e.; a hyper-producing strain with improved growth morphology (Parekh *et al.*, 2000; Peterson & Nevalainen, 2012; Ramzan *et al.*, 2013).

UV-irradiation is frequently used as it is efficient and it is easy to take effective safety precautions against it compared to gamma rays and nitrosoguanidine (Rani & Prasad, 2012). Many hyper-producing industrial strains currently in use have undergone several rounds of mutagenesis. The classical example is the significant 500-fold increase of penicillin production by *Penicillium chrysogenum* since its original isolation (Rowlands, 1984; Peñalva *et al.*, 1998). Mutagenesis has also been used along with knock-out or knock-in technology to produces strains with low protease activity in order to reduce the proteolytic degradation of recombinant proteins (Nevalainen *et al.*, 2003; Visser *et al.*, 2011; Nevalainen & Peterson, 2014). These

efforts have played a role in increasing protein yields and subsequently reducing the cost of production.

1.5 Transformation of filamentous fungi

Transformation strategies are important to perform genetic manipulation of different organisms, including filamentous fungi. This method involves the introduction of heterologous DNA into host cells. The DNA may either be integrated into the host genome or maintained as an autonomously-replicating plasmid (Tilburn *et al.*, 1983; Fincham, 1989; Ruiz-Díez, 2002). Filamentous fungi usually integrate incoming heterologous DNA in an ectopic manner, rather than homologous recombination. The site of integration may occur randomly or at a region that is homologous to the expression vector (Berka & Barnett, 1989; Ruiz-Díez, 2002). Current methods used for transforming filamentous fungi include, biological methods (protoplast-mediated transformation using various cell wall degrading enzymes and *Agrobacterium tumefaciens*-mediated transformation) and physical methods (electroporation, shock waves, biolistic, vacuum infiltration and agitation with glass beads) (Olmedo-Monfil *et al.*, 2004; Jiang *et al.*, 2013; Rivera *et al.*, 2014). Every method has its advantages and limitations, however, the classical protoplast-mediated transformation is the most preferred method.

Expression vectors often contain a transformation selection marker sequence which enables the screening and visualization of positive transformants and to control plasmid stability in the host (Rai & Padh, 2001; Meyer, Wanka, *et al.*, 2011; Anyaogu & Mortensen, 2015). Selection markers may either confer resistance to a particular antibiotic (e.g.; hygromycin B and phleomycin) (referred to as a dominant marker) or complement a nutritional deficiency (auxotrophy) in a host strain (referred to as an auxotrophic marker) (Moore, 2007; Visser *et al.*, 2011; Schoberle *et al.*, 2013; Rivera *et al.*, 2014).

1.6 Research aims and objectives

The major goal of this study is to identify an indigenous wild-type fungal isolate from the CMW culture collection with the natural ability to produce high levels of extracellular protein(s).

The objectives of the study are listed below:-

1. Screen the existing CMW culture collection for an isolate that produces high levels of an extracellular protein
2. Identify the highly expressed protein
3. Optimize growth conditions for high level protein production
4. Isolate and amplify the gene encoding the identified protein
5. *In silico* identification of regulatory gene elements associated with the gene

CHAPTER 2: GENERAL MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemical reagents

All chemical reagents used were of analytical or molecular biology grade and obtained from a range of suppliers.

Table 2.1. List of chemicals used in the study

Chemicals	Supplier
10000× GelRed™	Biotium, USA
Acetone	Merck-Saarchem, South Africa
Acrylamide/bis-acrylamide, 30% solution	Sigma-Aldrich, USA
Agar	Difco, USA
Ammonium nitrate (NH_4NO_3)	Sigma-Aldrich, USA
Ammonium persulphate (APS)	Sigma-Aldrich, USA
Bacto-peptone	Merck, Germany
Beta-mercaptoethanol (βME)	Sigma-Aldrich, USA
Big Dye® Terminator v3.1	Applied Biosystems, USA
Bromophenol blue	Sigma-Aldrich, USA
Calcium chloride ($CaCl_2$)	Merck-Saarchem, South Africa
Coomassie brilliant blue R250	Sigma-Aldrich, USA
Dextrose	Sigma-Aldrich, USA
D-glucose	Merck, Germany
Diaminoethane tetraacetic acid (EDTA)	Merck-Saarchem, South Africa
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, USA
Dithiothreitol (DTT)	Sigma-Aldrich, USA
Ethanol	Illovo Sugar, South Africa
Ferrous sulphate heptahydrate ($FeSO_4 \cdot 7H_2O$)	Merck-Saarchem, South Africa
Gelatin	Carl Roth chemicals, Germany
Glacial acetic acid	Merck-Saarchem, South Africa
Glycerol	Merck-Saarchem, South Africa
Glycine	Merck-Saarchem, South Africa
Hydrochloric acid (HCl)	Merck-Saarchem, South Africa
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Sigma-Aldrich, USA
LB agar	Lab M Limited, UK
LB broth	Biolab Merck, South Africa
Magnesium chloride hexahydrate ($MgCl_2 \cdot 6H_2O$)	Merck, Germany
Magnesium sulphate heptahydrate ($MgSO_4 \cdot 7H_2O$)	Merck, Germany
Magnesium sulphate, anhydrous ($MgSO_4$)	Sigma-Aldrich, USA
Malt extract	Biolab Merck, South Africa
Manganese chloride ($MnCl_2$)	Merck-Saarchem, South Africa
Methanol	Merck-Saarchem, South Africa
3-(N-morpholino) propanesulphonic acid (MOPS)	Merck-Saarchem, South Africa
Potassium acetate	Merck-Saarchem, South Africa
Potassium chloride (KCl)	Sigma-Aldrich, USA
Potassium hydrogen phosphate, anhydrous (K_2PO_4)	Sigma-Aldrich, USA
Potassium hydroxide (KOH)	Merck, Germany
Potassium nitrate (KNO_3)	Merck, Germany
Potato dextrose broth	Difco, USA
Rubidium chloride (RbCl)	Merck-Saarchem, South Africa
Saccharose	Merck-Saarchem, South Africa

Chemicals	Supplier
SeaKem LE Agarose	Lonza, USA
Sodium chloride (NaCl)	Sigma-Aldrich, USA
Sodium hydroxide (NaOH)	Sigma-Aldrich, USA
Sodium lauryl sulphate (SDS)	Sigma-Aldrich, USA
Sodium nitrate ($NaNO_3$)	Sigma-Aldrich, USA
Sucrose	Merck-Saarchem, South Africa
SYBR® Safe	Life Technologies, USA
Trichloroacetic acid (TCA)	Merck, Germany
Tris (hydromethyl) aminomethane (Tris base)	Merck, Germany
Tryptone powder	Merck, Germany
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal)	Sigma-Aldrich, USA
Yeast extract	Difco, USA

2.1.2 Buffers and solutions

Table 2.2. Buffers and solutions used in this study

Buffers and solutions	Content/Manufacturer	Purpose
0.5 M EDTA (pH 8.0)	37.22% [w/v] EDTA, adjust pH with 0.1 M NaOH	DNA/RNA storage buffer
0.5 M Tris-HCl (pH 6.8)	6% [w/v] Tris base, adjust pH with 0.1 M HCl	SDS-PAGE
1 × Sample buffer	25% [v/v] 0.5 M Tris-HCl (pH 6.8), 20% [v/v] glycerol, 4% [w/v] SDS, 0.012% [w/v] Bromophenol blue, 3.1% [w/v] DTT or 4% [v/v] βME, store aliquots at - 80°C	SDS-PAGE
1 M Tris (pH 8.0)	12.14% [w/v] Tris base, adjust pH with 0.1 M HCl	DNA storage Buffer
1.5 M Tris-HCl (pH 8.8)	18.15% [w/v] Tris base, adjust pH with 0.1 M HCl	SDS- PAGE
1× TAE (pH 8)	0.2% [w/v] Tris base, 0.5% [v/v] glacial acetic acid, 1% [v/v] 5 M EDTA	Agarose gel electrophoresis
10 mM Tris-EDTA (TE) buffer	1% [v/v] 1M Tris-HCl (pH 8.0), 0.2% [v/v] 0.5 M EDTA (pH 8), 99.2% [v/v] distilled water (dH_2O), adjust pH to 8.0 for DNA. Autoclave to sterilize, aliquot and store at - 20°C	DNA/RNA storage buffer
10% APS	10% [w/v] APS	SDS-PAGE
20% TCA solution	20% [w/v] TCA	Protein precipitation
6 × Sample buffer	70% [v/v] 0.5 M Tris-HCl (pH 6.8), 30% [v/v] glycerol, 10% [w/v] SDS, 0.012% [w/v] Bromophenol blue, 9.3% [w/v] DTT or 6% [v/v] βME, store aliquots at -80°C	SDS-PAGE
6× GelRed™-loading buffer solution	30× GelRed™ added into 5 ml of 6× Loading buffer	Co-loading for band visualization
6× Loading buffer	0.25% [w/v] Bromophenol blue, 40% [v/v] glycerol	Agarose Gel electrophoresis
Coomassie stain	0.125% [w/v] Coomassie brilliant blue R250, 50% [v/v] methanol, 10% [v/v] Glacial acetic acid	SDS-PAGE staining
Destaining solution	10% [v/v] Methanol, 10% [v/v] Glacial acetic acid	SDS-PAGE destaining
IPTG	0.1 M, filter-sterilized. Stored at - 20°C	Blue-white screening
SDS solution	10% [w/v] SDS	SDS-PAGE
SYBR® Safe	Life Technologies (USA)	DNA Gel Stain

Buffers and solutions	Content/Manufacturer	Purpose
X-Gal	2% [w/v] dissolved in DMSO, covered in foil. Store at - 20°C	Blue-white screening
1× Running buffer	0.25 mM Tris-HCl, 2 M glycine, 1% [w/v] SDS	SDS-PAGE

2.1.3 Media used in this study

A. Cultivation media for *E. coli*

i. Luria broth (LB) (per litre)

Tryptone powder	10 g
Yeast extract	5 g
NaCl	10 g

ii. SOB broth (per litre)

Tryptone powder	20 g
Yeast extract	5 g
NaCl	0.58 g
KCl	0.19 g
After autoclaving add	
1M $MgCl_2 \cdot 6H_2O$	10 ml
1 M $MgSO_4 \cdot 7H_2O$	10 ml

iii. SOC (per litre)

Add 3.60 g of D-glucose in SOB

The components listed were dissolved in dH_2O and portioned into flasks. The flasks were closed by cotton plugs, covered with foil and autoclaved at 121°C for 20 min. The sterile media was stored at room temperature or 4°C. Required antibiotics were added before usage.

B. Cultivation media for fungal isolates

The different growth media used for protein production are listed in Table 2.3.

Table 2.3. List of liquid media used for protein production

Media	Composition
Carboxymethyl cellulose (CMC)	1.5 (w/v) carboxymethyl cellulose sodium salt; 0.1% (w/v) NH_4NO_3; 0.1% (w/v) K_2HPO_4; 0.05% (w/v) $MgSO_4 \cdot 7H_2O$; and 0.1% (w/v) yeast extract
Casein-rich (CR)	0.5% (w/v) Casein; 0.1% (w/v) D-glucose; 0.01% (w/v) $(NH_4)_2SO_4$; 0.05% (w/v) $MgSO_4 \cdot 7H_2O$; 0.02% (w/v) KH_2PO_4; and 0.0001% (w/v) $FeSO_4 \cdot 7H_2O$
Czapek-Dox (CZK)	0.3% (w/v) $NaNO_3$; 1.0% (w/v) K_2HPO_4; 0.05% (w/v) KCl; 0.05% (w/v) $MgSO_4 \cdot 7H_2O$; 0.001% (w/v) $FeSO_4 \cdot 7H_20$; and 3% (w/v) sucrose
Gelatin-rich (GR)	0.5% Gelatin; 0.1% (w/v) D-glucose; 0.01% (w/v) $(NH_4)_2SO_4$; 0.05% (w/v) $MgSO_4 \cdot 7H_2O$; 0.02% (w/v) KH_2PO_4; and 0.0001% (w/v) $FeSO_4 \cdot 7H_2O$

Media	Composition
Malt extract (ME)	2% (w/v) malt extract (from malted barley containing carbohydrates, vitamins, and peptides)
Malt-yeast extract (MYE)	0.3% (w/v) yeast extract; 1% (w/v) D-glucose; 0.3% (w/v) malt extract; and 0.5% (w/v) bactopeptone
Potato D-glucose (PD)	0.4% (w/v) potato starch (from infusion) and 2% (w/v) D-glucose
Synthetic nutrient-poor (SNA)	0.1% (w/v) KH_2PO_4; 0.1% (w/v) KNO_3; 0.05% (w/v) $MgSO_4 \cdot 7H_2O$; 0.05% (w/v) KCl; 0.02% (w/v) D-glucose; and 0.02% (w/v) sucrose

All agar versions of the liquid media were prepared with the addition of 1.5% (w/v) agar. All components of the media were dissolved by heating with frequent agitation in dH_2O. All media was sterilized by autoclaving at 121°C for 20 min. Manufacturers/suppliers of the media components are listed in Table 2.1.

2.1.4 Kits

Table 2.4. Kits used in this study

Kit	Manufacturer	Purpose
Big Dye® Terminator v3.1 sequencing	Applied Biosystems, USA	Sanger Sequencing
Direct-Zol™ RNA Miniprep	Zymo Research, USA	Total RNA extraction
GeneJET Gel Extraction and DNA CleanUp Micro	Thermo Scientific, USA	DNA purification from PCR or DNA extraction from agarose gel
GeneJET PCR Purification Kit	Thermo Scientific, USA	DNA purification from ligation reactions
NucleoSpin® Gel and PCR Clean-up	Macherey-Nagel, Germany	Purification of restriction digest
RNA Clean & Concentrator™-5	Zymo Research, USA	Total RNA purification
Verso™ 1-Step RT-PCR ReddyMix™	Thermo Scientific, USA	RT-PCR amplification of RNA
ZR Fungal/Bacterial DNA Miniprep™	Zymo Research, USA	Chromosomal DNA isolation
Zyppy™ Plasmid Miniprep	Zymo Research, USA	Plasmid isolation

2.1.5 Standards

Table 2.5. Standards used in this study

1kb GeneRuler™	Fermentas (Lithuania)
PageRuler™ Prestained Protein Ladder	Thermo Scientific – Fermentas (USA)

2.1.6 Antibiotics

Liquid media (at room temperature) or solid media (at ~55°C) were supplemented aseptically with the appropriate antibiotics shown in Table 2.5. Antibiotic stock solutions were prepared by dissolving the antibiotics in the appropriate solvent followed by filter sterilization using 0.22

µm filters and stored as 500 µl aliquots at - 20°C. All antibiotics (Table 2.5) were obtained from Sigma-Aldrich, USA.

Table 2.6. Antibiotics used in this study

Antibiotic	Stock solution (mg/ml)	Solvent	Final concentration (µg/ml)
Ampicillin (Amp)	100	ddH_2O	100
Chloramphenicol (Chl)	10	Ethanol	10
Streptomycin sulphate (Strep)	50	ddH_2O	50

2.1.7 Enzymes

The various enzymes used in this study for DNA manipulations along with the suppliers are shown in Table 2.7.

Table 2.7. Enzymes used in this study

Enzymes	Function	Supplier
Big Dye® Terminator v3.1	Sequencing PCR	Applied Biosystems, United Kingdom
DreamTaq™ DNA polymerase	Standard and inverse PCR	Fermentas Life Sciences, Lithuania
Q5™ High-fidelity DNA polymerase	Site-Finding PCR	New England BioLabs, United Kingdom
Restriction endonucleases	Restriction enzyme digestion	Fermentas Life Sciences, Lithuania
T4 DNA Ligase	DNA ligation	Fermentas Life Sciences, Lithuania

2.1.8 Primers

Table 2.8. Primers used in this study

Purpose and name of primer		
ITS region primers	***Sequence* (5'→3')**	***Reference***
ITS1-F	GTCGTAACAAGGTTAACCTGCGG	Gardes & Bruns (1993)
ITS4	TCCTCCGCTTATTGATATGC	White *et al.* (1990)

Cdsp amplification	***Sequence* (5'→3')**	***Source***
CPR-For	GGTCGAGTACGTCGAGCAGGATGC	This study
CPR-Rev	GGGGAGGCCATGGAGGTACCAGAG	This study
CPRF	ATGCGCGTTTCTGCTCTCCTCTCC	This study
CPRR	TTACACAAGGTGGGCGAGGACGTT	This study

pGEM®-T vector sequencing primers	***Sequence* (5'→3')**	***Source***
M13 forward	GTTTTCCCAGTCACGAC	Promega (USA)
M13 reverse	CAGGAAACAGCTATGAC	Promega (USA)

Inverse PCR	*Sequence* (5'→3')	*Source*
5'-facing primers		
ICPR3 (5'-oriented)	GACACCGGCGTTGACGATCTCGTC	This study
INEST-1 (5'-oriented)	GCGTTGGACTGAGTGGCCGAGATG	This study
ICPR1 (5'-oriented)	GCATCCTGCTCGACGTACTCGACC	This study
3'-facing primers		
ICPR2 (3'-facing)	GAGCCTCTTGGGGAGCCAACTTTG	This study
INEST-2 (3'-facing)	GTTGGTGCTACGGGATCGGATGAC	This study
ICPR4 (3'-facing)	CTCCACCACCATGTCTGGTACCTC	This study

SiteFinding-PCR	*Sequence* (5'→3')	*Source*
SiteFinder	CACGACACGCTAAACACACCACCTCGCACAGCGT CCTCAAGCGGCCGCNNNNNNGCTC	This study
SFP1	CACGACACGCTAAACAC	Tan *et al.*, 2005
SFP2	ACTCAACACACCACCTCGCACAGC	Tan *et al.*, 2005

Complete cdsp gene PCR	**Sequence (5'→3')**	**Source**
CdspF	ACCTCCCACGGCGTATAACCACCC	This study
CdspR	GTAGTGAACCCGAGAACAAGGTCA	This study

2.1.9 *E. coli* strain used in the study

Table 2.9. The strain of *E. coli* used in this study

Bacterial strain	Relevant genotype	Source
GeneHogs®	*F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZ M15 Δlac 74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG fhuA::IS2* (confers phage T1 resistance)	Invitrogen, USA

2.1.10 Cloning vector

Table 2.10. Vector used in this study

Vector system	Relevant characteristics	Antibiotic resistance	Source
pGEM®-T Easy vector system (3015 bp) (Appendix A)	Contains the following sequence reference points: (i) T7 RNA Polymerase transcription initiation site, (ii) SP6 RNA Polymerase transcription initiation site, (iii) T7 RNA Polymerase promoter, (iv) SP6 RNA Polymerase promoter, (v) multiple cloning site, (vi) Lac Z start codon, operon sequence and operator, (vii) β-lactamase coding region, and (viii) phage f1 region	Ampicillin (100 µg/ml)	Promega, USA

2.2 Methods

2.2.1 Growth and maintenance of fungal isolates

Indigenous fungal isolates obtained from the CMW culture collection (Forestry and Agricultural Biotechnology Institute (FABI)), University of Pretoria, South Africa, were grown on malt extract agar (MEA; Section 2.1.3) plates by FABI culture collection staff. A mycelial plug (less than 5 × 5 mm^2) of fresh-growing hyphae (from edge of fungal culture) was aseptically transferred to the centre of a fresh MEA plate and incubated at 28°C for 7 days and stored at room temperature. Agar plugs containing mycelia from pure cultures were stored short-term in 1 ml of dH2O at 4°C or long-term in 500 µl of 50% glycerol at - 80°C.

2.2.2 Screening of fungal isolates for protein production

Experiments were carried out to identify a fungal isolate capable of producing extracellular protein(s) into liquid growth medium (ME). Mycelial plugs from 7-day-old fungal plates (Section 2.2.1) were aseptically inoculated into 50 ml ME liquid medium in 250 ml Erlenmeyer flasks. The cultures were maintained in a 28°C orbital shaker incubator (180 rpm) for up to 10 days. Cultures showing any signs of bacterial contamination (cloudy appearance) were re-inoculated onto MEA plates containing relevant antibiotics (Section 2.1.6). Once purified, those cultures were re-screened again in ME. Culture supernatants (2 ml) were harvested after 3, 5, 7 and 10 dpi, and mycelia were precipitated by centrifugation at 9,400 x *g* for 5 min. Screened isolates are listed in Appendix B.

2.2.3 TCA precipitation of extracellular proteins

Extracellular proteins were precipitated from culture supernatants with trichloroacetic acid (TCA). An aliquot of the culture supernatant (1 ml) was mixed with 333 µl of TCA (20% w/v) and incubated on ice for 1 hr to precipitate the proteins. Extracellular proteins were collected by centrifugation at 16,000 x *g* for 10 min at 4°C in 1.5 ml Eppendorf tubes. The pellet was washed twice with ice-cold acetone (500 µl) and centrifuged after each wash. The protein containing pellet was thereafter air-dried at room temperature for 30 min to remove any

remaining acetone. The pellet was thereafter resuspended in 25 µl of 1× SDS sample buffer (Table 2.2, Section 2.1.2), and denatured by heating at 100°C for 5 min. The denatured protein sample was centrifuged at 11,000 x *g* for 1 minute.

2.2.4 SDS-PAGE analysis

The extracellular protein profiles were analysed on SDS-PAGE. The identification of potential candidates was based on the presence of dark-blue bands indicating high levels of protein production. SDS-PAGE was used to separate the proteins based on their size against a standard protein molecular ladder according to the method described by Laemmli (1970). In this study, a separating gel (10% or 12%) and stacking gel (4%) were prepared with buffers and chemicals listed in Section 2.1.1 and Section 2.1.2. SDS-PAGE was performed with the Mini-protean®-3 system (Bio-Rad, USA). Components of the gels are shown in Table 2.11.

Table 2.11. Components of SDS-Polyacrylamide gel

	Separation gel (10 ml)		Stacking gel (5 ml)
Composition	***10%***	***12%***	***4%***
dH_2O	4.10 ml	3.40 ml	3.00 ml
Resolving gel buffer (1.5 Tris-HCl; pH 8.8)	2.50 ml	2.50 ml	-
Stacking gel buffer (0.5M Tris-HCl; pH 6.8)	-	-	1.25 ml
Acrylamide/bis-acrylamide 30% solution	3.30 ml	3.70 ml	700 µl
10% SDS	100 µl	100 µl	50 µl
10% APS	50 µl	50 µl	20 µl
TEMED	5 µl	5 µl	5 µl

* Measurements are adequate for two 0.75 mm gels (i.e.; 4.2 ml/gel)

Protein samples (20 µl) were loaded onto 12% SDS-PAGE gels and electrophoresed 200 V for ~ 45 min in 1× running buffer (Table 2.2) until the dye front reaches the bottom of the gel. The gel was stained with Coomassie stain (Ausubel *et al.*, 1998) (Table 2.2) by heating, in a microwave oven for 30 sec and cooled on a rotary shaker at 50 rpm for 1 hr at room temperature. The gel was rinsed with dH_2O and destained in destaining solution (Table 2.2) by heating in a microwave oven for 30 sec. The sizes of the proteins were compared to that of the protein molecular ladder (PageRuler™ Prestained Protein Ladder; Thermo Scientific

(USA)) used. The gel was analysed using a Bio-Rad's Gel Doc™ XR$^+$ system (Bio-Rad, USA) with the Coomassie Fluor™ orange settings.

2.2.5 Peptide sequencing by LC-MS/MS

A Coomassie-brilliant blue stained SDS-PAGE band, indicating high protein expression was excised from the gel and sent for LC-MS/MS (liquid chromatography-tandem mass spectrometry) peptide sequencing at the Council for Scientific and Industrial Research (CSIR; Pretoria, RSA). Briefly, the peptide mixture generated from tryptic digestion was separated using an HPLC coupled to a mass spectrometer according to manufacturer's instructions. Peptide sequences from the obtained MS/MS spectra dataset was identified using the MASCOT search tool (Matrix Science Ltd., UK) against the NCBInr (National Centre for Biotechnology Information; non-redundant) sequence database. Peptides with a high probability (MASCOT scores exceeding threshold; $p < 0.05$) were denoted as "hits". The retrieved peptide sequences were then analysed using the Uniprot pipeline (http://www.uniprot.org/blast/) and NCBI BLASTP (protein-protein Blast) database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (McCormack *et al.*, 1997; Steen & Mann, 2004; Gusakov *et al.*, 2010).

2.2.6 Identification of potential fungal isolate

The high protein-producing fungal isolate was identified by culture, morphology and molecular characterization. In this study, isolate CMW 17970 was identified as a potential fungal candidate for protein production.

2.2.6.1 Morphological studies

Morphological characteristics of fungal candidate, CMW 17970 were examined on potato dextrose agar (PDA), yeast potato dextrose agar (YEPD) and malt extract agar (MEA). Agar plugs from the margin of developing cultures were inoculated in the centre of plates and incubated at 25°C. The colour of the cultures were identified using a mycological colour chart (Rayner, 1970) as a guide. Microscopic structures were observed using an UB203i digital

microscope (UOP, China). Samples were placed in a water droplet on a glass slide, covered with a cover slip, and observed. Light micrograms were obtained under 40x or 100x magnifications with a Zeiss Axiovision microscope (Carl Zeiss Ltd, Germany). Cultural and morphological characterization of the isolate on PDA and MEA was compared with previous descriptions in published literature (Schroers *et al.*, 1999; Schroers, 2001).

2.2.6.2 ITS amplification

Sequencing of the ITS (internal transcribed spacer) region/locus was used to confirm identification of the isolate. Sequencing of the ITS region is a common PCR-based technique used to perform fungal molecular systematics and identification of fungi to the species level. A small amount (tip of sterile syringe needle) of hyphae from an actively growing 5-day old MEA culture was aseptically obtained and resuspended in 100 µl aliquot of the PrepMan® ultra sample preparation reagent in a 1.5 ml Eppendorf tube. The tube was vortexed for 30 sec, followed by heating at 100°C for 10 min in a heat block. The sample was thereafter cooled for 2 min and centrifuged at 12,000 x *g* for 2 min. The supernatant (25 µl) was transferred into a sterile 1.5 ml Eppendorf tube.

The ITS locus was PCR-amplified using the ITS1-F/ITS4 universal primer set (Inqaba Biotech, RSA; Table 2.8). A PCR reaction containing the following reagents was prepared: 2 µl of chromosomal DNA, 0.125 µl of DreamTaq™ DNA polymerase (500 U) (Thermo Scientific, USA), 2.5 µl of dNTP mix (2 mM), 5 µl of DreamTaq™ Buffer (10×), 1.25 µl of ITS1-F primer (10 µM), 1.25 µl of ITS4 (10 µM) and 12.875 µl of sterile distilled water to obtain a final reaction volume of 25 µl. PCR amplification was done using a T100™ Thermal cycler (Bio-Rad, USA) and cycling conditions were: initial denaturation at 94°C for 3 min, followed by 30 cycles, each of which consisted of: 30 s at 94°C, 30 s at 50°C and 30 s at 72°C. A final extension step was kept at 72°C for 8 min, and stored at 4°C.

2.2.7 Optimization of CMW 17970 (*C. rosea* 17970) growth conditions for optimal protein production

2.2.7.1 Inoculum preparation

The type and concentration of inoculum are among the many factors that influence protein production. In this study, mycelia were initially used as inoculum type during screening. The effect of spore (conidia) inoculum and concentration were also investigated. The effect of inoculum on extracellular protein productivity was compared by SDS-PAGE analysis. Fresh spore (conidia) suspensions were harvested from 2-week-old cultures on MEA plates. The plates were submerged in 10 ml of ME and dislodged by scraping with a glass spreader under aseptic conditions. The plate was left flooded for 15 min. The suspension, containing spores was transferred to a sterile 15 ml Falcon tube and vortexed at maximum speed for 1 min to break up spore aggregates. Hyphae were removed by filtering the spore suspension through sterile cheese cloth. The filtered spore suspension (100 µl) was plated aseptically onto MEA petri dishes (145 mm × 20 mm). The plates were then incubated at 28ºC for a week and spores from those plates were obtained as described above.

The concentration of harvested spores was determined with a Neubauer-improved hemocytometer (Marienfeld, Germany) and a digital microscope (model UB203i; UOP, China) at 40X magnification. A stock suspension with a concentration of ~5 x 10^{10} conidia/ml was obtained and inoculated into shake flasks (250 ml) containing ME (50 ml) liquid media to a final concentration of 1 x 10^{4} to 1 x 10^{9} conidia/ml. Cultures were grown at 28°C for 5 days on a rotary shaker (180 rpm). The best type and inoculum concentration were chosen for subsequent optimization experiments. Spore suspensions were alternatively stored in 20-50% sterile glycerol [v/v] at -80°C.

2.2.7.2 Growth media

Conidia (1 x 10^{7} conidia/ml) were inoculated in eight different liquid media (50 ml) listed in Table 3.2. Cultures were grown at 28°C, with shaking on a rotary shaker (180 rpm) for 5 days.

The effect of different media (Section 2.1, Table 2.3) on protein production was compared following SDS-PAGE analysis. The best medium was chosen for subsequent experiments.

2.2.7.3 Effect of growth conditions on protein production

a) Optimum pH studies

The pH of ME was increased from pH 4.8 (unadjusted pH) to different pH values (pH 7.0, 8.0, 8.5, 9.0 and 10.0) using 1M NaOH. The flasks (containing 50 ml of ME) were inoculated with 1 x 10^7 conidia/ml. The effect of pH on protein production was assessed in cultures grown at 28°C, with shaking on a rotary shaker (180 rpm) for 5 days. Extracellular proteins were compared on SDS-PAGE gels. The pH of culture supernatants was monitored daily on pH test strips. Two sets of universal pH indicator papers were used with pH scales of 0-14.0 (1.0 pH interval), and 7.0-14.0 (0.5 pH interval) obtained from Sigma-Aldrich (USA).

b) Optimum temperature studies

Shake flasks containing ME (50 ml) were inoculated with 1 x 10^7 conidia/ml and grown for 5 days with shaking (180 rpm) and incubated at different temperatures (28°C, 29°C, 30°C, 31°C, 32°C, 33°C, and 37°C). Extracellular protein production from the cultures was compared on SDS-PAGE gels.

c) *Protein production under optimum conditions*

Optimum protein production was investigated in ME under optimum conditions: inoculum level (1 x 10^7 conidia/ml), pH 9 and temperature (32°C) over a number of days (3, 4, 5, 7, 10 and 12 dpi). Extracellular proteins were compared on SDS-PAGE gels

2.2.8 Chromosomal DNA isolation

An inoculum concentration of 1 x 10^7 conidia/ml from CMW 17970 (*C. rosea* 17970) (Section 2.2.7.1) was grown in ME on a rotary shaker (180 rpm) at 28°C for 6 days. A culture sample (250 µl) was centrifuged at 9,400 x *g* for 5 min and chromosomal DNA was isolated from 50 mg (wet weight) of mycelia using the ZR fungal/bacterial Miniprep™ kit (Zymo Research, USA) according to the manufacturer's instructions for cultured fungi. Chromosomal DNA was finally eluted in 25-30 µl of sterile Milli-Q water (Millipore) in a 1.5 ml Eppendorf tube. The concentration and purity of extracted DNA (1 µl) was quantified using a spectrophotometer

(Section 2.3.10) and verified by gel electrophoresis (Section 2.2.11). Chromosomal DNA was stored at - 20°C.

2.2.9 RNA isolation and purification

Mycelia (~50 mg) from an actively growing 7-day-old MEA culture was aseptically obtained and total RNA was extracted using the TRI Reagent® Solution (Sigma-Aldrich, USA) and the Direct-Zol™ RNA Miniprep kit according to the manufacturer's instructions. RNA was finally eluted in 30 µl of RNase-free water. The RNA was further purified using the RNA Clean & Concentrator™-5 (Zymo Research, USA) according to manufacturer's instructions. The concentration and purity of the purified RNA was determined by using a spectrophotometer (Section 2.2.10) and immediately used to perform Reverse Transcription-PCR (Section 2.2.18).

2.2.10 Spectrophotometric analysis of DNA and RNA

A volume (1 µl) of DNA and RNA was analysed with a NanoDrop® ND-1000 UV-Vis Spectrophotometer (Thermo Scientific, USA) using the 'Nucleic Acids' software application settings (NanoDrop Technologies Inc., USA). DNA or RNA concentration was determined using the formula, OD_{260} nm x 50 ng/µl, and OD_{260} nm x 44 ng/µl, respectively. The nuclei acids analysed were regarded pure when the ratio between absorbance at 260 nm and 280 nm ($A_{260/280}$) was between 1.8 and 2.

2.2.11 Agarose gel electrophoresis

Agarose gels (0.8%-1.5% (w/v)) were prepared by dissolving agarose in 1× TAE buffer (Table 2.2) by heating in a microwave. Nucleic acid samples were visualized by mixing with 6× GelRed™-loading buffer solution (Table 2.2) using a 5:2 ratio of sample to 6× GelRed™-buffer solution. Samples were thereafter loaded into the wells of the cast gels and separated together with a 1 kb GeneRuler™ (Fermentas, Lithuania) molecular weight marker by electrophoresis at 90-100 V in 1× TAE buffer. Bands were visualized on a UV transilluminator or viewed and

photographed using the Molecular Imager® Gel Doc™ XR+ system (Bio-Rad, South Africa) with settings for the visualization agent used.

2.2.12 PCR amplification of the partial gene fragment

Following the identification of the overproduced protein by LC-MS/MS, the peptide sequences identified were aligned against published protein sequences that are highly homologous using the BLASTP program (NCBI website; http://www.blast.ncbi.nlm.nih.gov/Blast.cgi) (Section 2.2.5). The nucleotide sequences encoding those proteins were downloaded from GenBank and aligned with our *in silico* reverse translated peptide sequences with MAFFT version 7 (Katoh, 2013). A forward and reverse primer pair (CPR-For/CPR-Rev; Section 2.1.8, Table 2.8) was then designed from the conserved (consensus) nucleotide sequences.

A standard 50 µl PCR reaction containing the following reagents was prepared: 25 ng of chromosomal DNA, 1.25 U/µl DreamTaq™ DNA polymerase (500 U; Thermo Scientific, USA), 0.2 µM of dNTP mix, 1× of DreamTaq™ Buffer and 0.5 µM of each primer. The PCR amplification was performed using a T100™ Thermal cycler (Bio-Rad, USA) and cycling conditions were: initial denaturation at 95°C for 3 min, followed by 28 cycles, each of which consisted of: 30 s at 95°C, 30 s at 58°C and 50 s at 72°C. A final extension step was set at 72°C for 10 min, and thereafter cooled to 4°C.

2.2.13 Purification of PCR products

PCR products were purified from solution or agarose gels using the GeneJET Gel Extraction and DNA CleanUp Micro kit (Thermo Scientific, USA) according to manufacturer's instructions. Pure DNA fragments were eluted in 10 µl of sterile Milli-Q water (Millipore). The purity and concentration of the purified DNA fragments were determined spectrophotometrically (Section 2.2.10) and by gel electrophoresis (Section 2.2.11), respectively. Samples were thereafter stored in 600 µl tubes at - 20°C.

2.2.14 Cloning of PCR products

PCR amplicons contained 3' A-overhangs generated by the DreamTaq™ DNA polymerase (Thermo scientific, USA) that enabled A/T cloning into the pGEM®-T easy vector system (Promega, USA; Section 2.1.10; Appendix A) according to the manufacturer's instructions. Purified PCR products (Section 2.2.13) were ligated with the pGEM®-T easy to a molar ratio of 3:1 (Sambrook *et al.*, 1989). The amount of PCR product (insert) required for the ligation reaction was calculated using the formula below:

$$insert\ required\ [\ x\ ng] = \left[\frac{vector[ng]\ \times insert[bp]}{vector\ [bp]}\right] \times \frac{3}{1}$$

In a final volume of 10 µl, 0.3 Weiss/µl of T4 DNA ligase (Promega, USA), 1X Rapid Ligation buffer (Promega, USA), vector (50 ng), insert (x ng) and the required volume of ddH_2O, were mixed in a sterile 600 µl tube. The reaction was incubated overnight (16 hrs) at 16°C. The T4 DNA ligase in the reaction was thereafter deactivated by incubation at 65°C for 10 min. The ligation reaction was evaluated on a 1% (w/v) TAE agarose gel (Section 2.2.11).

2.2.15 *E. coli* transformation

2.2.14.1 Preparation of electrocompetent *E. coli* GeneHogs® cells

The frozen surface of the native *E. coli* GeneHogs® cells' glycerol stock was gently scraped with a sterile inoculation loop, aseptically streaked onto a fresh LB agar (Section 2.1.3) plate and incubated at 37°C overnight. A single colony was inoculated into 15 ml LB liquid media (Section 2.1.3) and the culture was maintained in a 37°C orbital shaker incubator (250 rpm) for 16 hrs (overnight). Fresh LB broth (200 ml) was inoculated with the overnight culture (10 ml) at a starting OD_{600} of 0.05. The culture was thereafter incubated at 37°C with agitation at 250 rpm until the OD_{600} of 0.5-0.7 was reached. The cells were equally portioned into eight pre-chilled 250 ml Falcon tubes (each containing ~25 ml of culture) and kept on ice for 30 min prior to being harvested by centrifugation at 10,000 x *g* for 15 min at 4°C. The supernatant was poured off and the cell pellets were resuspended and washed twice with 12.5 ml of 10%

(v/v) ice-cold glycerol and centrifuged after each wash. The cells were gently resuspended in 150 µl ice-cold GYT medium (10% (v/v) glycerol, 0.125% (w/v) yeast extract and 0.25% (w/v) tryptone). Aliquots (50 µl) of the competent cells were stored in 1.5 ml Eppendorf tubes at -80°C.

2.2.14.2 Transformation of *E. coli* GeneHogs® cells

Electrocompetent cells (50 µl) were thawed on ice for 2 min and the ligation reaction containing the insert-pGEM®-T easy plasmid construct (1 µl; Section 2.2.14) was added directly to the cells followed by incubation on ice for ~1 minute. The mixture (51 µl) was thereafter pipetted into a pre-chilled 0.1 cm gap MicroPulser® electroporation cuvette (Bio-Rad Laboratories, USA). The electric pulse was supplied by a Gene Pulser (Bio-Rad, USA) with parameters set at 1.8 kV, 25 µF and 200 Ω. The electroporated cells were pipetted into 1 ml of pre-warmed (37°C) SOC media (Section 2.1.3) and incubated at 37°C with shaking (175 rpm) for 1 hr. Aliquots of the transformed culture (50, 100 and 150 µl) were then plated onto three separate LB agar plates supplemented with Ampicillin (100 µg/ml) as well as IPTG (0.2 mM) and X-Gal (40 µg/ml) to facilitate blue-white screening to identify recombinant clones when using the pGEM®-T easy vector system. The plates were incubated overnight (16 hrs) at 37°C.

2.2.16 Identification of recombinant clones

Positive colonies (white) and a few blue colonies (Section 2.2.14.2) were sub-cultured onto fresh LB plates containing Ampicillin (100 µg/ml), IPTG (0.2 mM) and X-Gal (40 µg/ml) and incubated overnight (16 hrs) at 37°C. The colonies were analysed for the presence of the DNA insert using a combination of methods:

2.2.15.1 Colony screening PCR

White *E. coli* colonies (positive transformants) were randomly picked and resuspended in 50 µl of sterile Milli-Q water (Millipore) in 1.5 ml Eppendorf tubes. Samples were boiled at 100°C for 5 min, followed by centrifugation at 11,000 x *g* for 2 min. PCR was performed contained

the following reagents: 1× of DreamTaq™ Buffer, 0.2 µM of dNTP mix, 0.5 µM of each sequence specific primer, 1.25 U/µl DreamTaq™ DNA polymerase (500 U) (Thermo Scientific, USA), 4 µl of DNA template and sterile Milli-Q water (Millipore) up to 25 µl. The PCR amplification was performed using a T100™ Thermal cycler (Bio-Rad, USA) and cycling conditions were: initial denaturation at 95°C for 3 min, followed by 30 cycles, each of which consisted of: 30 s at 95°C, 30 s at specific annealing temperature and 50 s at 72°C. A final extension step was set at 72°C for 10 min, and thereafter cooled to 4°C. PCR products were analysed on a 0.8-1% (w/v) agarose TAE gel (section 2.3.11) for the presence of the cloned PCR product.

2.2.15.2 Screening of positive clones by restriction enzyme digestions

Positive colonies (white) were inoculated into LB liquid medium (5 ml) supplemented with 100 µg/ml of Ampicillin and grown overnight (16 hrs) with shaking (150 rpm). Plasmid DNA was isolated from the cultures using the Zyppy™ Plasmid Miniprep kit (Zymo Research, USA) according to the manufacturer's instructions. The extracted plasmid construct was eluted in 30 µl of sterile Milli-Q water (Millipore). The concentration of the purified plasmid DNA was determined spectrophotometrically (Section 2.2.10). The isolated insert-plasmid constructs were digested in a reaction volume of 10 µl consisting of plasmid DNA, 1 U/µl *Eco* RI and 1× buffer *Eco* RI in sterile 600 µl Eppendorf tubes. The reaction tubes were incubated for 1 hr in a 37°C heating block. The digested products were visualized on a 0.8% or 1% (w/v) TAE agarose gels (Section 2.2.11). The purified plasmid DNA was stored at - 20°C.

2.2.17 Chromosome walking to obtain the complete gene fragment

2.2.16.1 *Inverse PCR*

Inverse PCR (Ochman *et al.*, 1988) was used to amplify the 5' and 3' flanking regions of the partial gene fragment (Sections 2.2.12) from chromosomal DNA. A schematic representation of the inverse PCR method is shown in Figure 2.1.

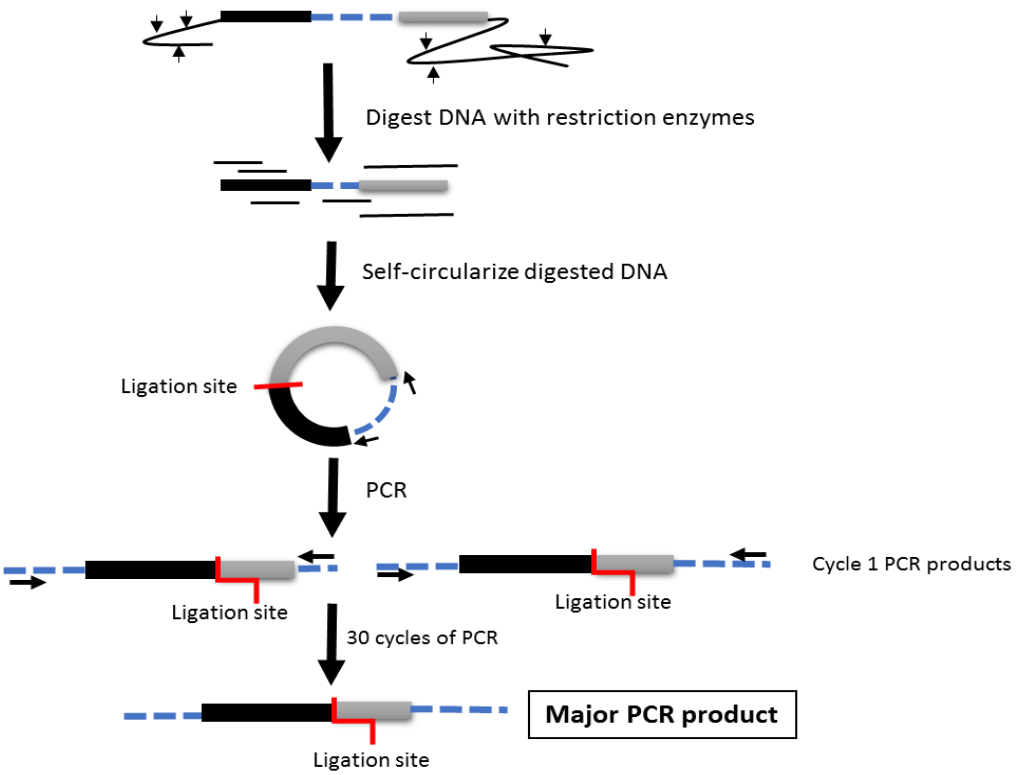


Figure 2.2. A schematic representation of the inverse PCR technique. The known gene region is shown as a dashed blue line between the black and grey boxes which represent the upstream and downstream flanking regions of the gene, respectively. Chromosomal DNA is digested with a specific restriction enzyme (restriction sites are shown as black arrows), self-ligated under conditions which will allow the generation of single circles of the template, and amplified using PCR. Gene specific primers, pointing in opposite directions (shown as black arrows) were designed to anneal to the known gene region. The direction of DNA synthesis using PCR is depicted by black arrows. Image modified and adapted from Ochman *et al.* (1988).

To generate templates for Inverse PCR, chromosomal DNA (~5 µg; Section 2.2.8) was digested in a 50 µl reaction consisting of 1× restriction enzyme buffer and 1-5 U/µl of different restriction endonucleases (Table 2.12). The reaction was incubated at 37°C for at least 16 hrs, followed by incubation at 65°C for 20 min to inactivate the restriction enzyme. Aliquots (3 µl) of the reaction were evaluated on a 1% (w/v) agarose gel (Section 2.2.11). The remaining

reaction was spin-column purified using the NucleoSpin® Gel and PCR Clean-up kit, according to manufacturer's instructions and eluted to a final volume of 50 µl.

Table 2.12. Restriction endonucleases used to digest chromosomal DNA

Name	Sequence	Overhang	Name	Sequence	Overhang
Alu I	AGCT	5' - AG (blunt)	*Nco* I	CCATGG	5' – CATG
Bam HI	GGATCC	5´ - GATC	*Nsp* I	RCATGY	5' – CATG
Bgl II	AGATCT	5´ - GATC	*Sal* I	GTCGAC	5' – TCGA
Eco RI	GAATTC	5´ - AATT	*Xho* I	CTCGAG	5´ - TCGA

The purified digested fragments (2-10 µl) were self-circularized in a 100 µl ligation reaction containing 1× T4 DNA ligase buffer and 1 U of T4 DNA ligase. The reaction was incubated for 16 hrs at 12°C. The T4 DNA ligase in the reaction was thereafter inactivated by heating at 65°C for 10 min. DNA was purified using the GeneJET PCR Purification Kit (Thermo Scientific, USA) according to manufacturer's instructions. The purified circularised DNA was eluted in 10 µl of sterile Milli-Q water (Millipore) and its concentration and purity was determined using a NanoDrop® ND-1000 UV-Vis Spectrophotometer (Thermo Scientific, USA) (Section 2.2.10). Inverse PCR was performed in a 50 µl reaction containing 1× of DreamTaq™ Buffer, 0.2 µM of dNTP mix, 0.5 µM of each inverse primer pair (Table 2.8), ~10 ng/µl of circularized DNA, and 1.25 U/µl DreamTaq™ DNA polymerase (500 U) (Thermo Scientific, USA). The inverse PCR cycle consisted of an initial denaturation of 94°C for 5 min followed by 30 cycles that each contained a denaturation step (94°C for 30 sec), an annealing step (annealing temperature of specific primer pair for 60 sec) and an elongation step (72°C for 2.5 min) followed by an extended final elongation step of 72°C for 10 min in a T100™ Thermal cycler (Bio-Rad, USA). An aliquot (10 µl) of the PCR reaction was visualized on a 1% agarose gel (Section 2.2.11) and the major band was excised from the agarose gel and gel purified using the GeneJET Gel Extraction and DNA CleanUp Micro kit (Thermo Scientific, USA) according to manufacturer's instructions. The purified PCR products were ligated into the pGEM®-T easy vector (Section 2.2.14), transformed into *E. coli* GeneHogs® (Section 2.2.15.2), screened for positive recombinants (Section 2.2.16) and sequenced (Section 2.2.19).

2.2.16.1 *SiteFinding-PCR*

SiteFinding-PCR (Tan *et al.*, 2005) was also used to obtain the 5' and 3' flanking sequences adjacent to the partial gene fragment obtained in Section 2.2.12. The method is shown in Figure 2.2.

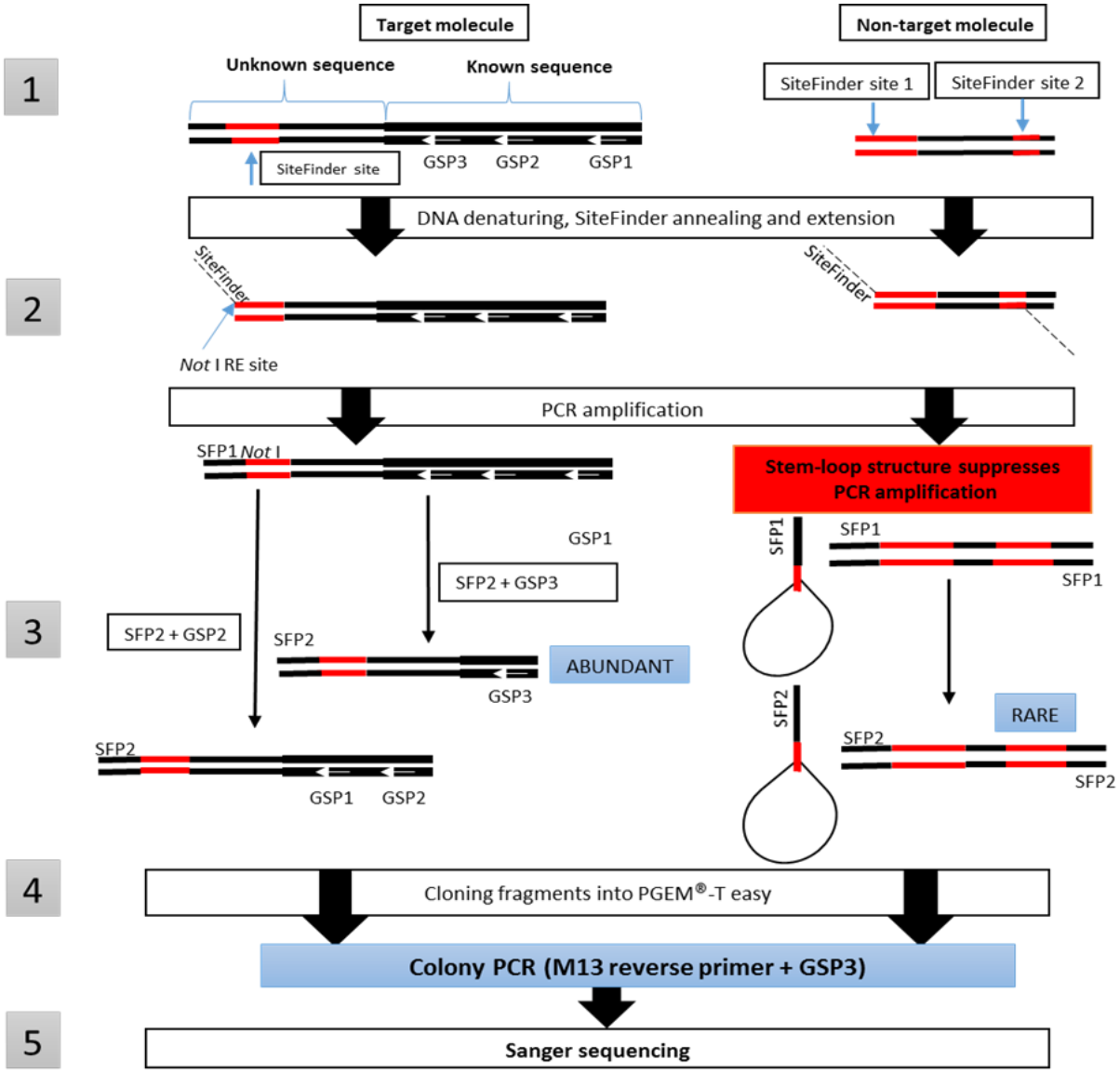


Figure 2.3. A schematic outline of the SiteFinding-PCR method for chromosome walking. 1. Chromosomal DNA templates, showing the flanking region of the know gene sequence and non-target sequence. 2. Site-Finder primer anneals to chromosomal DNA with *Not* I anchor at a low temperature. 3. Nested PCR to amplify the targeted gene sequence with SiteFinder primers (SFPs) and gene specific primers (GSPs). Amplification of non-target products is suppressed by the formation of stem-loop structures. 4. Gel-purified target fragments are cloned into pGEM®-T easy vector. 5. Sequencing of gene fragment with vector-specific primer (M13 reverse) and GSP3. Image modified and adapted from Tan *et al.* (2005).

The oligonucleotide 5'-NNNNNNGCTC-3' at the 3' end of our SiteFinder (Table 2.8) was designed to target 5'-GCTC-3' sites to increase primer specificity and decrease miss-priming (Tan *et al.*, 2005). Two SiteFinding-PCR reactions were prepared, each including 1× One Taq

standard reaction buffer, 200 µM of mixed dNTP solution, 0.5 U of One Taq DNA polymerase (New England Biolabs, USA), 0.2 µM of SiteFinder and ~50 ng of chromosomal DNA (Section 2.2.11). The final volume was brought to 20 µl with sterile Milli-Q water (Millipore). The reaction was subjected to a single cycle PCR cycle (Table 2.13) on a T100™ Thermal cycler (Bio-Rad, USA).Three 5'-facing and 3'-facing gene-specific primers designed for Inverse PCR (Table 2.8) were used to amplify the 5'-flanking and 3'-flanking regions respectively. A primer mixture (5 µl) consisting of 0.08 µM of SiteFinder primer (SFP1), 0.32 µM of 'GSP1' (ICPR3 for upstream and ICPR2 for downstream gene amplification), and 1× One Taq standard reaction buffer, was added to the SiteFinding reaction to a final volume of 25 µl. The PCR reaction carried out using the primary thermal conditions (Table 2.13). The primary PCR reactions were thereafter purified using the GeneJET Gel Extraction and DNA CleanUp Micro kit (Thermo Scientific, USA) according to manufacturer's instructions (Section 2.2.13).

An aliquot (1 µl) of the purified PCR products from the two reactions were diluted and mixed into 1 mL of sterile Milli-Q water (Millipore). An aliquot (2 µl) from each diluted reaction was then combined with a nested PCR reaction mixture (48 µl) consisting of 1× One Taq standard reaction buffer, 25 µM of mixed dNTP solution, 0.8 U of One Taq DNA polymerase (New England Biolabs, USA), and 0.2 µM of the internal SiteFinder primer (SFP2), and 3.2 µM of internal 'GSP2' (INEST-1 and INEST-2 for upstream and downstream gene amplification, respectively). The same PCR reaction was prepared, however 'GSP2' primers were replaced with 'GSP3' primers (ICPR1 for upstream and ICPR4 for downstream amplification). PCR of the four reactions was carried using cycling conditions described on Table 2.13.

Table 2.13. SiteFinding-PCR cycling program

Reaction	Cycle(s) no.	Thermal and time conditions
SiteFinding	1	92°C (2 min), 95°C (1 min), 25°C (1 min), ramp to 68°C (over 3 min), 68°C (10 min)
Primary	1	94°C (30 s)
	30	95°C (10 s), 68°C (6 min)
	1	68°C (5 min)
Nested	1	94°C (30 s)
	30	95°C (10 s), 68°C (6 min)
	1	68°C (5 min)

An aliquot (10 µl) form each PCR reaction was visualized on a 1% agarose gel (Section 2.2.8) and the band generated from the second nested PCR was excised from the agarose gel and purified using the GeneJET Gel Extraction and DNA CleanUp Micro kit (Thermo Scientific, USA) according to manufacturer's instructions. The purified PCR products were ligated into the pGEM®-T easy vector (Section 2.2.13), subsequently transformed into *E. coli* GeneHogs® (Section 2.2.14.2), screened for positive recombinants (Section 2.2.15) and sequenced (Section 2.2.19).

2.2.18 Reverse Transcription-PCR (RT-PCR)

To amplify the ORF of the gene without interrupting intronic sequences, concentrated RNA (Section 2.2.17) was isolated and used as a template to obtain the cDNA of the coding region. Reverse transcription and PCR was carried out in one reaction mix with the aid of a Verso™ 1-Step RT-PCR ReddyMix™ Kit (Thermo Scientific, USA). The sample preparation and thermal cycling conditions were set according to manufacturer's specification indicated in Table 2.14 and Table 2.15, respectively.

Table 2.14. Components of the RT-PCR reaction mix for a 50 µl reaction

	Volume	Final concentration
Verso Enzyme Mix	1 µl	-
1-Step PCR ReddyMix (2X)	25 µl	1X
Forward Primer (10 µM)	1 µl	0.2 µM
Reverse Primer (10 µM)	1 µl	0.2 µM
RT Enhancer	2.5 µl	-
Water (PCR grade)	14.5 µL	-
RNA Template	5 µL	1 ng
Total volume	50 µl	

Table 2.15. 1-Step RT-PCR thermal cycling program

Reaction	Cycle(s) no.	Thermal and time conditions
cDNA Synthesis	1	42°C (25 min)
Verso Inactivation	1	95°C (2 min)
Denaturation Annealing Extension	35	95°C (30 s) Annealing temp. (50 s) 72°C (1.5 min)
Final Extension	1	72°C (5 min)

2.2.19 Automated nucleotide sequencing

In this study, nucleotide sequences of cloned fragments were determined by automated nucleotide sequencing. Sequencing reactions were prepared using the Big Dye™ Terminator version 3.1 cycle sequencing premix kit (Applied Biosystems, USA) in 600 µl tubes. For DNA sequences with more than 600 bp the volume of Big Dye version 3.1 was doubled. Sequencing was performed on an ABI PRISM® DT3100 Genetic Analyzer (PE Applied Biosystems, USA).

Table 2.16. Sequencing reaction mix

	Amount per reaction	
Reagents	**< 600 bp**	**> 600 bp**
Big Dye v3.1 reaction mix	0.5 µl	1 µl
5 x Sequencing buffer	2.4 µl	2.4 µl
DNA template (60-100 ng)	4 µl	4 µl
Universal primer (10 µM)	1 µl	1 µl
ddH2O	4.1 µl	3.6 µl
Total	12 µl	12 µl

Table 2.17. PCR program for sequencing

Reaction	Cycle(s) no.	Thermal and time conditions
Denaturation	1	94°C (5 sec)
Denaturation Annealing Extension	25	94°C (30 s) 50°C (30 s) 60°C (4 min)
Hold	4	∞

To prepare the sample for sequencing, the PCR reaction was subjected to DNA precipitation using ethanol. To each sample, 3 volumes of 100% (v/v) ethanol and one fifth of the volume of 3 M sodium acetate (pH 5.2) were added. The reaction mixtures were incubated on ice for 30 min and centrifuged at 4°C for 30 min at 16,000 x *g*. The supernatant was discarded and the precipitate was washed twice with 500 µl of cold 70% (v/v) ethanol, followed by centrifugation at 4°C for 5 min at 16,000 x *g* after each wash step. The DNA pellet was air-dried and analysed by the DNA Sequencing Facility at University of Pretoria, RSA (http://seqserve.bi.up.ac.za). Alternatively, purified PCR fragments which were difficult to

transform were sequenced at the Central Analytical Facility (CAF) at Stellenbosch University (http://academic.sun.ac.za/saf/).

2.2.20 Sequence analysis

The forward and reverse nucleotide sequences from sequenced data (Section 2.2.19) were visualized by using Chromas Lite version 2.1.1 (Chromas Lite, Technelysium, Australia; http://www.technelysium.com.au/chromas_lite.htm) and aligned using BioEdit sequence alignment editor version 7.2.5 (Hall, 1999). The consensus sequence was compared with sequences in the the GenBank database using the basic local alignment search tool (BLAST) program on the NCBI website (http://www.blast.ncbi.nlm.nih.gov/Blast.cgi; megablast or blastn algorithms).

The putative protein sequence was assessed for the presence of a secretion signal sequence using SignalP version 4.1 (http://www.cbs.dtu.dk/services/signalP). Putative post-translational modification sites (e.g.; acetylation, amidation, phosphorylation, etc.) and conserved motifs (e.g.; enzyme active sites) were also identified through comparisons with the PROSITE (release 20.115, 5 March 2015) web database (http://prosite.expasy.org/; Sigrist *et al.*, 2013). The protein family and its predicted function were determined through multiple sequence alignments (generated using hidden Markov Models) with the Pfam database (release 27.0, March 2013) (http://pfam.xfam.org; Punta *et al.*, 2012) and the NCBI-CDD (conserved domain database; http://www.ncbi.nlh.nih.gov/structure/cdd/cdd.shtml).

Nucleotide sequences upstream of the start codon (ATG) were assessed for general conserved transcription control motifs, including CAAT-boxes (5'-CCAAT), TATA-boxes (5'-TATAAA), and the initiator (Inr) element (5'-YYANWYY-3'). Putative transcription control factors for nitrogen regulation (5'-GATA), carbon regulation (5'-SYGGRG; abbreviated as CreA), stress response (5'-AGGGG; abbreviated as STRE), and pH regulation (5'-GCCARG; abbreviated as PacC) within the promoter region were also investigated. The mentioned DNA regulatory motifs of interest were identified by using a combination of different *in silico* motif

prediction software i.e.; ScerTF (http://stormo.wustl.edu/ScerTF), and MEME suite (GOMO algorithm, http://meme.ncbr.net/meme/). The same bioinformatics approach was applied to analyse the downstream regions for consensus putative polyadenylation sites, and conserved signal elements which are important for mRNA processing. Conserved putative RNA-binding motifs were determined by using the MEME suite (http://meme.ncbr.net/meme/) and RBPDB (http://rbpdb.ccbr.utoronto.ca/index.php) motif search and annotation tools.

CHAPTER 3: RESULTS

3.1 Screening for extracellular proteins

One hundred and fifty five isolates (Appendix B) from the CMW culture collection were screened in ME liquid media. Examples of protein profiles of 7-day-old culture supernatants resolved on Coomassie-stained 12% SDS-PAGE from 23 fungal isolates are shown in Figure 3.1. Screening targeted wild-type fungi with a natural ability to produce high levels of extracellular proteins in culture supernatants. The 23 CMW isolates shown, produced different protein profiles.

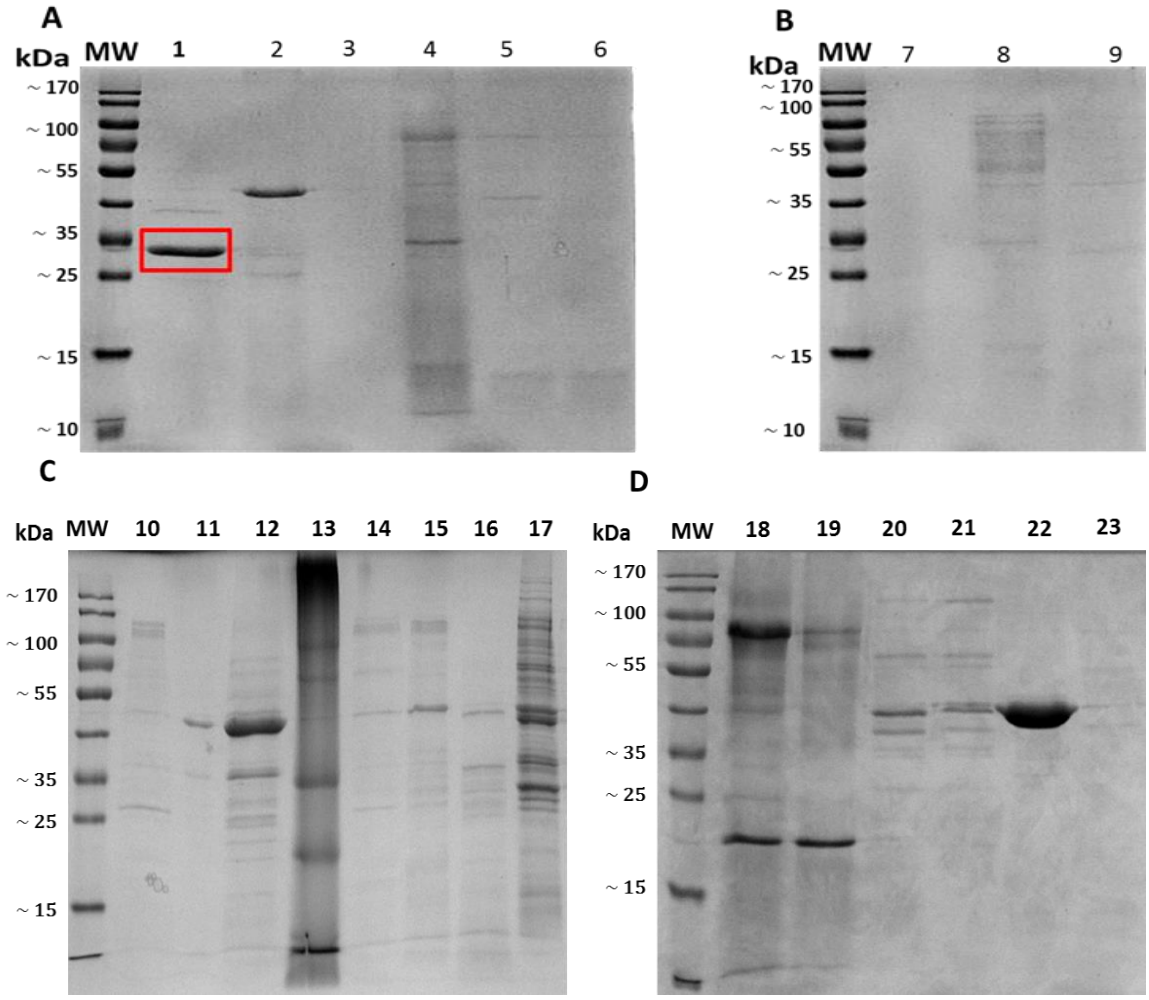


Figure 3.1. Extracellular protein profile of 23 fungal isolates resolved on 12% SDS-PAGE gels. *Gel A*: Lane 1, CMW 17970; Lane 2, CMW 17972; Lane 3, CMW 37667; Lane 4, CMW 5690; and Lane 5, CMW 1166. *Gel B*: Lane 6, CMW 226; Lane 7, CMW 13980; Lane 8, 3788; and Lane 9, CMW 242. *Gel C*: Lane 10, CWM 7130; Lane 11, CMW 794; Lane 12, CMW 5690; Lane 13, CMW 592; Lane 14, CMW 207; Lane 15, CMW 8217; Lane 16, CMW 13229; and Lane 17, CMW 25892. *Gel D*: Lane 18, CMW 19090; Lane 19, CMW 19091; Lane 20, CMW 13232; Lane 21, CMW 13233; Lane 22, CMW 6261 and Lane 23, CMW 6260. MW is the PageRuler™ molecular weight marker. Each lane represents extracellular protein from 1 ml of culture supernatant.

The gel images indicate that most of the isolates produced low levels of protein, and some had very little protein in the sample. Four isolates: CMW 17970 (Gel A, Lane 1), CMW 17972

(Gel A, Lane 2), CMW 5690 (Gel C, Lane 12), and CMW 6261 (Gel D, Lane 22) had strong bands, indicating higher protein production. Apart from CMW 17970, all three of the other isolates were contaminated with bacteria. These isolates were purified through antibiotic treatment and re-screened in ME medium. Following SDS-PAGE analysis, the dark bands originally obtained were lost, indicating that the proteins were produced by the bacteria. Therefore only CMW 17970 was selected as the best candidate for protein production in ME liquid medium.

3.2 Protein identification

The prominent extracellular protein (~35 kDa) produced by CMW 17970 was identified by LC-MS/MS sequencing. The MS/MS spectra data was deposited into the UniProtKB/SwissProt protein database, and results were analysed using the BLAST algorithm incorporated on the UniProt website (http://www.uniprot.org/blast/). Table 3.3 shows putative peptide/protein sequences obtained.

Two putative fungal proteins were identified using BLAST search. The first identified protein produced 80 MS/MS spectra representing partial peptide sequences homologous to the ~33 kDa cuticle-degrading serine protease (PrC) by *Clonostachys rosea* (Uniprot accession number: C6K2H8; 386aa) published by Li *et al.* (2006). The second protein's peptide data matched a phosphomannomutase-containing domain from a phosphoglucomutase protein by *Colletotrichum graminicola* (UniProt accession number: E3QD33; 540aa) identified from only 3 MS/MS spectra. Peptide sequences from fungal protein, hit number 1 (Table 3.3) were aligned against that of the putative PrC protein from *C. rosea* (Figure 3.2). With sequence coverage of over 40%, the putative PrC protein was confirmed as the most likely protein compared to that of *C. graminicola* (below 5% for the phosphoglucomutase protein). The protein identified from CMW 17970 will be referred to, as Cdsp (cuticle-degrading serine protease).

Table 3.1. Identification of peptides/protein from the gel slice by LC-MS/MS

Hit no.	Uniprot KB code	Protein name	Organism	Peptide sequence (Most important ones)	Confidence scores (%)
1	C6K2H8	Cuticle-degrading serine protease (PrC)	*C. rosea*	AANYSPASAASAC	99.00
				AANYVDSSSTDGNGHGTHVAGTVGGTTYGVAK	99.00
				ASFAANYVDSSSTDGNGHGTHVAGTVGGTTYGVAK	99.00
				ASPHVAGLGAYYLGLGR	99.00
				VEYVEQDAIVSISATQTGAPWGIAR	99.00
				ITSVPSGTPNVLAHLV	91.22
				LSNTNTGSTTYTYDDSAGDGTCAFIIDTGIYTSHS	99.00
				SGTSMASPHVAGLGAYYLGLGR	95.11
				TYDDSAGDGTCAFIIDTGIYTSHSDFGGR	99.00
				VAGLGAYYLGLGR	99.00
				VLDSSGSGTTSGVIAGMNYVTN	99.00
				TGAPWGIAR	12.46
				YYLGLGR	23.68
				GAYYLGLGR	48.80
				IVSTGLTGR	1.16
				LYAVKVLDSSGSGTTSGVIAGMNYVTN	20.83
				VLDSSGSGTTSGVIAGMNYVTN	69.00
				YVDSSSTDGNGHGTHVAGTVGGTTYGVAK	52.08
2	E3QD33	Phosphoglucomutase	*Colletotrichum graminicola*	VAGIVKRYGGL	98.99
				DELANKVAGIVK	4.5×10^{-5}
				IDLSAPAR	4.5×10^{-5}

```
  1  MRVSALLSIL PLVAAAPAKR EEVAPLHVPR DVEVIPGKYI VKLKEGVVSI SSTISSIEAK
 61  PDFEYEGGFQ GFVRIYTSLV SQTRNMQITD NSSHRLVEYV EQDAIVSISA TQTGAPWGIA
121  RLSNTNTGST TYTYDDSAGD GTCAFIIDTG IYTSHSDFGG RASFAANYVD SSSTDGNGHG
181  THVAGTVGGT TYGVAKKTKL YAVKVLDSSG SGTTSGVIAG MNYVTNSAGT YSCPKGVVVN
241  MSLGGGYSAS LNTAANNIVS AGYFLAVAAG NSAANAANYS PASAASACTV GATTSSDALA
301  SYSNYGSIVD ILAPGSSVLS AYNNGGTATL SGTSMASPHV AGLGAYYLGL GRASASGLCS
361  YIVSTGLTGR ITSVPSGTPN VLAHLV
```

Figure 3.2. CMW 17970 peptide fragments identified from LC/MS-MS sequencing aligned to the amino acid sequence of the PrC protein by *C. rosea*. Red indicates low scoring peptides (<50% confidence), yellow for moderate scoring peptides (>50% confidence), and green for high scoring peptides (>95% confidence).

3.3 Identification of CMW 17970

3.3.1 Morphological studies of *C. rosea* 17970

Morphological studies were carried out to help identify the fungal isolate (Section 2.2.6.1). Microscopic and macroscopic description of the CMW 17970 isolate were confirmed and described using Schroers *et al.* (1999) and Schroers (2001) as aids. Figure 3.3 shows the macroscopic morphology of the isolate.

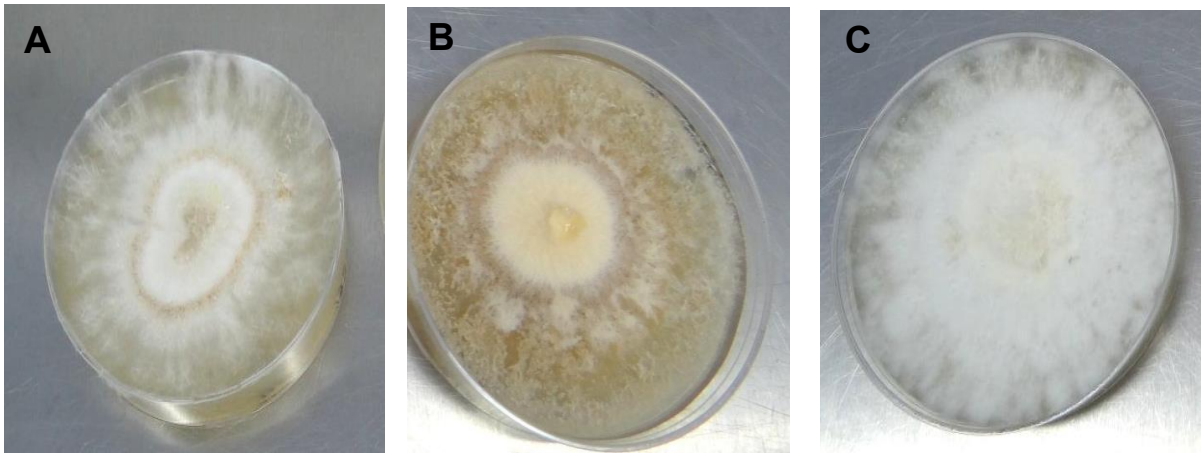


Figure 3.3. Morphological characteristics of CMW 17970. Examples of the fungal isolate on MEA (A); PDA (B) and YPDA (C) after 2 weeks incubation at 25°C.

Colony on MEA (Figure 3. 3A): Fast-growing at 25°C, with a colony diameter of ~27 mm after 7 days. Colony was mostly white with a yellow ring forming around the site of inoculation. Aerial hyphal strands were present 2 weeks post inoculation (wpi). The reverse side pigment of the agar plate was cream in colour. Brown granulose-like structures were visualized 2 wpi, and the number of granules increased after 4 weeks.

Colony on PDA (Figure 3.3B): The colony was white-to-yellowish and had a dense cotton-like surface texture. Colony diameter was ~25 mm after 7 days. Aerial hyphal strands were also present after 2 weeks. The reverse plate colouration is white-to-light yellow. The orange-to-white granule-like structures developed in 2 weeks, and increased in number after 4 weeks.

Colony on YPDA (Figure 3.3C): Fast-growing fungal culture, with yellow-to-white dense cotton-like surface structure. The colony diameter reached ~30 mm after 7 days. The surface and reverse colony colouration was light yellow-to-white.

CMW 17970 displayed two types of conidiophores. Primary (verticillium-type) conidiophores lacked branches with divergent phialides. Conidia from these conidiophores were slightly curved and rounded at the distal end (Schroers *et al.*, 1999). In contrast, secondary conidiophores (Figure 3.4) existed alone or grouped together. Described in literature as penicillate; they possessed branches and phialides which were tightly clustered together and tapered towards the tip. Phialides measured 10-16 µm x 2-3.2 µm. Conidia from this type of conidiophores were shortened and less rounded (Figure 3.4). Conidia measured 3.5-6.0 µm x 2-3.8 µm. Stipes measured 20-200 µm long. Cultural and morphological characteristics revealed CMW 17970 to be similar to *C. rosea* described by (Schroers *et al.*, 1999; Schroers, 2001).

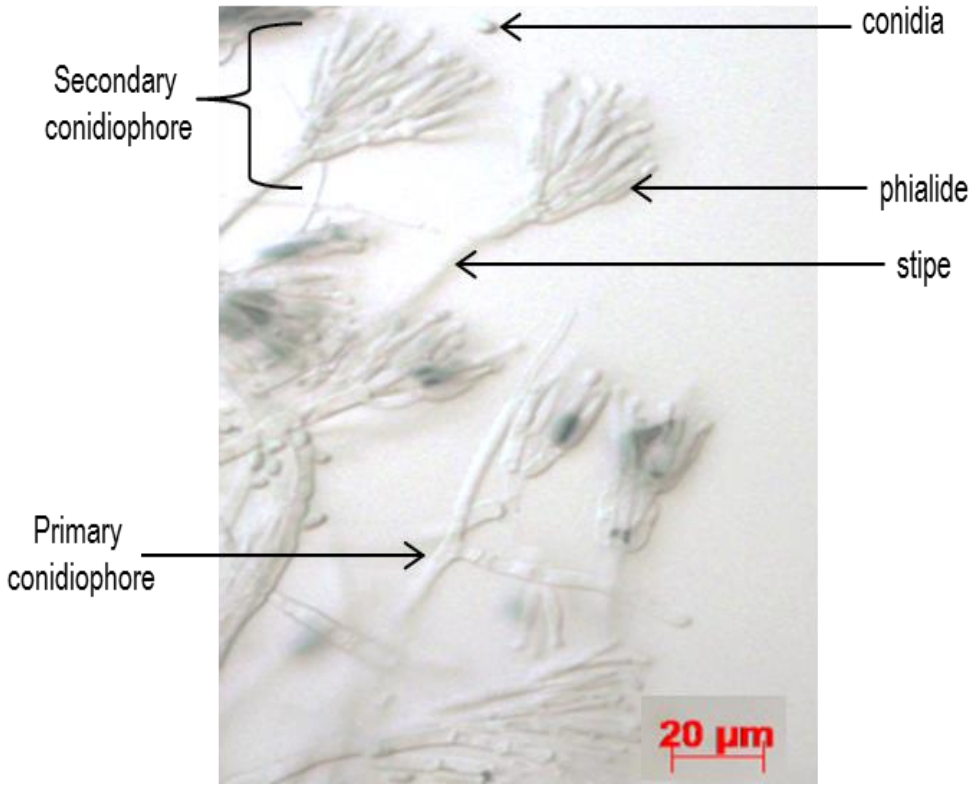


Figure 3.4. Microscopic structures of CMW 17970. Labelled micrograph of primary conidiophores, secondary conidiophores and conidia are shown. Scale bar: 20 µm. Image analysed using the Zeiss AxioVision 4.8. Software (Carl Zeiss; Germany)

3.3.2 Molecular identification of CMW 17970 as *C. rosea*

In order to verify the identity of CMW 17970, the sequence of the ITS region was analysed *in silico* (Section 2.2.6.2). The results confirmed that the CMW 17970 isolate was 100% identical to *C. rosea* ATT017 (NCBI accession number: HQ607798), which was isolated from soil in Texas, USA (Rodrigues *et al.*, 2011). The isolate also showed homology (99%) to other *C. rosea spp.* and with its telemorph *Bionectria ochroleuca* CBS113336 strain (NCBI accession number: EU552110). The ITS sequence is shown in Appendix C.

3.4 Optimization of growth conditions

Following the identification of the Cdsp protein, culture-dependent methods were investigated to optimize the amount of protein produced and secreted by *C. rosea* 17970 (Section 2.2.7). The various parameters tested included initial inoculum type and concentration, growth media, temperature, and pH. Optimization was done by the single-parameter approach i.e.; one variable was investigated per experiment. The experiments were carried out in duplicate.

3.4.1 Selection of optimum inoculum

The effect of inoculum concentration on protein production was investigated in shake flasks containing ME liquid media (Section 2.2.7.1). Figure 3.5 shows the various growth morphologies observed in shake flask cultures inoculated with different spore concentrations (1×10^4 – 1×10^9 conidia/ml). Pellets were dominant for cultures with spore inoculum concentration below 1×10^6 conidia/ml. The structure of the pellets varied between the inoculum concentrations tested. Pellets from concentrations of 1×10^4 conidia/ml and below, had a “hairy” mycelial surface with a smooth-looking compact core. The same pellet morphology was observed when ME was inoculated with mycelial plugs instead of conidia (result not shown). However, compact shiny pellets with less “hair” were observed for cultures with initial inoculum concentration between 1×10^5 and 1×10^6 conidia/ml. Inoculum

concentration of 1 x 10^7 conidia/ml and higher produced filamentous morphology (freely dispersed mycelia) (Figure 3.5).

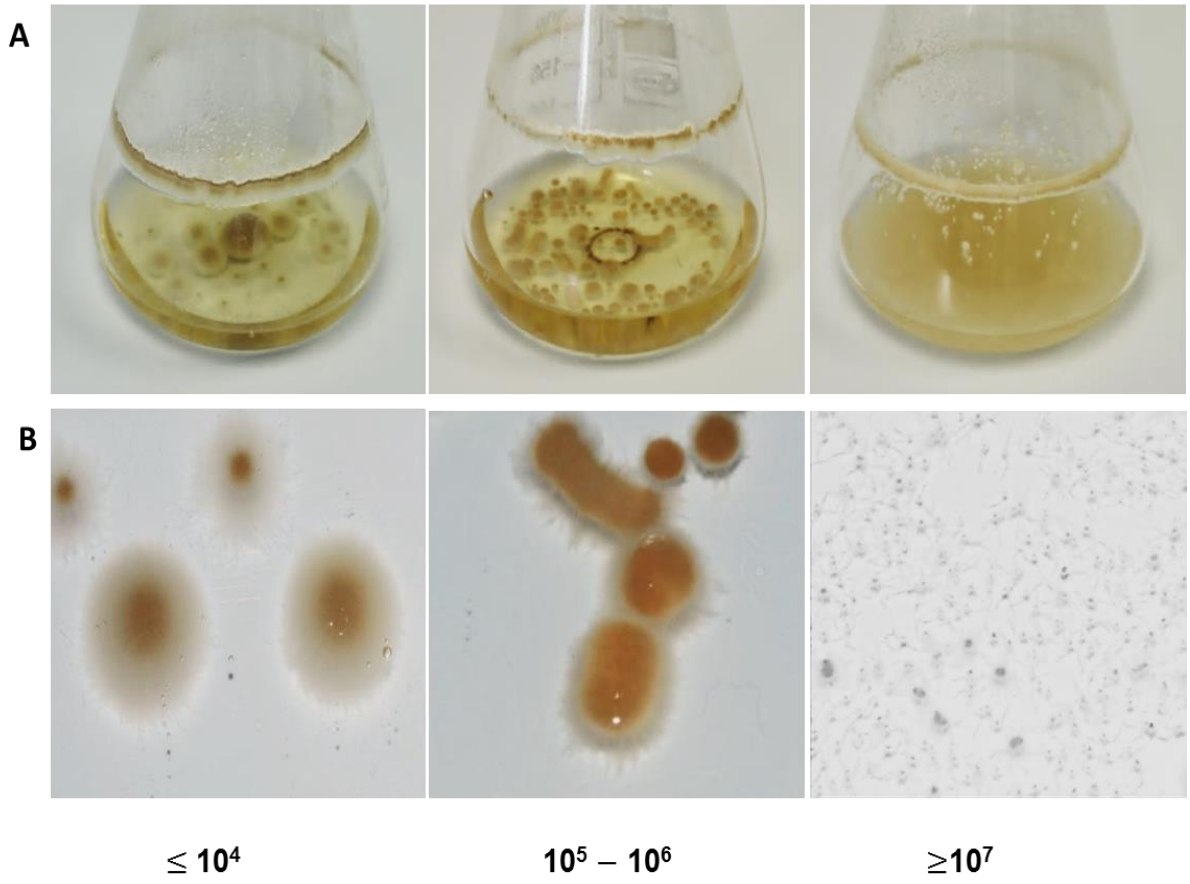


Figure 3.5. Different growth morphology of *C. rosea* 17970 cultures inoculated with different spore concentrations. (A) Shake flasks and (B) close-up photographs of culture in A.

The protein profiles of culture supernatants from the different inoculum concentrations were compared for days 3, 5 and 7 on 12% SDS-PAGE gels. However, strong protein bands of the expected size (~35 kDa) was already visualized from day 5-day-old cultures (Figure 3.6), much earlier than the original screening experiment (7-day-old cultures) when a mycelial plug was used for the initial inoculum (Figure 3.1A). The most prominent band corresponding to elevated protein production was observed for 1 x 10^7 conidia /ml in ME liquid media. This result confirmed the importance of determining the inoculum type (mycelia or spores), spore concentration, and growth morphology for protein productivity. The optimum spore inoculum concentration (1 x 10^7 conidia/ml) was used for subsequent experiments.

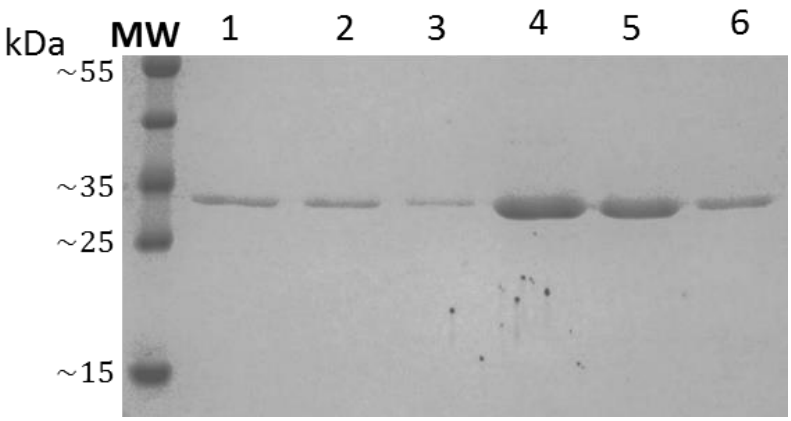


Figure 3.6. Extracellular protein profile of 5-day-old ME cultures inoculated with different concentrations of conidia. Lane 1, 1 x 10^4 conidia/ml; Lane 2, 1 x 10^5 conidia/ml; Lane 3, 1 x 10^6 conidia/ml; Lane 4, 1 x 10^7 conidia/ml; Lane 5, 1 x 10^8 conidia/ml; and Lane 6, 1 x 10^9 conidia/ml. MW is the PageRuler™ molecular weight marker. Each lane represents extracellular protein from 1 ml of culture supernatant.

3.4.2 Selection of liquid medium for high level protein production

The effect of medium composition on protein production (Section 2.2.7.2) was investigated by growing *C. rosea* 17970 in different liquid media (Table 3.2). The protein profiles from each medium inoculated with the optimum inoculum concentration (1 x 10^7 conidia/ml) are shown in Figure 3.7. Optimum protein production was observed when using ME liquid media.

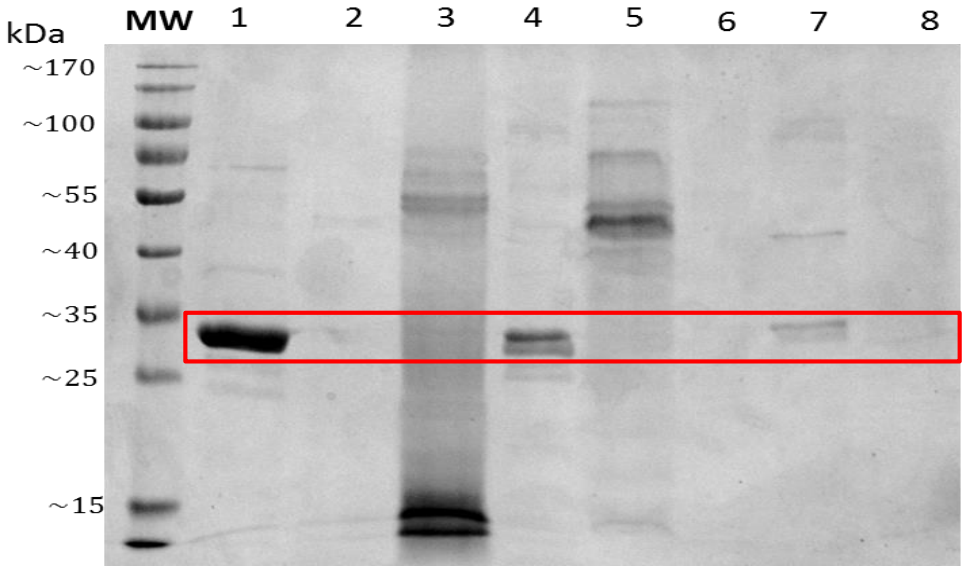


Figure 3.7. SDS-PAGE protein profile of *C. rosea* 17970 5-day-old culture supernatants with an inoculum level of 1 x 10^7 conidia/ml. Lanes 1 to 8 represent different liquid media used: Lane 1, Malt extract; Lane 2, Synthetic nutrient poor; Lane 3, Potato dextrose; Lane 4, Malt yeast extract; Lane 5, Gelatin-rich; Lane 6, Czapek-dox; Lane 7, Casein-rich and Lane 8, Carboxymethyl cellulose. MW is the PageRuler™ molecular weight marker. Each lane represents extracellular protein from 1 ml of culture supernatant.

The gel image shows that the expected ~35 kDa band was present in ME, malt yeast extract (MYE) and casein-rich (CR) culture supernatants, however the band was only prominent in ME. No bands were visualized for synthetic nutrient poor (SNA), czapek-dox (CZD) and carboxymethyl cellulose (CMC) media. The principal carbon and nitrogen sources found in the different media are shown in Table 3.4.

It was found that the composition of ME was best suited for the production of Cdsp. The presence of freely dispersed mycelial growth was only observed in ME, whereas pellets were observed in all other media regardless of spore inoculum concentration (data not shown), confirming the impact of morphology on protein production for *C. rosea* 17970.

Table 3.2. Carbon and nitrogen sources present in liquid media used in this study

Media	Main Carbon Source	Main Nitrogen Source
ME	**Maltose**	peptides, vitamins, amino acids, and purines
SNA	Sucrose	Potassium nitrate
PD	D-glucose	None
MYE	D-glucose/Maltose	Bactopeptone

Media	Main Carbon Source	Main Nitrogen source
GR	Gelatin	Ammonium sulphate
CZD	Sucrose	Sodium nitrate
CMC	CMC	Ammonium nitrate
CR	Casein	Ammonium sulphate

ME liquid medium is the best growth medium (in bold).

3.4.3 Effect of initial pH

The effect of initial pH on protein production was examined (Section 2.2.7.3a) over pH ranges 4.8 to 10.0. The optimal initial inoculum concentration of 1 x 10^7 conidia/ml was used for this experiment. Protein profiles of culture supernatants from 3-day-old cultures representing pH values 4.8, 7.0, 8.0, 8.5, 9.0, and 10. 0 are shown in Figure 3.8. The optimum pH for Cdsp production was 9.0. The dominant growth morphology of *C. rosea* 17970 during this experiment was freely dispersed mycelia, which corresponded to the inoculum level used. The pH of the different cultures was monitored daily for pH shifts during

growth. It was found that pH values changed after 3 dpi and remained fairly constant. For example, pH 4.8 shifted up to ~ pH 6, and pH 9 shifted down to ~ pH 7.

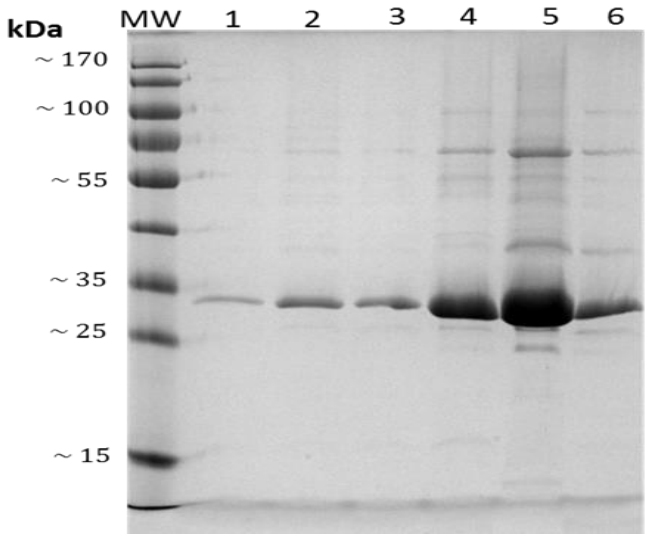


Figure 3.8. Protein profile of *C. rosea* 17970 3-day-old culture supernatants at different initial pH values. Lane 1, pH 4.8; Lane 2, pH 7.0; Lane 3, pH 8.0; Lane 4, pH 8.5; Lane 5, pH 9.0 and Lane 6, pH 10.0. MW is the PageRuler™ molecular weight marker. Each lane represents extracellular protein from 0.5 ml of culture supernatant.

3.4.4 Effect of temperature

The effect of temperature on protein production was investigated (Section 2.2.7.3b). An optimum inoculum concentration of 1 x 10^7 conidia/ml was used and the pH of ME was unadjusted (pH 4.8) for this experiment. The resulting SDS-PAGE analysis of extracellular

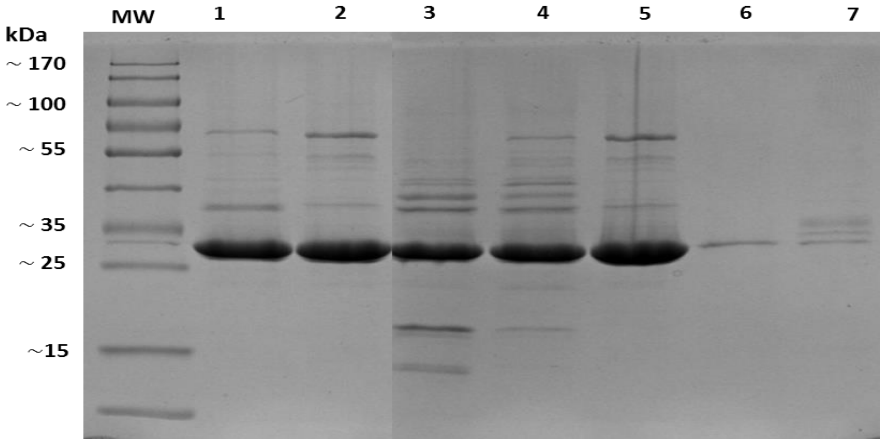


Figure 3.9. SDS-PAGE analysis of culture supernatant of 5-day-old cultures of *C. rosea* 17970 incubated at different temperatures./ Lane 1, 28°C; Lane 2, 29°C, Lane 3, 30°C, Lane 4, 31°C; Lane 5, 32°C; Lane 6, 33°C; and Lane 7, 37°C). MW is the PageRuler™ molecular weight marker. Each protein profile represents extracellular protein from 0.5 ml of culture supernatant.

protein content is shown in Figure 3.9. The results show a broad temperature range (28°C–32°C) for Cdsp production. Slightly more protein was observed for the culture incubated at 32°C.

3.4.5 Growth at optimum conditions

C. rosea 17970 was grown at optimum conditions as determined in previous experiments (Section 2.2.7.3c) in ME liquid medium: inoculum concentration of 1 x 10^7 conidia/ml, initial pH 9.0 of ME, and incubation at 32°C. The protein profiles are shown in Figure 3.10. Protein production, represented by the prominent ~35 kDa band, was very consistent for cultures incubated up to 12 days.

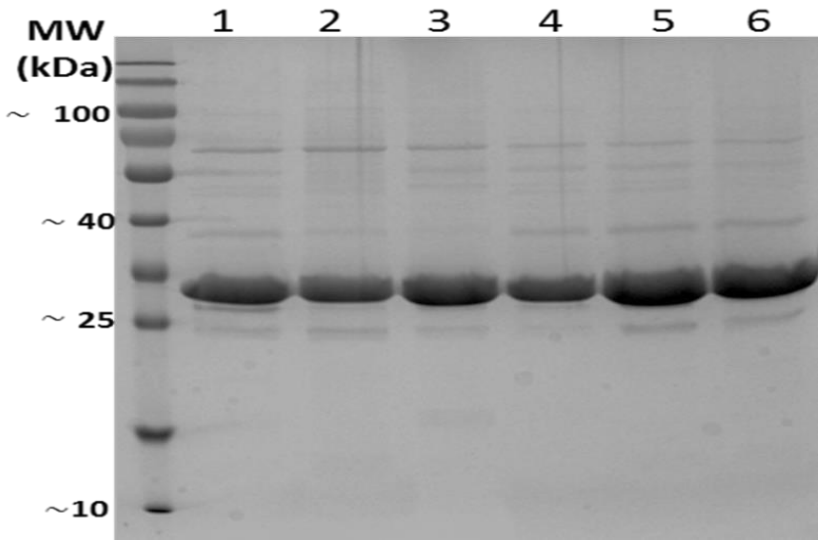


Figure 3.10. Secreted protein profile of *C. rosea* 17970 at optimum conditions. Lane 1, 3 dpi; Lane 2, 4 dpi; Lane 3, 5 dpi; Lane 4, 7 dpi; Lane 5, 10 dpi; and Lane 6, 12 dpi. MW is the PageRuler™ molecular weight marker. Each protein profile represents extracellular protein from 0.5 ml of culture supernatant.

3.5 Amplification and sequencing of the partial *cdsp* gene fragment

PCR amplification from *C. rosea* 17970 chromosomal DNA (Section 2.2.9, Figure 3.11A) using primers CPR-For and CPR-Rev (Section 2.1.8) resulted in an 841 bp fragment (Figure 3.11B). The fragment was purified from a 1% (w/v) agarose gel (Section 2.2.12), ligated into the pGEM®-T Easy vector (Section 2.2.14) and transformed into electrocompetent *E. coli* GeneHogs® cells (Section 2.2.15). Recombinant colonies were screened with PCR (Section 2.2.16.1) and *Eco* RI restriction digests (Section 2.2.16.2). Positive clones were sequenced

(Section 2.2.19) using M13F and M13R sequencing primers (Section 2.1.8, Table 2.8). The sequenced data confirmed the insert to be an internal fragment of the *cdsp* gene when compared with other serine protease genes (data not shown). The *cdsp* partial gene fragment (Figure 3.12) showed moderate homology (70%) with both *PrC* gene from *B. ochroleuca* (NCBI accession number: GQ149467) and *Sep* gene from *C. rosea* (NCBI accession number: KM199682) genes through BLASTN analysis.

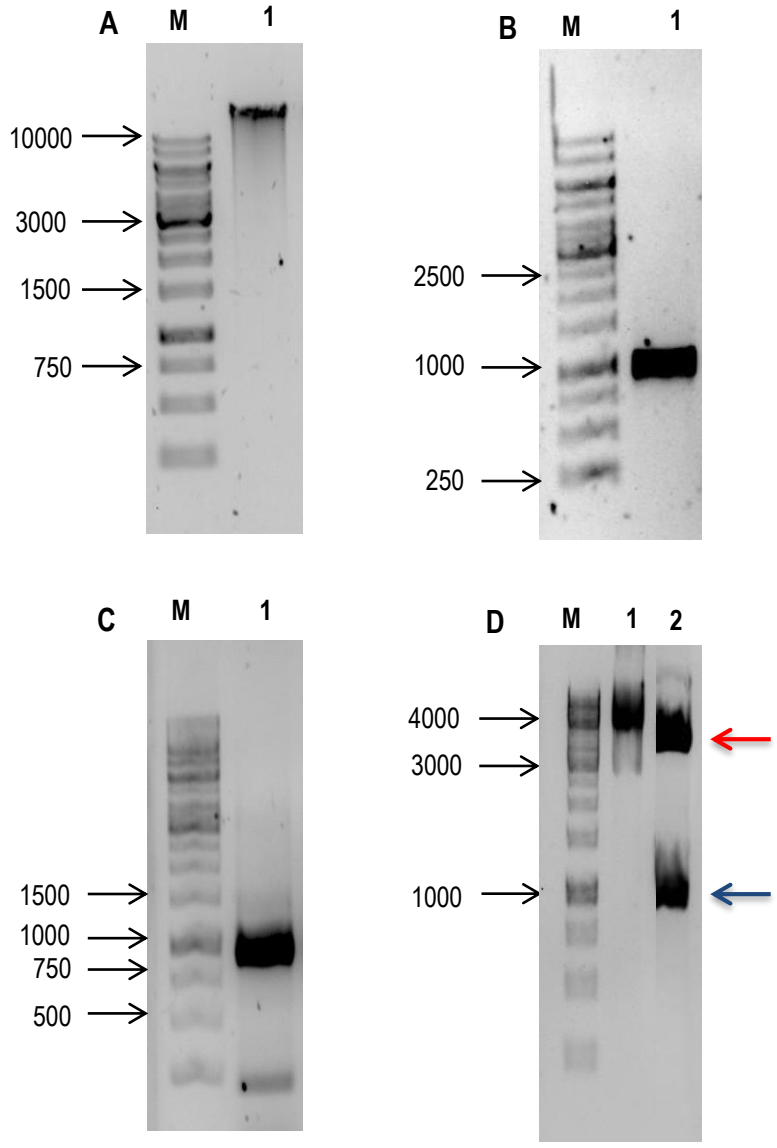


Figure 3.11. Experiments performed to obtain the sequence of the partial *cdsp* gene fragment. *Gel A*: Lane 1, chromosomal DNA. *Gel B*: Lane 1, partial *cdsp* gene fragment (841 bp). *Gel C*: Lane 1, colony PCR of the partial *cdsp* gene fragment (841 bp). Gel D: Lane 1, undigested plasmid DNA (~ 3,856 bp); and Lane 2, *Eco* RI digested plasmid DNA showing the released gene insert (~ 859 bp) is indicated by a blue arrow and the remaining pGEM®-T Easy vector (~ 2,997 bp) is indicated by a red arrow. M is the GeneRuler 1 kb DNA Ladder.

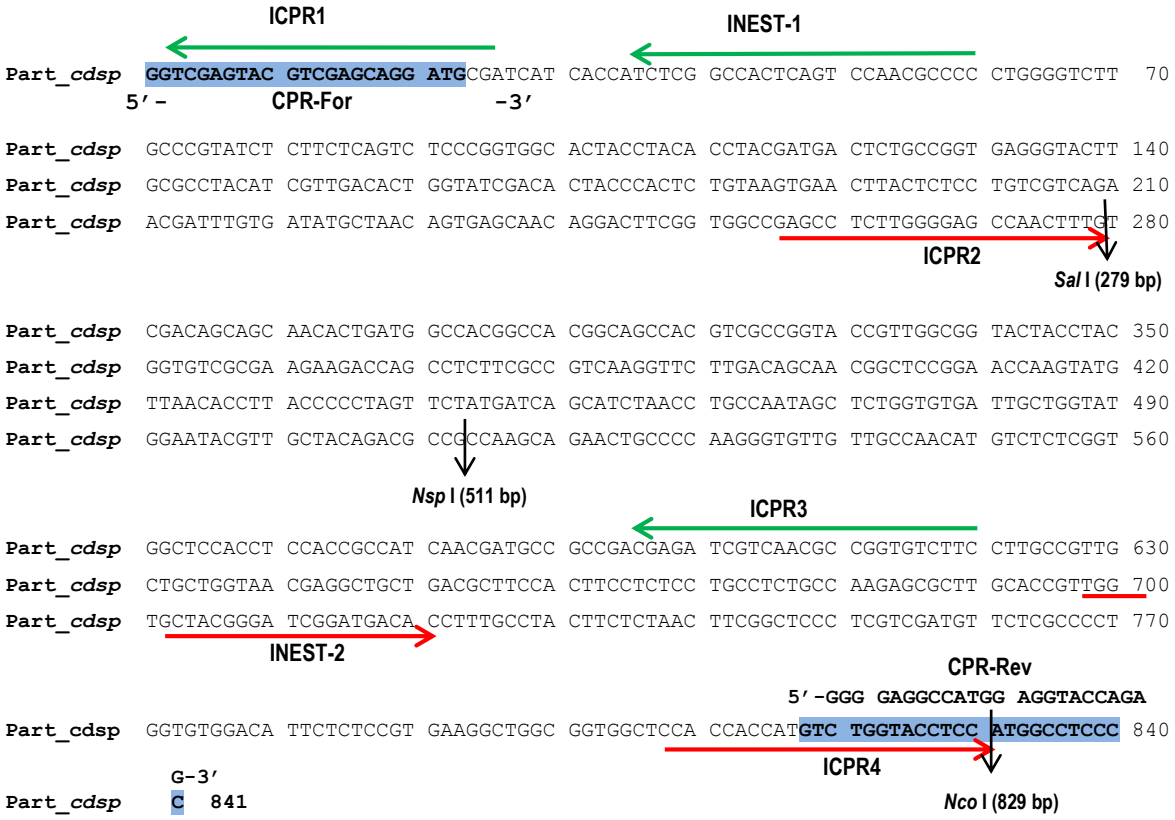


Figure 3.12. The sequence of the partial *cdsp* gene. The forward and reverse primers (CPR-For/CPR-Rev) used to amplify the partial gene are indicated and their respective primer annealing sites are highlighted in blue. The partial sequence was used a template to design three 5'-facing primers (ICPR1, INEST-1 and ICPR3; represented by green arrows) and three 3'-facing primers (ICPR2, INEST-2 and ICPR4; represented by red arrows) used to amplify the upstream and downstream gene regions, respectively.

3.6 Chromosome walking to obtain the nucleotide sequences of regions upstream and downstream of the *cdsp* gene fragment

3.6.1 Inverse PCR

Inverse PCR (Section 2.2.16.1) of the regions adjacent to the *cdsp* partial gene fragment provided fragments which were not larger than 2.5 kb (Figure 3.13). The *Sal* I-generated (1,220 bp; from INEST-1/ICPR2 primer pair) fragment was successfully cloned into pGEM®-T easvy vector (Section 2.2.14) and transformed into *E. coli* GeneHogs® competent cells (Section 2.2.15). The plasmid DNA was isolated, purified and subsequently sequenced (Section 2.2.19) using M13F and M13R sequencing primers (Section 2.1.8, Table 2.8). Other fragments: *Bgl* II (2,500 bp; INEST-1/ICPR2), *Nco* I (1,752 bp; INEST-1/INEST-2), and *Nsp* I (2,320 bp; INEST-1/ICPR2) were directly sequenced with the same primer set,

and/or with an upstream facing primer and corresponding downstream facing nested primer, as transformation efficiency was very low for the larger fragments. Fragments greater than 1,200 bp could not be fully sequenced, therefore SiteFinding-PCR was additionally used to compare with inverse PCR, which further optimized chromosome walking of the gene fragment. *Eco* RI, *Alu* I and *Xho* I (not shown) did not produce fragments with all possible primer combinations.

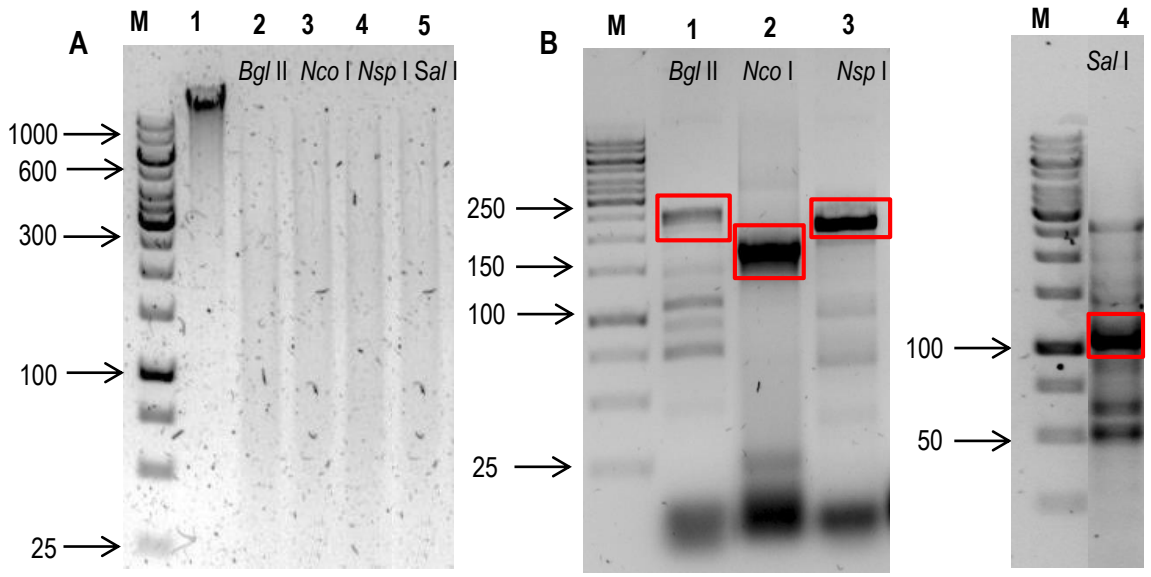


Figure 3.13. Inverse PCR. *Gel A*: Lane 1, *C. rosea* 17970 chromosomal DNA; Lane 2-4 shows the digestion of chromosomal DNA with different restriction endonucleases: Lane 2, *Bgl* II; Lane, 3 *Nco* I; Lane 4, *Nsp* I; and Lane 5, *Sal* I. *Gel B*: The different restriction endonucleases used to generate self-circularized templates for Inverse PCR led to visualization of fragments of different sizes, from which the strongest band (in red boxes) were gel purified and subsequently sequenced. M is the GeneRuler 1 kb DNA Ladder.

3.6.2 SiteFinding-PCR

SiteFinding-PCR (Section 2.2.16.2) was used to improve chromosome walking of the partial *cdsp* gene fragment. Two groups of three primers: ICPR3, INEST-1 (nested primer 1) and ICPR1 (nested primer 2), were used to amplify the sequence upstream of the 5′ region and primers ICPR4, INEST-2 (nested primer 1) and ICPR2 (nested primer 2), for the sequence downstream of the 3′ region (Figure 3.14). Flanking sequences, 525 bp and 495 bp at the 5’ and 3’ end, respectively, were obtained after the second nested PCR was performed (Figure 3.14). The fragments were gel purified, and directly sequenced using the SFP2

primer and a nested primer (for corresponding 5' or 3' sequence orientation) used to obtain the fragment (Section 2.2.19).

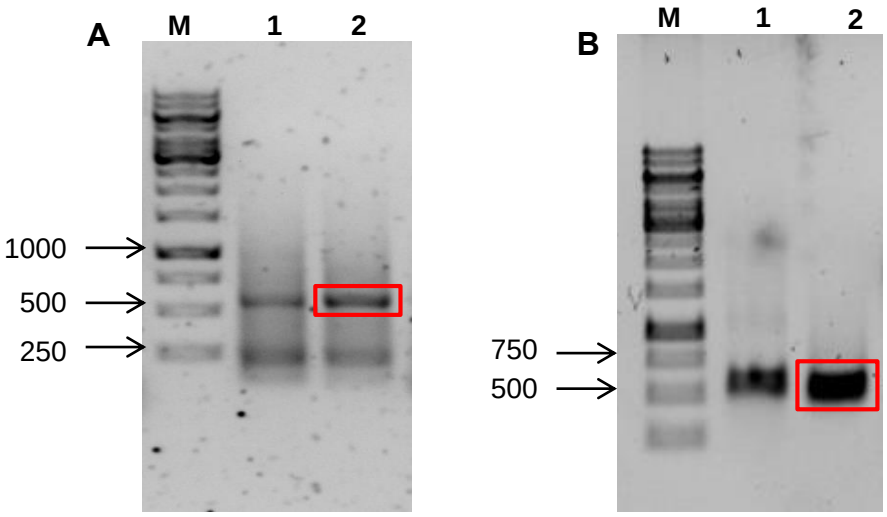


Figure 3.14. SiteFinding-PCR. *Gel A*: 5' chromosomal walking: Lane 1, First nested PCR; and Lane 2, Second nested PCR. *Gel B:* 3' chromosomal walking: Lane 1, First nested PCR; and Lane 2, Second nested PCR. Fragments which were gel extracted are shown in red boxes. M is the GeneRuler 1 kb DNA Ladder.

3.7 Sequence assembly of the complete *cdsp* gene and regulatory regions

Sequenced fragments from Inverse PCR and SiteFinding-PCR were assembled into a 2,923 bp long sequence. PCR amplification of the complete *cdsp* gene fragment was carried out using the gene specific primer pair: CPRF and CPRR (Section 2.1.8, Table 2.8) designed from the ATG start to the TAA stop codons. A fragment of 1,300 bp was amplified using *C. rosea* 17970 chromosomal DNA as template (Figure 3.15). The ORF (1,140 bp) encoding the *cdsp* gene was also amplified using a RT-PCR approach (Section 2.2.18), which identified the presence of three introns. Both fragments were visualized on a 1% (w/v) agarose gel (Figure 3.11), and subsequently sequenced (Section 2.2.19). Sequencing results for both fragments were identical to the assembled fragment.

The assembled sequence (Figure 3.16) harboured the putative motif of the peptidase S8 family (positions 110-359) identified using the NCBI conserved domain database. The cDNA sequence encodes a polypeptide of 379 amino acid residues. The predicted protein showed high similarity with serine protease amino acid sequences of *C. rosea* (Sep;

AIT70965; 99%), *C. rosea* (PrC; ACS66684; 94%) and *Acremonium chrysogenum* (Cuticle-degrading protease-like protein; KFH40949; 64%). A 15 amino acid secretion signal peptide sequence (MRVSALLSILPLVAA) was identified at the N-terminus using SignalIP. Three active site motifs: AFIIDTGIytsH (positions 137-148; aspartic acid (D) active site at position 141), HGThVAGtVGG (positions 172-182, histidine (H) active site at position 172), and GTSmAsPhVAG (positions 325-335, serine (S) active site at position 327), were found in the putative Cdsp protein sequence using PROSITE. Similarly, an Asn glycosylation motif (NMSL) at position 233-236 was identified.

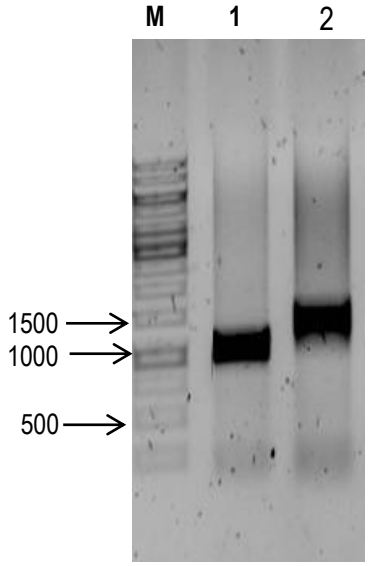


Figure 3.15. RT-PCR and PCR amplification of the complete *cdsp* ORF. Lane 1, 1,141 bp cDNA fragment; and Lane 2, 1,300 bp DNA fragment. M is the GeneRuler 1 kb DNA Ladder.

The 1,114 bp fragment, upstream of the *cdsp* ATG start codon harbours a predicted 50 bp long promoter region, 5'-CGAGACCTTCTCAAAGTCACTCTTGTGGGCCA-TCTGAATC**A**ATCATCTTT-3' (bold letter 'A' represents the putative transcription start site) at position 1034-1084, identified *in silico* using the Neural Network Promoter Prediction tool. A number of general transcription control motifs were identified upstream of the ATG start codon using different software programs (Section 2.2.20). The upstream fragment harboured one putative TATA-box (5'-TATAAA) motif at position 972 bp, and two CAAT boxes (5'-CCAAT) at positions 52 and 746 bp in the 5'-flanking region. Transcription control

binding sites and their predicted functional roles, i.e.; four putative GATA motifs (5'-GATA; at positions 572, 593, 867 and 936 bp) for nitrogen regulation, one CreA motif (5'-SYGGRG; position 113 bp) for carbon regulation, three STRE motifs (5'-AGGGG; positions 403, 408, 46 bp) associated with stress response and one PacC motif (5'-GCCARG; position 283 bp) for ambient pH regulation were also identified. Similarly, the 509 bp *cdsp* downstream region revealed one consensus yeast polyadenylation signal (5'-TATATA) at position 2611, and only three consensus sequences linked to mRNA 3' end processing signals in yeast genes, downstream of the TAA stop codon were identified. Graber *et al.* (1999) identified four types of yeast mRNA 3'-end processing sequence clusters (*cis* elements) with putative functional roles in the efficiency (type 1), positioning (type 2), pre-cleavage (type 3) and downstream (type 4) processing of mRNA (Yamanishi *et al.*, 2013). Only type 2 signal (5'-ATGTAT, position 2886) was found between two type 3 signals (5'-AAAAAA, position 2643; and 5'-TTTCTT, position 2917, respectively).

```
   1 ACCTCCCACG GCGTATAACC ACCCTTTTCC AGCTGGACCA TTGACAGTGC CCCAATGTGC
                                                          CAAT-box

  61 CGAGATCCTT CTTGACAGTC AACTAGCTGC CTATGGTGAT CAATTACCCG TCCCGGGGTT
                                                          CreA

 121 CCCCATTAGA TCGGGGGAAA TCATTGGCAG ACCCCCGGTC GGAAGATGGG AAGGTGACAT
 181 ATGGCACTAT TGTCCCCTTT CTCGCTTTCC AGCTGCTTCT AACTTCCTAT TTGGACCCCG
 241 CGTCTTGACA AATGGTCCAA GAGGTTTCTT TTCTAGTTTG GAGCCAAGCA TTTTTGTTGG
                                                   PacC

 301 CTAATCGATT CATATCTCAA GTGTTAACCC AACCCAGGAG CATCTTATCA TGTGCGCCGT
 361 CAGGGTGTTA TCATGACGTT CCATTCTCTA AGAGAATCCC GCAGGGGAGG GGCCTGGCCT
                                                   STRE (X2)

 421 GGACCTGGCC CCCAAAGCCA AACACGAAAC TACCTATGCT GAACAAGGGG AAGTCACGGG
                                                   STRE

 481 TGAATTTGCC GAGTAGCCAT GCTTAAGCTT TTGTCTCTGG TAGTGTGACA CCCGAGTCGA
 541 TTCGAAGTGG TGTTTGGTAG TAACAACGCG GGATACCAAC GAACCTACAC GAGATGATC
                                   GATA transcription factor  GATA transcription factor

 601 TCTTATCAGA AATCGCTGAT CGGCACTCAT AGTACCATTA TTCTAGAGCT CAGCGGCAAA
 661 ACAGTCCTGC TTACCAGCCA AAAACCGGAT TATACAGCCC GTATCTGTGG GTTTGTCGTA
 721 TATGACTGTC TGATCAATAT CTTGACCAAT ATCAATCCAT TGGTCCAAGC TTCTTTAGGG
                                   CAAT-box

 781 AGGAGGACCC TTTGTCAACG TAGGACCTGT CAGTATGCAG TAGCCCTGCG TCCAGCAGCC
 841 TGCCGTATTT GGTTGATTGC AACCAAGATA CGAACATGAC TGCACGCGAG TTGACAAGGG
                                   GATA transcription factor

 901 AAAAGGGAAG GTGAGGGCGT TCGCGAGTAT TACATGATAG ACGGGACAAC AAAGACCTCA
                                   GATA transcription factor

 961 CGATCCTCAA CTATAAAGAC TACCCGTTCT CCTCTCCATC TCTTCTTTTT CTTTCTTTCT
                TATA-box

1021 CCTCACCAGC AATCGAGACC TTCTCAAAGT CACTCTTGTG GGCCATCTGA ATCAATCATC
                                                   CPRF

1081 TTTTGTGTAC TTCACCAAGC AACAAACCGT CAACATGCGC GTTTCTGCTC TCCTCTCCAT

   1                                              M  R  V  S  A  L  S  I

1141 CCTGCCTCTG GTTGCCGCGG CACCAGCCAA GCGCGAGGAAGT CGCTCCTC TCCACGTCCC

  10  L  P  L  V  A  A  A  P  A  K  R  E  E  V  A  P  L  H  V  P

1201 CCGCGACGTC GAGGTCATCC CCGGAAAGTA CATTGTCAAG CTCAAGGAGG GTGTTGTTTC
  30  R  D  V  E  V  I  P  G  K  Y  I  V  K  L  K  E  G  V  V  S

1261 CATCTCCTCC ACCATCTCGT CCATTGAGGC CAAGCCCGAC TTTGAGTACG AGGGTGGCTT
  50  I  S  S  T  I  S  S  I  E  A  K  P  D  F  E  Y  E  G  G  F

1321 CCAGGGTTTC GCTGGTGCCC TGACTGAGGC TGAGGTCCAG GCCCTCCGTG AGAGCCCCGA
  70  Q  G  F  A  G  A  L  T  E  A  E  V  Q  A  L  R  E  S  P  E
      Intron 1
1381 Ggtaagatcc cccatcacac cactcaagcc aatatgctaa cgtcccttct tcagGTCGAG
  90                                                              V  E
1441 TACGTCGAGC AGGATGCCAT CGTCTCCATC TCCGCCACCC AGACCGGAGC TCCATGGGGC
  92  Y  V  E  Q  D  A  I  V  S  I  S  A  T  Q  T  G  A  P  W  G
1501 ATCGCTCGTC TCTCCAACAC CAACACTGGT TCCACCACCT ACACCTACGA CGACTCTGCT
 112  I  A  R  L  S  N  T  N  T  G  S  T  T  Y  T  Y  D  D  S  A
                                                                 Intron 2
1561 GGTGACGGCA CCTGCGCCTT CATCATTGAC ACCGGCATCT ACACCTCCCA CACgtaagat
 132  G  D  G  T  C  A  F  I  I  D  T  G  I  Y  T  S  H  T
1621 tcccccatca caccactcaa gccaatatgc taacgtccct tcttcagCGA CTTCGGTGGC
 150                                                    D  F  G  G
1681 CGTGCCTCCT TCGCCGCCAA CTACGTTGAC AGCAGCAGCA CCGATGGCAA TGGCCACGGC
 154  R  A  S  F  A  A  N  Y  V  D  S  S  S  T  D  G  N  G  H  G
```

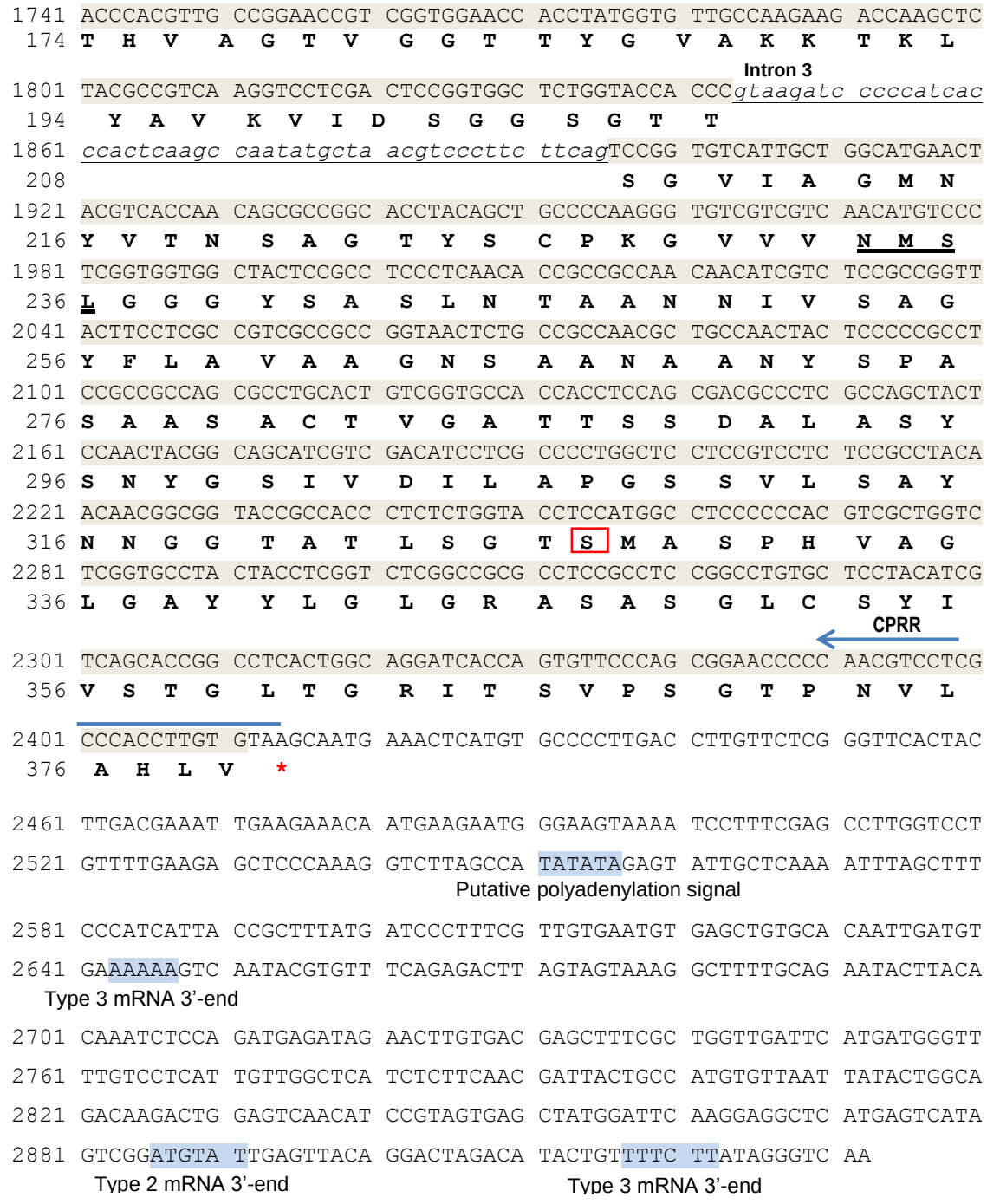


Figure 3.16. The complete assembled nucleotide sequnece of the *cdsp* gene with regulatory regions. The metheonine start codon (position 1115; in red) and stop codon (position 1907; indicated by a red asterisk(*)) are shown. The putaitve transcription start site is highlighted in yellow. Consensus transcription regulatory motifs were identified, i.e.; one TATA-box (position 972; bold and double underlined), two CAAT-boxes (positions 52 and 746 ; bold, italicized and boxed); one CreA (positionn 113; bold, wave underlined); one PacC (position 283; dash underlined); three STRE (positions 403, 408 and 466; italicized and underlined), and four GATA transcription factors (positon 572, 593, 867 and 936; bold and italicized). The three introns: intron 1 (position 305-372), intron 2 (position 598-933) and intron 3 (1844-1895) are shown in lower case, italicized and underlined. The 15-amino acid long signal peptide is highlighted in green within the coding region. A putative polyadelation signal (position 2550) and sequences downstream of *cdsp* associated with mRNA 3'-end processing (positions 2643,2886 and 2917) are highlighted in light blue. The conserved serine protease catalytic triad (Asp, position 141; His, position 172; and Ser, position 327) and putative glycosylation motif (NMSL, positions 233-236) within the putative Cdsp protein sequence is represented in red boxes and thick underlined, respectively.

CHAPTER 4: DISCUSSION AND CONCLUDING REMARKS

Screening of fungal isolates (155) from the CMW culture collection in ME liquid medium identified CMW 17970 as a possible candidate for high levels of protein production. Malt extract was chosen for screening because it is a general purpose medium which supports the growth of most fungi and yeasts. The original isolate was found on the South African endemic grass-like plant, *Restio filiformis* in the Fynbos, and was identified as *Erythrogloeum sp.* However, the isolate obtained from the culture collection was a mixture of two different fungi and after purification; the protein producing isolate was identified as *C. rosea* using culture, morphology, and molecular characterization.

LC-MS/MS peptide sequencing and analysis identified the ~35 kDa protein, produced by the *C. rosea* 17970 isolate, as a putative extracellular cuticle-degrading serine protease (Cdsp) (Li *et al.*, 2006; Liang *et al.*, 2011). Following the identification of Cdsp, culture dependent experiments were carried out to develop culture conditions for improved protein production. Initially, mycelial plugs (< 5 mm^2) were used as inoculum in the screening experiment, during which the Cdsp protein band was observed after 7 dpi. Spore suspensions were then prepared and used as inoculum to compare the effect of spores versus mycelia as inoculum type for protein production. It was found that spore inoculum concentration of 1×10^7 conidia/ml was optimum for Cdsp production, and also shortened production to 5 dpi, two days faster than using mycelia as inoculum. It was also found that the concentration of spores used as inoculum impacted on the type of growth morphology produced. Freely dispersed mycelia were associated with spore inoculum concentration of $\geq 10^7$ conidia/ml and pellets were observed for spore concentrations ($\leq 10^6$ conidia/ml), as well as for mycelial inocula. Previous investigations on the effect of inocula on protein production were carried out with many other fungi. For example, the production of penicillin from *P. chrysogenum* requires freely dispersed mycelia (inocula $> 10^4$ spores/ml), whereas production of citric acid from *A. niger* favours pelleted morphology (inocula $< 10^4$ spores/ml) (Vecht-Lifshitz *et al.*, 1990; Papagianni, 2004; Wang *et al.*, 2005).

Different media were then tested for improved protein production. It was found that the type of media used also affected the growth morphology of *C. rosea* 17970 and Cdsp protein production. From the eight media tested, ME was found to be the best. *C. rosea* 17970 grew as freely dispersed mycelia only in ME, whereas the Cdsp protein-limiting pellets were observed in all other media (data not shown).

The production of Cdsp was found to be optimum when ME was adjusted to pH 9.0 before inoculation. However, the pH of the culture shifted down to ~ pH 7 after 3 dpi, and remained constant thereafter. Similarly, pH shifts were observed when *C. rosea* 17970 was grown at lower pH values, i.e.; an initial pH of 4.8 increased to ~6 after 3 days. The growth medium for protein production is required to buffer the culture against pH changes, however ME is regarded to have a low buffering capacity (Knapp & Eveleigh, 1973) possibly causing the observed pH shifts. The pH shift from 9.0 to ~7 could also be due to the accumulation of dissolved bicarbonate ions formed from high levels of carbon dioxide gas (Papagianni, 2004). The effect of temperature on Cdsp protein production was the last parameter tested. It was found that *C. rosea* 17970 produced high levels of protein over a broad temperature range (between 28 and 32°C). When grown at optimum conditions: inoculum of 1 x 10^7 conidia/ml, initial pH of 9, and incubation at 32°C; Cdsp was produced consistently at high levels from day 3, up to day 12 post inoculation.

Nucleotide sequences resulting from the BLAST analysis and *in silico* reverse translation of the identified Cdsp peptide sequences were aligned with *PrC* (NCBI accession number: GQ149467) and *Sep* (NCBI accession number: KM199682) genes. The multiple sequence alignment identified conserved nucleotide regions which were used to design primers to amplify an 841 bp long internal *cdsp* gene fragment. Two chromosome walking methods, Inverse PCR (Ochman *et al.*, 1988) and SiteFind PCR (Tan *et al.*, 2005), were successfully used to obtain a 2,923 bp long fragment which contains the complete *cdsp* (1,141 bp) gene as well as upstream and downstream regulatory regions. The cDNA obtained from *C. rosea* 17970 RNA, confirmed the actual size of the *cdsp* gene (1,300 bp) and identified three introns.

The deduced Cdsp protein sequence (379 amino acids) showed high similarity with other serine protease amino acid sequences from *C. rosea* (Sep; AIT70965; 99%), *C. rosea* (PrC; ACS66684; 94%) and *Acremonium chrysogenum* (Cuticle-degrading protease-like protein; KFH40949; 64%) through BLASTN analysis. Moreover, the protein harbours a conserved signature motif of the peptidase S8 family found in many other serine proteases when screening the NCBI-CDD database. Briefly, serine proteases belong to a family of enzymes characterized by the presence of a serine group in their active site (Siezen & Leunissen, 1997; Rao *et al.*, 1998). The two largest groups of serine proteases belong to the chymotrypsin (S1) and subtilisin (S8) subfamilies. Both subfamilies possess a serine, histidine and aspartate (Ser/His/Asp) catalytic triad that possesses conserved spatial arrangement in the sequence, however, the order of the residues in the amino acid sequence and tertiary structure can differ (Rawlings & Barrett, 1993; Zanphorlin *et al.*, 2011). The conserved serine, histidine and aspartate active site motifs of serine proteases were identified in Cdsp via PROSITE and a 15 amino acid long signal peptide was identified using SignalP, indicating that this protein is a secreted protease.

Flanking regions of most functional genes in eukaryotic organisms consist of various important gene regulatory elements. The upstream 5' region (1,114 bp) of the *cdsp* coding region was identified as a putative promoter region and through *in silico* sequence analysis, important 5'-flanking regulatory motifs were identified. The promoter harbours a well-defined TATA and two CAAT boxes (Unkles, 1992; Ato, 2005). In addition, DNA-binding sites for transcription factors that are involved in the regulation of different pathways, i.e.; stress response (STRE), ambient pH regulation (PacC), carbon (CreA) and nitrogen (GATA) metabolism were also identified.

Similar to the promoter region, the region downstream (509 bp) of the *cdsp* TAA stop codon was analysed for conserved elements associated with translational processing and regulation. Inspection of the 3' region did not include any consensus eukaryotic polyadenylation signals (5'-AATAAA-3'), however a 5'-TATATA-3' signal that is mostly frequent in yeast was identified (Guo & Sherman, 1995). In addition, several yeast mRNA 3' end processing signals were

detected in the putative 3' untranslated region. Due to the limited information on filamentous fungal terminator elements and their respective role in protein production (Su *et al.*, 2011), a thorough analysis was not possible. Therefore, this area of study will have to be further investigated.

As with current commercial filamentous fungal isolates (e.g.; *Trichoderma* spp., *Aspergillus* spp. and *M. thermophila*), the identification of a new isolate (such as *C. rosea* 17970) is a small milestone compared to the experimental work required to determine its capability for protein production. The required experiments include:

(i) Validation of the promoter (P*cdsp*), signal peptide sequence and terminator (T*cdsp)* regions for heterologous protein production of a reporter protein (e.g.; GFP) in a constructed expression cassette. This step is significant, as it will demonstrate the strength of the regulatory and secretory elements for potential research or commercial use.

(ii) Optimization of growth conditions for medium-to-large scale production of Cdsp via submerged fermentation could confirm the versatility of *C. rosea* 17970 as an expression host.

(iii) Strain development strategies (i.e., UV-light mutagenesis and screening for hyper-expressing clones) could produce strains with enhanced secretion capability. The performance of the mutants would be compared with the parent strain (*C. rosea* 17970) cultured under the same conditions.

(iv) Development of a *C. rosea* 17970 transformation method would be a critical milestone for the development of the future value of this study.

Taken together, the results of this study have illustrated that *C. rosea* 17970 may have physiological and genetic potential to be developed into a host-vector system. Previous studies

using *C. rosea* have largely concentrated on optimizing its efficacy as a biocontrol agent (Roberti et al., 2008; Toledo et al., 2011; Lahlali & Peng, 2014) rather than understanding its secretion and physiological capability for protein production. This work is the first attempt to investigate an isolate belonging to this species as a potential protein expression system. In conclusion, this thesis has produced a potential fungal isolate for protein production, provided some understanding regarding the development of a new fungal expression host/vector system and recommended research for future development.

REFERENCES

Adrio, J. L. & Demain, A. L., 2014. Microbial Enzymes: Tools for Biotechnological Processes. *Biomolecules*, **4**, pp.117–139.

Agrawal, V. & Bal, M., 2012. Strategies for Rapid Production of Therapeutic Proteins in Mammalian Cells. *BioProcess International*, **10**, pp.32–48.

Ahamed, A. & Vermette, P., 2009. Effect of Culture Medium Composition on *Trichoderma Reesei*'s Morphology and Cellulase Production. *Bioresource Technology*, **100**, pp.5979–87.

Ahmed, M., Laing, M. D. & Nsahlai, I. V, 2014. Use of *Clonostachys Rosea* against Sheep Nematodes Developing In Pastures. *Biocontrol Science and Technology*, **24**, pp.389–398.

Altman, L. K., 1982. A New Insulin Given Approval for Use in U.S. New York Times.

Altschul, S. F, Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **215**, pp.403–410.

AMFEP, 2014. List of Commercial Enzymes from the Association of Manufacturers and Formulators of Enzyme Products (AMFEP) Available at: http://www.amfep.org/content/list-enzymes [Accessed May 10, 2015].

Andrie, R. M., Martinez, J. P. & Ciuffetti, L. M., 2005. Development of ToxA and ToxB Promoter-Driven Fluorescent Protein Expression Vectors for Use in Filamentous Ascomycetes. *Mycologia*, **97**, pp.1152–1161.

Anne, J., Maldonado, B., Van, I. J., Van, M. L. & Bernaerts, K., 2012. Recombinant Protein Production and *Streptomycetes*. *Journal of Biotechnology*, **158**, pp.159–167.

Anyaogu, D. C. & Mortensen, U. H., 2015. Heterologous Production of Fungal Secondary Metabolites in *Aspergillus*. *Frontiers in Microbiology*, **6**.

Archer, D. B., Jeenes, D. J., MacKenzie, D. A., Brightwell, G., Lambert, N., Lowe, G., Radford, S. E. *et al*., Hen Egg White Lysozyme Expressed in, and Secreted From, *Aspergillus niger* is Correctly Processed and Folded. *Biotechnology,* **8**, pp.741–745.

Archer, D. B., Peberby, J. F. & Peberdy, J. E., 1997. The Molecular Biology of Secreted Enzyme Production by Fungi. *Critical Reviews in Biotechnology*, **17**, pp.273–306.

Aro, N., Ilmen, M., Saloheimo, A. & Penttila, M., 2003. ACEI of *Trichoderma reesei* is a Repressor of Cellulase and xylanase Expression. *Applied Environmental Microbiology*, **69**, pp.56–65.

Aro, N., Saloheimo, A., Ilmen, M. & Penttila, M., 2001. ACEII, A Novel Transcriptional Activator Involved in Regulation of Cellulase and Xylanase Genes of *Trichoderma reesei*. *The Journal of Biological Chemistry*, **276**, pp.24309–24314.

Ato, M. K., 2005. An Overview of the CCAAT-Box Binding Factor in Filamentous Fungi: Assembly, Nuclear Translocation, and Transcriptional Enhancement. *Bioscience, Biotechnology, and Biochemistry*, **69**, pp.663–672.

Ausubel, F., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., 1998. Current Protocols in Molecular Biology, New York, USA: John Wiley and Sons Inc.

Bachran, C., Abdelazim, S., Fattah, R. J., Liu, S. & Leppla, S. H. 2013. Recombinant Expression and Purification Of A Tumor-targeted Toxin in *Bacillus anthracis*. *Biochemical and Biophysical Research Communications*, **430**, pp.150–155.

Ballance, D. J., 1986. Sequences Important For Gene Expression in Filamentous. *Yeast*, **2**, pp.229–236.

Barr, D. J. S., 1978. Taxonomy and Phylogeny of Chytrids. *Biosystems*, **10**, pp.153–165.

Barrios-González, J., 2012. Solid-State Fermentation: Physiology of Solid Medium, Its Molecular Basis and Applications. *Process Biochemistry*, **47**, pp.175–185.

BBC Research, 2014a. In Report BIO009 F - Global Markets for Bioengineered Protein Drugs, Wellesley, USA.

BBC Research, 2014b. In Report BIO030 H - Global Markets for Enzymes in Industrial Applications, Wellesley, USA.

Bennett, J. W., 1998. Mycotechnology: The Role of Fungi in Biotechnology. *Journal of Biotechnology*, **66**, pp.101–107.

Bergquist, P., Valentino, T., Cziferszky, A., Paula, F., Azevedo, M. & Nevalainen, H., 2002. Expression of Xylanase Enzymes from Thermophilic Microorganisms in Fungal Hosts. *Extremophiles : Life under Extreme Conditions*, **6**, pp.177–84.

Berka, R. M. & Barnett, C. C., 1989. The Development of Gene Expression Systems for Filamentous Fungi. *Biotechnology Advances*, **7**, pp.127–154.

Bill, R. M., Henderson, P. J. F., Iwata, S., Kunji, E. R. S., Michel, H., Neutze, R., Newstead, S. *et al*., 2011. Overcoming Barriers to Membrane Protein Structure Determination. *Nature Biotechnology*, **29**, pp.335–340.

Bill, R. M., 2014a. Playing Catch-Up with *Escherichia coli* : Using Yeast to Increase Success Rates in Recombinant Protein Production Experiments. *Frontiers in Microbiology*, **5**.

Bill, R. M., 2014b. Recombinant Protein Subunit Vaccine Synthesis in Microbes : A Role for Yeast? *Journal of Pharmacy and Pharmacology*, **63**, pp.319–328.

Bodén, M. & Bailey, T. L., 2008. Associating Transcription Factor-Binding Site Motifs with Target GO Terms and Target Genes. *Nucleic Acids Research*, **36**, pp.4108–4117.

Bohlin, C., Jonsson, L. J., Roth, R. & van Zyl, W. H., 2006. Heterologous Expression of *Trametes versicolor* Laccase in *Pichia pastoris* and *Aspergillus niger*. *Applied Biochemistry and Biotechnology*, **129**, pp.195–214.

Bouws, H., Wattenberg, A. & Zorn, H., 2008. Fungal Secretomes – Nature's Toolbox for White Biotechnology. *Applied Microbiology and Biotechnology*, **80**, pp.381–388.

Brunt, J. V., 1986. Fungi: The Perfect Hosts? *Nature Biotechnology*, **4**, pp.1057–1062.

Buske, F. A., Bodén, M., Bauer, D. C. & Bailey, T. L., 2010. Assigning Roles to DNA Regulatory Motifs Using Comparative Genomics. *Bioinformatics*, **26**, pp.860–866.

Cereghino, G. P. & Cregg, J. M., 1999. Applications of Yeast in Biotechnology: Protein Production and Genetic Analysis. *Current Opinion in Biotechnology*, **10**, pp.422–427.

Chatterton, S., Jayaraman, J. & Punja, Z. K., 2008. Colonization of Cucumber Plants by the Biocontrol Fungus *Clonostachys rosea f. catenulata*. *Biological Control*, **46**, pp.267–278.

Chatterton, S. & Punja, Z. K., 2009. Chitinase and β-1,3-glucanase Enzyme Production by The Mycoparasite *Clonostachys rosea f. catenulata* Against Fungal Plant Pathogens. *Canadian Journal of Microbiology*, **55**, pp.356–367.

Chen, X. I., Liang, Y., Hua, J., Tao, L., Qin, W. & Chen, S., 2010. Overexpression of bacterial ethylene-forming enzyme gene in *Trichoderma reesei* enhanced the production of ethylene. *International Journal of Biological Sciences*, **6**, pp.96–106.

Cheon, S. A., Kim, H., Oh, D. B., Kwon, O. & Kang, H. A., 2012. Remodeling of the Glycosylation Pathway in the Methylotrophic Yeast *Hansenula polymorpha* To Produce Human Hybrid Type N-Glycans. *Journal of Microbiology*, **50**, pp.341–348.

Choi, J. H. & Lee, S. Y., 2004. Secretory and Extracellular Production of Recombinant Proteins Using *Escherichia coli*. *Applied Microbiology and Biotechnology*, **64**, pp.625–635.

Chou, C. P., 2007. Engineering Cell Physiology to Enhance Recombinant Protein Production in *Escherichia coli*. *Applied Microbiology and Biotechnology*, **76**, pp.521–532.

Cohen, S., Chang, A. C., Boyer, H. W., Helling, R. B., 1973. Construction of Biologically Functional Bacterial Plasmids *In Vitro*. *Proceedings of the National Academy of Sciences of the USA*, **70**, pp.3240–3244.

Cohen, S., 2013. DNA Cloning: A Personal View after 40 Years. *Proceedings of the National Academy of Sciences of the USA*, **110**, pp.15521–15529.

Conde, R., Cueva, R., Pablo, G., Polaina, J. & Larriba, G., 2004. A Search for Hyperglycosylation Signals in Yeast Glycoproteins. *Journal of Biological Chemistry*, **279**, pp.43789–43798.

Conesa, A., Punt, P. J., van Lujik, N., & van den Holden, C. A. M. J. J., 2001. The Secretion Pathway in Filamentous Fungi : A Biotechnological View. *Fungal Genetics and Biology*, **33**, pp.155–171.

Coradetti, S. T., Craig, J. P., Xiong, Y., Shock, T., Tian, C. & Glass, N. L., 2012. Conserved and Essential Transcription Factors for Cellulase Gene Expression in Ascomycete Fungi. *Proceedings of the National Academy of Sciences of the USA*, **109**, pp. 7397-7402.

Corchero, J. L., Gasser, B., Resina, D., Smith, W., Parrilli, E., Vázquez, F., Abasolo, I. *et al*., 2013. Unconventional Microbial Systems for the Cost-Efficient Production of High-Quality Protein Therapeutics. *Biotechnology Advances*, **31**, pp.140–153.

Cota, L. V., Maffia, L. A., Mizubuti, E. S. G. & Macedo, P. E. F., 2009. Biological Control by *Clonostachys rosea* as a Key Component in the Integrated Management of Strawberry Gray Mold. *Biological Control*, **50**, pp.222–230.

Curran, B. & Begeja, V., 2011. The Biotechnological Exploitation of Heterologous Protein Production in Fungi. In K. Kavanagh, ed. *Fungi: Biology and Applications*. London, UK: John Wiley & Sons, Inc., pp. 205–227.

Curran, K. A., Morse, N. J., Markham, K. A., Wagman, A. M., Gupta, A. & Alper, H. S., 2015. Short Synthetic Terminators for Improved Heterologous Gene Expression In Yeast. *ACS Synthetic Biology*, **4**, pp. 824-832.

Curran, K. A., Karim, A. S., Gupta, A. & Alper, H. S., 2013. Use of Expression-Enhancing Terminators in *Saccharomyces cerevisiae* to Increase mRNA Half-Life and Improve Gene Expression Control for Metabolic Engineering Applications. *Metabolic Engineering*, **19**, pp.88–97.

Dai, Z., Mao, X., Magnuson, J. K. & Lasure, L. L., 2004. Identification of Genes Associated with Morphology in *Aspergillus niger* by Using Suppression Subtractive Hybridization. *Applied and Environmental Microbiology*, **70**, pp.2474–2485.

Davey, J. & Ladds, G., 2011. Heterologous Expression of GPCRs in Fission Yeast. *Methods in Molecular Biology*, **746**, pp.113–131.

Davies, R. W., 1994. Heterologous Gene Expression and Protein Secretion in *Aspergillus*. *Progress in Industrial Microbiology*, **29**, pp.527–560.

Demain, A. L., 2007. The Business of Biotechnology. *Industrial Biotechnology*, **3**, pp.269–283.

Demain, A. L. & Vaishnav, P., 2009. Production of Recombinant Proteins by Microbes and Higher Organisms. *Biotechnology Advances*, **27**, pp.297–306.

Demir, A., Aytar, P., Gedikli, S., Çabuk, A., Arisoy, M., Aytar, P., Çabuk, A. *et al*., 2011. Laccase Production with Submerged and Solid State Fermentation: Benefit and Cost. *Hacettepe Journal of Biology & Chemistry*, **39**, pp.305–313.

Dijl, J. M. Van & Hecker, M., 2013. *Bacillus subtilis*: From Soil Bacterium To Super-Secreting Cell Factory. *Microbial Cell Factories*, **12**.

Driouch, H., Hänsch, R., Wucherpfennig, T., Krull, R. & Wittmann, C. l., 2012. Improved Enzyme Production by Bio-Pellets of *Aspergillus niger*: Targeted Morphology Engineering using Titanate Microparticles. *Biotechnology & Bioengeneering*, **109**, pp.462–471.

Driouch, H., Sommer, B. & Wittmann, C., 2010. Morphology Engineering of *Aspergillus niger* for Improved Enzyme Production. *Biotechnology & Bioengeneering*, **105**, pp.1058–1068.

Dubé, E., Shareck, F., Hurtubise, Y., Daneault, C. & Beauregard, M., 2008. Homologous Cloning, Expression, and Characterisation of a Laccase from *Streptomyces coelicolor* and Enzymatic Decolourisation of an Indigo Dye. *Applied Microbiology and Biotechnology*, **79**, pp.579–603.

Durand, H., Clanet, M. & Tiraby, G., 1988. Genetic Improvement of *Trichorderma reesei* for Large-Scale Cellulase Production. *Enzyme and Microbial Technology*, **10**, pp.341–345.

El-Enshasy, A. H., 2007. Filamentous Fungal Cultures – Process Characteristics, Products, and Applications. In S. Yang, ed. *Bioprocessing For Value-Added Products From Renewable Resources, New Technologies And Applications*. London, UK: Elsevier, pp. 225–261.

Elinbaum, S., Ferreyra, H., Ellenrieder, G. & Cuevas, C., 2002. Production of *Aspergillus terreus* Alpha-L-Rhamnosidase by Solid State Fermentation. *Letters in Applied Microbiology*, **34**, pp.67–71.

Emmerstorfer, A., Wriessnegger, T., Hirz, M. & Pichler, H., 2014. Overexpression of Membrane Proteins from Higher Eukaryotes in Yeasts. *Applied Microbiology and Biotechnology*, **98**, pp.7671–7698.

Favaro, L., Jooste, T., Basaglia, M., Jose, S. H., Saayman, M., Görgens, J. F., Casella, S. *et al*., 2013. Designing Industrial Yeasts for the Consolidated Bioprocessing Of Starchy Biomass to Ethanol. *Bioengineered*, **4**, pp.97–102.

Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E. & Villaverde, A., 2009. Microbial Factories for Recombinant Pharmaceuticals. *Microbial Cell Factories*, **8**.

Ferrer-Miralles, N. & Villaverde, A., 2013. Bacterial Cell Factories for Recombinant Protein Production; Expanding the Catalogue. *Microbial Cell Factories*, **12**.

Fincham, J. R. S., 1989. Transformation in Fungi. *Microbiological Reviews*, **53**, pp.148–170.

Finkelstein, D. B., 2013. Transformation. In D. B. Finkelstein, ed. *Biotechnology of Filamentous Fungi: Technology and Products*. South Wales, UK: Newnes, pp. 113–156.

Fleißner, A. & Dersch, P., 2010. Expression and Export: Recombinant Protein Production Systems for Aspergillus. *Journal of Applied Microbiology*, **87**, pp.1255–1270.

Fowler, T. & Berka, R. M., 1991. Gene Expression Systems for Filamentous Fungi. *Current Opinion in Biotechnology*, **2**, pp.691–697.

Gardes, M. & Bruns, T. D., 1993. ITS Primers with Enhanced Specificity of Basidiomycetes: Application to the Identification of Mycorrhizae and Rusts. *Molecular Ecology*, **2**, pp.113–118.

Gasser, B., Prielhofer, R., Marx, H., Maurer, M., Nocon, J., Steiger, M., Puxbaum, V. *et al*., 2013. *Pichia pastoris*: Protein Production Host and Model Organism for Biomedical Research. *Future Microbiology*, **8**, pp.191–208.

Gasser, B. & Mattanovich, D., 2007. Antibody Production with Yeasts and Filamentous Fungi: On the Road to Large Scale? *Biotechnology Letters*, **29**, pp.201–212.

Genentech, 1982. First Recombinant DNA Product Approved by the Food and Drug Administration. Press Release.

Gerngross, T. U., 2004. Advances in the Production of Human Therapeutic Proteins in Yeasts and Filamentous Fungi. *Nature Biotechnology*, **22**, pp.1409–1414.

Gibbs, P. A., Seviour, R. J. & Schmid, F., 2000. Growth of Filamentous Fungi in Submerged Culture: Problems and Possible Solutions. *Critical Reviews in Biotechnology,* **20**, pp.17–48.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F. *et al*., 1996. Life with 6000 Genes. *Science*, **274**, pp.563–547.

Gomi, K., Akeno, T., Minetoki, T., Ozeki, K., Kumagai, C., Okazaki, N. & Iimura, Y., 2000. Molecular Cloning and Characterization of a Transcriptional Activator Gene, Amyr, Involved in the Amylolytic Gene Expression in *Aspergillus oryzae*. *Bioscience, Biotechnology, and Biochemistry*, **64**, pp.816–827.

Gong, Y., Hu, H., Gao, Y., Xu, X. & Gao, H., 2011. Microalgae as Platforms for Production of Recombinant Proteins and Valuable Compounds: Progress and Prospects. *Journal of Industrial Microbiology*, **38**, pp. 1879-1890.

Graber, J., Cantor, C. R., Mohr, S. C., Smith, T. F., 1999. Genomic Detection of New Yeast Pre-mRNA 3'-end-Processing Signals. *Nucleic Acids Research*, **27**, pp.888–894.

Graber, J. H., McAllister, G. D. & Smith, T. F., 2002. Probabilistic Prediction of *Saccharomyces cerevisia*e mRNA 3'-Processing Sites. *Nucleic Acids Research*, **30**, pp.1851–1858.

Grimm, L., Kelly, S., Krull, R. & Hempel, D. C., 2005. Morphology and Productivity of Filamentous Fungi. *Applied Microbiology and Biotechnology*, **69**, pp.375–384.

Guo, Z. & Sherman, F., 1995. 3'-End-Forming Signals of Yeast mRNA. *Molecular and Cellular Biology*, **15**, pp.5983–5990.

Gusakov, A. V, Semenova, M. V & Sinitsyn, A.P., 2010. Mass Spectrometry in the Study of Extracellular Enzymes. *Journal of Analytical Chemistry*, **65**, pp.1446–1461.

Gwynne, D. I., Buxton, F. P., Williams, S. A., Sills, A. M., Johnstone, J. A., Buch, J K., Guo, Z. M. *et. a*l., 1989. Development of an Expression System in *Aspergillus nidulans*. *Biochemical Society Transactions*, **17**, pp.338–340.

Hall, T.A., 1999. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, pp.95–98.

Hamilton, S. R., Davidson, R. C., Sethuraman, N., Nett, J. H., Jiang, Y., Rios, S., Bobrowicz, P. *et al.*, 2006. Humanization of Yeast to Produce Complex Terminally Sialylated Glycoproteins. *Science*, **313**, pp.1441–1443.

Hamilton, S. R, Bobrowicz, P., Bobrowicz, B., Davidson, R. C., Li, H., Mitchell, T., Nett, Juergen H. *et a*l., 2003. Production of Complex Human Glycoproteins in Yeast. *Science*, **301**, pp.1244–1246.

Hankin, L. & Anagnostakis, S. L., 1975. The Use of Solid Media for Detection of Enzyme Production by Fungi. *Mycologia*, **67**, pp.597–607.

Harkki, A., Mäntylä, A., Penttilä, M., Muttilainen, S., Bühler, R., Suominen, P., Knowles, J. *et al*., 1991. Genetic Engineering of *Trichoderma* to Produce Strains with Novel Cellulase Profiles. *Enzyme and Microbial Technology*, **13**, pp.227–233.

Hawksworth, D. L., 2012. Global Species Numbers of Fungi: Are Tropical Studies and Molecular Approaches Contributing to a More Robust Estimate? *Biodiversity and Conservation*, **21**, pp.2425–2433.

Hayakawa, Y., Ishikawa, E., Shoji, J., Nakano, H. & Kitamoto, K., 2011. Septum-Directed Secretion in the Filamentous Fungus *Aspergillus oryzae*. *Molecular Microbiology*, **81**, pp.40–55.

Hegde, R. S. & Bernstein, H. D., 2006. The Surprising Complexity of Signal Sequences. *Trends in Biochemical Science*, **31**, pp.563–567.

von Heijne, G., 1990. The Signal Peptide. *Journal of Membrane Biology*, **115**, pp.195–201.

Hitzeman, R. A., Hagie, F. E., Levine, H. L., Goeddel, D. V., Ammerer, G. & Hall, B. D., 1981. Expression of a Human Gene for Interferon in Yeast. *Nature*, **293**, pp.717–722.

Idiris, A., Tohda, H., Sasaki, M. & Okada, K., 2010. Enhanced Protein Secretion from Multiprotease-Deficient Fission Yeast by Modification of Its Vacuolar Protein Sorting Pathway. *Applied Microbiology and Biotechnology*, **85**, pp.667–677.

Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F., Boyer, H. W., 1977. Expression in *Escherichia coli* of a Chemically Synthesized Gene for the Hormone Somatostatin. *Science*, **198**, pp.1056–1063.

Ito, Y., Yamanishi, M., Ikeuchi, A., Imamura, C., Tokuhiro, K., 2013. Characterization of Five Terminator Regions that Increase the Protein Yield of a Transgene in *Saccharomyces cerevisiae*. *Journal of Biotechnology*, **168**, pp.486–492.

Jensen, D. F., Knudsen, I. M. B., Lübeck, M., Mamarabadi, M., Hockenhull, J. & Jensen, B., 2007. Development of a Biocontrol Agent for Plant Disease Control with Special Emphasis on the Near Commercial Fungal Antagonist *Clonostachys rosea* strain “IK726.” *Australasian Plant Pathology*, **36**, pp.95– 101.

Jeong, T., Son, Y., Ryu, H., Koo, B., Jeong, S., Hoang, P., Hang, B. *et a*l., 2014. Soluble Expression and Partial Purification of Recombinant Human Erythropoietin from *E. coli*. *Protein Expression and Purification*, **95**, pp.211–218.

Jiang, D., Zhu, W., Wang, Y., Sun, C., Zhang, K. & Yang, J., 2013. Molecular Tools for Functional Genomics in Filamentous Fungi : Recent Advances and New Strategies. *Biotechnology Advances*, 31, pp.1562–1574.

Johnson, I, 1983. Human Insulin from Recombinant DNA technology. *Science*, **219**, pp.632–637.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & Madden, T. L., 2008. NCBI BLAST: A Better Web Interface. *Nucleic Acids Research*, **36**, pp.5–9.

Jørgensen, M. S., Skovlund, D. A., Johannesen, P. F. & Mortensen, U. H., 2014. A Novel Platform for Heterologous Gene Expression in *Trichoderma reesei* (Teleomorph *Hypocrea jecorina*). *Microbial Cell Factories*, **13**.

Jun, H., Kieselbach, T. & Jönsson, L. J., 2011. Enzyme Production by Filamentous Fungi : Analysis of the Secretome of *Trichoderma reesei* Grown on Unconventional Carbon Source. *Microbial Cell Factories*, **10**.

Kapp, K., Schrempf, S., Lemberg, M. K. & Dobberstein, B., 2009. Post-Targeting Functions of Signal Peptides. In R. Zimmermann, ed. *Protein Transport into the Endoplasmic Reticulum.* Austin, USA: Landes Bioscience.

Kapust, R. B. & Waugh, D. S., 1999. *Escherichia coli* Maltose-Binding Protein is Uncommonly Effective at Promoting the Solubility of Polypeptides to Which It Is Fused. *Protein Science*, **8**, pp.1668–1674.

Katoh, S., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, **30**, pp.772–778.

Kaufman, R. J., 1990a. Selection and Co-Amplification of Heterologous Genes in Mammalian Cells. *Methods in Enzymology*, **185**, pp.537–566.

Kaufman, R. J., 1990b. Vectors Used for Expression in Mammalian Cells. *Methods in Enzymology*, **185**, pp.487–511.

Khokhar, I., Haider, M. S., Mushtaq, S. & Mukhtar, I., 2013. Isolation and Screening of Highly Cellulolytic Filamentous Fungi. *Journal of Applied Sciences and Environmental Management*, **16**.

Kikuchi, Y., Itaya, H., Date, M., Matsui, K. & Wu, L. F., 2008. Production of *Chryseobacterium proteolyticum* Protein-Glutaminase using the Twin-Arginine Translocation Pathway in *Corynebacterium glutamicum*. *Applied Microbiology and Biotechnology*, **78**, pp.67–74.

Knapp, C. & Eveleigh, D. E., 1973. Improved pH Control of Fungal Culture Media. *Mycologia*, **65**, pp.1078–1086.

Kosawang, C., Karlsson, M., Jensen, D. F., Dilokpimol, A. & Collinge, D. B., 2014. Transcriptomic Profiling to Identify Genes Involved in *Fusarium* mycotoxin deoxynivalenol and Zearalenone Tolerance in the Mycoparasitic Fungus *Clonostachys rosea*. *BMC Genomics*, **15**.

Krull, R., Wucherpfennig, T., Eslahpazir, M., Walisko, R., Melzer, G., Hempel, Dietmar C. *et al*., 2013. Characterization and Control of Fungal Morphology for Improved Production Performance in Biotechnology. *Journal of Biotechnology*, **163**, pp.112–123.

Krzeslak, J., Braun, P., Voulhoux, R., Cool, R. & Quax, W., 2009. Heterologous Production of *Eschericia coli* Penicillin G Acylase in *Pseudomanas aeruginosa. Journal of Biotechnology,* **142**, pp.250–258.

Kuipers, O. P., de Ruyter, P. G., Kleerebezem, M. & de Vos, W. M., 1997. Controlled Overproduction of Proteins by Lactic Acid Bacteria. *Trends in Biotechnology*, **15**, pp.135–140.

Ladds, G., Davis, K., Hillhouse, E. W. & Davey, J., 2003. Modified Yeast Cells to Investigate the Coupling of G-Protein-Coupled Receptors to Specific G-Proteins. *Molecular Microbiology*, **47**, pp.781–792.

Laemmli, U.K., 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4*. Nature*, **226**, pp.680–685.

Lagoumintzis, G., Zisimopoulou, P., Trakas, N., Grapsa, E., Poulas, K. & Tzartos, S. J., 2014. Scale Up and Safety Parameters of Antigen Specific Immunoadsorption of Human Anti-Acetylcholine Receptor Antibodies. *Journal of Neuroimmunology*, **267**, pp.1–6.

Lahlali, R. & Peng, G., 2014. Suppression of Clubroot by *Clonostachys rosea* via Antibiosis and Induced Host Resistance. *Plant Pathology*, **63**, pp.447–455.

Leoni, C., Volpicella, M., De Leo, F., Gallerani, R. & Ceci, L. R.., 2011. Genome Walking in Eukaryotes. *FEBS Journal*, **278**, pp.3953–3977.

Letcher, P. M. & Powell, M. J., 2014. Hypothesized Evolutionary Trends in Zoospore Ultrastructural Characters in Chytridiales (Chytridiomycota). *Mycologia*, **106**, pp.379–396.

Li, C., Yang, Z., Zhang R. H. C., Zhang, D., Chen, S. & Ma, L., 2013. Effect of pH on Cellulase Production and Morphology of *Trichoderma reesei* and The Application in Cellulosic Material Hydrolysis. *Journal of Biotechnology*, **168**, pp.470–477.

Li, J., Wang, J., Wang, S., Xing, M., Yu, S. & Liu, G., 2012. Achieving Efficient Protein Expression in *Trichoderma reesei* by Using Strong Constitutive Promoters. *Microbial Cell Factories*, **11**.

Li, J., Yang, J. K., Huang, X. W. & Zhang, K. Q., 2006. Purification and Characterization of an Extracellular Serine Protease from *Clonostachys rosea* and Its Potential as a Pathogenic Factor. *Process Biochemistry*, **41**, pp.925–929.

Liang, L., Yang, J., Li, J., Mo, Y., Li, L., Zhao, X. & Zhang, K., 2011. Cloning and Homology Modeling of a Serine Protease Gene (PrC) from the Nematophagous Fungus *Clonostachys rosea*. *Annals of Microbiology*, **61**, pp.511–516.

Lim, D., Hains, P., Walsh, B., Bergquist, P. & Nevalainen, H., 2001. Proteins Associated with the Cell Envelope of *Trichoderma reesei*: A Proteomic Approach. *Proteomics*, **1**, pp.899–910.

Linn, S. & Arber, W., 1968. Host Specificity of DNA Produced by *Escherichia coli*, X. *In vitro* Restriction of phage fd Replicative Form. *Proceedings of the National Academy of Sciences of the USA*, **59**, pp.1300–1306.

Loir, Y. L., Azevedo, V., Oliveira, S. C., Freitas, D. A., Miyoshi, A., Bermúdez-Humarán, L. G. & Nouaille, S. *et al*., 2005. Protein Secretion in *Lactococcus lactis*: An Efficient Way to Increase the Overall Heterologous Protein Production. *Microbial Cell Factories*, **4**.

Lubertozzi, D. & Keasling, J. D., 2009. Developing *Aspergillus* as a Host for Heterologous Expression. *Biotechnology Advances*, **27**, pp.53–75.

Lv, D., Wang, W. & Wei, D., 2012. Plasmid Construction of Two Vectors for Gene Expression in *Trichoderma reesei*. *Plasmid*, **67**.

Ma, L., Zhang, J., Zou, G., Wang, C. & Zhou, Z., 2011. Improvement of Cellulase Activity in *Trichoderma reesei* by Heterologous Expression of a Beta-glucosidase Gene from *Penicillium decumbens*. *Enzyme and Microbial Technology*, **49**, pp.366–371.

Machida, M., 2008. Genome Sequence of *Aspergillus oryzae*. In S. Osmani & G. Goldman, eds. *Aspergilli*: *Genomics, Medical Aspects, Biotechnology, and Research Methods*. Florida, USA: CRC Press, pp.75–84.

MacRae, W. D., Buxton, F. P., Gwynne, D. I. & Davies, R. W., 1993. Heterologous Protein Secretion Directed by a Repressible Acid Phosphatase System of *Aspergillus niger*. *Gene*, **132**, pp.193–198.

Madhavan, A. & Sukumaran, R. K., 2014. Promoter and Signal Sequence from Filamentous Fungus Can Drive Recombinant Protein Production in the Yeast *Kluyveromyces lactis*. *Bioresource Technology*, **165**, pp.302–308.

Madhavan, A. & Sukumaran, R. K., 2015. Signal Peptides from Filamentous Fungi Efficiently Mediate the Secretion of Recombinant Proteins in *Kluyveromyces lactis*. *Biochemical Engineering Journal*, **102,** pp31–37.

Mahmoud, K., 2007. Recombinant Protein Production: Strategic Technology and a Vital Research Tool. *Research Journal of Cell and Molecular Biology*, **1**, pp.9–22.

Margolles-Clark, E., Harman, G. E. & Penttilä, M., 1996. Enhanced Expression of Endochitinase in *Trichoderma harzianum* with the cbh1 Promoter of *Trichoderma reesei. Applied Environmental Microbiology,* **62**, pp.2152–2155.

Marisch, K., Bayer, K., Cserjan-Puschmann, M., Luchner, M. & Striedner, G., 2013. Evaluation of Three Industrial *Escherichia coli* strains in Fed-Batch Cultivations during High-Level SOD Protein Production. *Microbial Cell Factories*, **12**.

Martinez, J. L., Liu, L., Petranovic, D. & Nielsen, J., 2012. Pharmaceutical Protein Production by Yeast: Towards Production of Human Blood Proteins by Microbial Fermentation. *Current Opinion in Biotechnology*, **23**, pp.965–971.

Matsuzawa, T., Tohda, H. & Takegawa, K., 2013. Ethanol-Inducible Gene Expression using gld1 (+) Promoter in the Fission Yeast *Schizosaccharomyces pombe*. *Applied Microbiology and Biotechnology*, **97**, pp.6835–6843.

Mattanovich, D. Branduardi, P., Dato, L., Gasser, B., Sauer, M. & Porro, D., 2012. Recombinant Protein Production in Yeasts. *Methods in Molecular Biology*, **824**, pp.329–358.

McCormack, A., Schieltz, D., Goode, B., Yang, S., Barnes, G., Drubin, D. & Yates, J. R., 1997. Direct Analysis and Identification of Proteins in Mixtures by LC/MS/MS and Database Searching at the Low-Femtomole Level. *Analytical Chemistry*, **69**, pp.767–776.

McQuilken, M. P., Gemmel, J. & Lahdenpera, M., 2001. *Gliocladium catenulatum* as a Potential Biological Control Agent of Damping-Off in Bedding Plants. *Journal of Phytopathology*, **149**, pp.171–178.

Merten, O. W., 2006. Introduction to Animal Cell Culture Technology — Past, Present and Future. *Cytotechnology*, **50**, pp.1–7.

Meyer, H. & Schmidhalter, D. R., 2012. Microbial Expression Systems and Manufacturing from a Market and Economic Perspective. In *Innovations in Biotechnology*. Rijeka, Croatia: InTech, pp.211–250.

Meyer, V., Wanka, F., van Gent, J., Arentshorst, M., van de Hondel, C. A. M. J. J. & Ram, A. F. J., 2011. Fungal Gene Expression on Demand: An Inducible, Tunable, and Metabolism-Independent Expression System for *Aspergillus niger. Applied and Environmental Microbiology*, **77**, pp.2975–2983.

Meyer, V., Wu, B. & Ram, A. F. J., 2011. *Aspergillus* as a Multi-Purpose Cell Factory: Current Status and Perspectives. *Biotechnology Letters*, **33**, pp.469–476.

Mikosch, T., Klemm, P., Gassen, H. G., van den Hondel, C. A. M. J. J. & Kemme, M., 1996. Secretion of Active Human Mucus Proteinase Inhibitor by *Aspergillus niger* After KEX2-like Processing of a Glucoamylase-Inhibitor Fusion Protein. *Journal of Biotechnology,* **52**, pp.97–106.

Møller, K., Jensen, B., Andersen, H. P., Stryhn, H. & Hockenhull, J., 2003. Biocontrol of *Pythium tracheiphilum* in Chinese cabbage by *Clonostachys rosea* under Field Conditions. *Biocontrol Science and Technology*, **13,** pp.171–182.

Moore, M. M., 2007. Genetic Engineering of Fungal Cells. *Biotechnology,* **3**, pp.36–63.

Mustalahti, E., Saloheimo, M. & Joensuu, J. J., 2011. Intracellular Protein Production in *Trichoderma reesei* (*Hypocrea jecorina*) with Hydrophobin Fusion Technology. *Nature Biotechnology,* **30**, pp.262–268.

Needham, P., Mikoluk, K., Dhakarwal, P., Khadem, S., Snyder, A. C., Subramanya, A. R. & Brodsky, J. L., 2011. The Thiazide-Sensitive NaCl Co-Transporter is Targeted For Chaperone-Dependent Endoplasmic Reticulum-Associated Degradation. *Journal of Biological Chemistry*, **286**, pp.43611–43621.

Nevalainen, H. & Peterson, R., 2014. Making Recombinant Proteins in Filamentous Fungi – Are We Expecting Too Much? *Frontiers in Microbiology*, **5**.

Nevalainen, H., Te'o, V. & Penttilä, M., 2003. Application of Genetic Engineering for Strain Improvement in Filamentous Fungi. In K. A. Dilip, ed. *Handbook of Fungal Biotechnology*. CRC Press, pp.193–208.

Nevalainen, H., Te'o, V. S. J. & Bergquist, P. L., 2005. Heterologous Protein Expression in Filamentous Fungi. *Trends in Biotechnology*, **23**, pp.468–74.

Nguyen, Q. D., Bujna, E., Styekó, G., Rezessy-Szabó, J. M. & Hoschke, Á., 2015. Fungal Biomolecules for the Food Industry. In V. K. Gupta, R. L. Mach, & S. Sreenivasaprasad, eds. *Fungal Biomolecules: Sources, Applications and Recent Developments*. West Sussex, UK: John Wiley & Sons, Inc., pp.11–38.

Ochman, H., Gerber, A. S., Hart, D. L. & Hartl, D. L., 1988. Genetic Applications of an Inverse Polymerase Chain Reaction. *Genetics*, **120**, pp.621–3.

Olmedo-Monfil, V., Cortés-Penagos, C. & Herrera-Estrella, A., 2004. Three Decades of Fungal Transformation. In P. Balbás & A. Lorence, eds. *Recombinant Gene Expression, Methods in Molecular Biology*. Totowa, USA: Humana Press Inc., pp.297–313.

Palomares, L. A., Estrada-Mondaca, S. & Ramírez, O. T., 2004. Production of Recombinant Proteins: Challenges and Solutions. In P. Balbás & A. Lorence, eds. *Methods in Molecular Biology*. Totowa, USA: Humana Press Inc., pp.15–52.

Papagianni, M., 2004. Fungal Morphology and Metabolite Production in Submerged Mycelial Processes. *Biotechnology Advances*, **22**, pp.189–259.

Parekh, S., Vinci, V. A. & Strobel, R.J., 2000. Improvement of Microbial Strains and Fermentation Processes. *Applied Microbiology and Biotechnology*, **54**, pp.287–301.

Partow, S., Siewers, V., Bjorn, S., Nielsen, J. & Maury, J., 2010. Characterization of Different Promoters for Designing a New Expression Vector in *Saccharomyces cerevisiae. Yeast,* **27**, pp.955–964.

Peciulyte, A., Anasontzis, G. E., Karlström, K., Larsson, P. T. & Olsson, L., 2014. Morphology and Enzyme Production of *Trichoderma reesei* Rut C-30 are Affected by the Physical and Structural Characteristics of Cellulosic Substrates. *Fungal Genetics and Biology*, **72**, pp.64–72.

Peñalva, M. A., Rowlands, R. T. & Turner, G., 1998. The Optimization of *Penicillin* Biosynthesis in Fungi. *Trends in Biotechnology*, **16**, pp.483–489.

Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H., 2011. SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nature Methods*, **8**, pp.785–786.

Peterson, R. & Nevalainen, H., 2012. *Trichoderma reesei* RUT-C30 – Thirty Years of Strain Improvement. *Microbiology*, **158**, pp.58–68.

Pointing, S. B., 1999. Qualitative Methods for the Determination of Lignocellulolytic Enzyme Production by Tropical Fungi. *Fungal Diversity*, **2**, pp.17–33.

Porcel, R. E. M., López, C. J. L., Pérez, S. J. A., Sevilla, F. J. M. & Chisti, Y., 2005. Effects of Pellet Morphology on Broth Rheology in Fermentations of *Aspergillus terreus*. *Biochemical Engineering Journal*, **26**, pp.139–144.

Porro, D., Gasser, B., Fossati, T., Maurer, M., Branduardi, P., Sauer, M. & Mattanovich, D., 2011. Production of Recombinant Proteins and Metabolites In Yeasts: When Are These Systems Better Than Bacterial Production Systems? *Applied Biochemistry and Biotechnology*, **89**, pp.939–948.

Posch, A. E., Herwig, C. & Spadiut, O., 2013. Science-Based Bioprocess Design for Filamentous Fungi. *Trends in Biotechnology*, **31**, pp.37–44.

Punt, P. J., van Biezen, N., Conesa, A., Albers, A., Mangnus, J. & van den Hondel, C. A. M. J. J., 2002. Filamentous Fungi as Cell Factories for Heterologous Protein Production. *Trends in Biotechnology*, **20**, pp.200–206.

Punt, P.J., Dingemanse, M. A., Kuyvenhoven, A., Soede, R. D., Pouwels, P. H. & van den Hondel, C. A. M. J. J., 1990. Functional Elements in the Promoter Region of the *Aspergillus nidulans* gpdA Gene Encoding glyceraldehyde-3-phosphate dehydrogenase. *Gene*, **93**, pp.101–109.

Punt, P. J., Zegers, N. D., Busscher, M., Pouwels, P. H. & van den Hondel, C. A. M. J. J., 1991. Intracellular and Extracellular Production of Proteins in *Aspergillus* under the Control of Expression Signals of the Highly Expressed *Aspergillus nidulans* gpdA Gene. *Journal of Biotechnology*, **17**, pp.19–33.

Punt, P. J., Veldhuisen, G. & van de Hondel, C. A. M. J. J., 1994. Protein Targeting and Secretion in Filamentous Fungi. *Antonie van Leeuwenhoek*, **65**, pp.211–216.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., *et al*., 2012. The Pfam Protein Families Database. *Nucleic Acids Research*, **40**, pp.290–301.

Qin, L. N., Cai, F. R., Dong, X. R., Huang, Z. B., Tao, Y., Huang, J. Z. & Dong, Z. Y., 2012. Improved Production of Heterologous Lipase in *Trichoderma reesei* by RNAi-Mediated Gene Silencing of an Endogenic Highly Expressed Gene. *Bioresource Technology*, **109**, pp.116–122.

Radzio, R. & Kück, U., 1997. Synthesis of Biotechnologically Relevant Heterologous Proteins in Filamentous Fungi. *Process Biochemistry*, **32**, pp.529–539.

Rai, M. & Padh, H., 2001. Expression Systems for Production of Heterologous Proteins. *Current Science*, **80**, pp.1121–1128.

Rakestraw, J. A., Sazinsky, S. L., Piatesi, A., Antipov, E. & Wittrup, K. D., 2009. Directed Evolution of a Secretory Leader for the Improved Expression of Heterologous Proteins and Full Length Antibodies in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, **103**, pp.1192–1201.

Ramzan, M., Asgher, M., Sheikh, M. A. & Bhatti, H. N., 2013. Hyperproduction of manganese peroxidase through chemical mutagenesis of *Trametes versicolor* IBL-04 and optimization of process parameters. *Bioresources*, **8**, pp.3953–3966.

Rani, M. & Prasad, N., 2012. Strain Improvement of *Aspergillus flavus* for Enhanced Production of Alkaline Protease Enzyme. *International Journal of Science Innovations and Discoveries*, **2**, pp.113–120.

Rao, D. N., Dryden, D. T. F. & Bheemanaik, S., 2014. Type III Restriction-Modification Enzymes: A Historical Perspective. *Nucleic Acids Research*, **42**, pp.45–55.

Rao, M. B., Tanksale, A. M. & Ghatge, M.S., 1998. Molecular and Biotechnological Aspects of Microbial Proteases. *Microbiology and Molecular Biology Reviews*, **62**, pp.597–635.

Rawlings, N. D. & Barrett, A. J., 1993. Evolutionary Families of Peptidases. *Biochemical Journal*, **218**, pp.205–218.

Rayner, R. W., 1970. A Mycological Colour Chart, Kew, Surrey, UK: Commonwealth Mycological Institute.

Retallack, D. M., Jin, H. & Chew, L., 2011. Reliable Protein Production in a *Pseudomonas fluorescens* Expression System. *Protein Expression and Purification*, **81**, pp.157–165.

Richey, M. G., Marek, E. T., Schardl, C. L. & Smith, D. A., 1989. Transformation of Filamentous Fungi with Plasmid DNA by Electroporation. *Phytopathology*, **79**, pp.844–847.

Rivera, A. L., Magaña-Ortíz, D., Gómez-Lim, M., Fernández, F., Loske, A. M. & Leonor, A., 2014. Physical Methods for Genetic Transformation of Fungi and Yeast. *Physics of Life Reviews*, **11**, pp.184–203.

Roberti, R., Veronesi, A. R., Cesari, A., Cascone, A., Di Berardina, I., Bertini, L. & Caruso, C., 2008. Induction of PR Proteins and Resistance by the Biocontrol Agent *Clonostachys rosea* in Wheat Plants Infected with *Fusarium culmorum*. *Plant Science*, **175**, pp.339–347.

Rodrigues, A., Mueller, U. G., Ishak, H. D., Bacci, M. & Pagnocca, F. C., 2011. Ecology of Microfungal Communities in Gardens of Fungus-Growing Ants (Hymenoptera: Formicidae): A Year-Long Survey of Three Species of Attine Ants in Central Texas. *FEMS Microbiology Ecology*, **78**, pp.244–55.

Rosano, G. L. & Ceccarelli, E. A., 2014. Recombinant Protein Expression in Microbial Systems. *Frontiers in Microbiology*, **5**.

Rose, S. & Van Zyl, W., 2002. Constitutive Expression of the *Trichoderma reesei* β-1,4-xylanase Gene (xyn2) and The β-1,4-endoglucanase Gene (egI) in *Aspergillus niger* in Molasses and Defined Glucose Media. *Applied Microbiology and Biotechnology*, **58**, pp.461–468.

Rowlands, R. T. T., 1984. Industrial Strain Improvement: Rational Screens and Genetic Recombination Techniques. *Enzyme and Microbial Technology*, **6**, pp.290–300.

Ruiz-Díez, B., 2002. Strategies for the Transformation of Filamentous Fungi. *Applied Microbiology*, **92**, pp.189–195.

Sagmeister, P., Schimek, C., Meitz, A., Herwig, C. & Spadiut, O., 2014. Tunable Recombinant Protein Expression with *E. coli* in a Mixed-Feed Environment. *Applied Microbiology and Biotechnology*, **98**, pp.2937–2945.

Sahdev, S., Khattar, S. & Saini, K., 2008. Production of Active Eukaryotic Proteins through Bacterial Expression Systems: A Review of the Existing Biotechnology Strategies. *Molecular and Cellular Biochemistry*, **307**, pp.249–264.

Saloheimo, M. & Pakula, T. M., 2012. The Cargo and the Transport System: Secreted Proteins and Protein Secretion in *Trichoderma reesei* (*Hypocrea jecorina*). *Microbiology*, **158**, pp.46–57.

Salovouri, I., Makarow, M., Rauvala, H., Knowles, J. & Kääriäinen, L. 1987. Low Molecular Weight High-Mannose Type Glycans in a Secreted Protein of the Filamentous Fungus *Trichoderma reesei*. *Nature Biotechnology*, **5**, pp.152–156.

Sambrook, J., Fritsch, E. F. & Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd ed., New York, USA: Cold Spring Harbor Laboratory Press.

Saritha, M., Arora, A. & Nain, L., 2012. Pretreatment of Paddy Straw with *Trametes hirsuta* for Improved Enzymatic Saccharification. *Bioresource Technology*, **104**, pp.459–465.

Sarrouh, B., Santos, T. M., Miyoshi, A., Dias, R. & Azevedo, V., 2012. Up-To-Date Insight on Industrial Enzymes Applications and Global Market. *Journal of Bioprocessing & Biotechniques*, **S4**.

Saunders, G., Picknett, T. M., Tuite, M. F. & Ward, M., 1989. Heterologous Gene Expression in Filamentous Fungi. *Trends in Biotechnology*, **7**, pp.283–287.

Schoberle, T. J., Nguyen-Coleman, C. K. & May, G. S., 2013. Plasmids for Increased Efficiency of Vector Construction and Genetic Engineering in Filamentous Fungi. *Fungal Genetics and Biology*, **58**, pp.1–9.

Schroers, H., Samuels, Gary J., Seifert, K. A. & Gams, W., 1999. Classification of the Mycoparasite *Gliocladium roseum* in *Clonostachys* as *C. rosea*, Its Relationship to *Bionectria ochroleuca,* and Notes on Other *Gliocladium*-like Fungi. *Mycologia*, **91**, pp.365–385.

Schroers, H., 2001. A Monograph of *Bionectria* (Ascomycota, Hypocreales, Bionectriaceae) and Its *Clonostachys* Anamorphs. *Studies in Mycology*, **46**, pp.1–214.

Schumann W. & Fereira L. C., 2004. Secretory Production of Recombinant Proteins in *Escherichia coli*. *Genetics and Molecular Biology*, **27**, pp.442–453.

Schuster, A. & Schmoll, M., 2010. Biology and Biotechnology of Trichoderma. *Applied Microbiology and Biotechnology*, **87**, pp.787–799.

Seiboth, B., Ivanova, C. & Seidl-Seiboth, V., 1996. *Trichoderma reesei*: A Fungal Enzyme Producer for Cellulosic Biofuels. INTECH Open Access Publisher.

Semba, H., Ichige, E., Imanaka, T., Atomi, H. & Aoyagi, H., 2010. Efficient Production of Active Form Recombinant Cassava Hydroxynitrile Lyase using *Escherichia coli* in Low-Temperature Culture. *Methods in Molecular Biolology*, **643**, pp.133–144.

Sette, L. D., Pagnocca, F. C. & Rodrigues, A., 2013. Microbial Culture Collections as Pillars for Promoting Fungal Diversity, Conservation and Exploitation. *Fungal Genetics and Biology*, **60**, pp.2–8.

Sharma, R., Katoch, M., Srivastava, P. S. & Qazi, G. N., 2009. Approaches for Refining Heterologous Protein Production in Filamentous Fungi. *World Journal of Microbiology and Biotechnology*, **25**, pp.2083–2094.

Shoji, J. Y., Kikuma, T. & Kitamoto, K., 2014. Vesicle Trafficking, Organelle Functions, and Unconventional Secretion in Fungal Physiology and Pathogenicity. *Current Opinion in Microbiology*, **20**, pp.1–9.

Sibthorp, C., Wu, H., Cowley, G., Wong, P. W. H., Palaima, P., Morozov, I. Y., Weedall, G. D. & Caddick, M. X., 2013. Transcriptome Analysis of the Filamentous Fungus *Aspergillus nidulans* Directed to the Global Identification of Promoters. *BMC Genomics*, **14**.

Siezen, R. J. & Leunissen, J. A. M., 1997. Subtilases: The Superfamily of Subtilisin like Serine Protease. *Protein Science*, **6**, pp.501–523.

Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L. & Xenarios, I., 2012. New and Continuing Developments at PROSITE. *Nucleic Acids Research*, **41**, pp.1–4.

Sinha, S. & Tompa, M., 2003. YMF: a Program for Discovery of Novel Transcription Factor Binding Sites by Statistical Overrepresentation. *Nucleic Acids Research*, **31**, pp.3586–3588.

Smith, D., 2012. Cap 4 Culture Collections. *Advances in Applied Microbiology*, **7**, pp.73–118.

Sørensen, H. P., 2010. Towards Universal Systems for Recombinant Gene Expression. *Microbial Cell Factories*, **9**.

Specht, E. A. & Mayfield, S. P., 2014. Algae-Based Oral Recombinant Vaccines. *Frontiers in Microbiology*, **5**.

Steen, H. & Mann, M., 2004. The ABC's (and XYZ's) of Peptide Sequencing. *Nature Reviews Molecular Cell Biology*, **5**, pp.699–711.

Strauss, J., Zeilinger, S., Hartler, G., Stoffler, G., Wolschek, M. & Kubicek, C. P., 1995. Cre1, The Carbon Catabolite Repressor Protein from *Trichoderma reesei*. *FEBS Letters*, **376**, pp.103–107.

Su, X., Schmitz, G., Zhang, M., Roderick, I. M. & Cann, I. K. O., 2011. Heterologous Protein Expression in Filamentous Fungi. *Advances in Applied Microbiology*, **81**, pp.1–161.

Sutton, J. C., Li, D., Peng, G., Yu, H., Zhang, P. & Valdebenito-Sanhueza, R. M., 1997. *Gliocladium roseum* A Versatile Adversary of *Botrytis cinerea* In Crops. *Plant Disease*, **81**, pp.316–328.

Tan, G., Gao, Y., Shi, M., Zhang, X., He, S. & Chen, Z., 2005. SiteFinding-PCR : A Simple and Efficient PCR Method for Chromosome Walking. *Nucleic Acid Research*, **33**, pp.1–7.

Teather, R. M. M. & Wood, P. J. J., 1982. Use of Congo Red-polysaccharide Interactions in Enumeration and Characterization of Cellulolytic Bacteria from the Bovine Rument. *Applied and Environmental Microbiology*, **43**, pp.777–780.

Terpe, K., 2006. Overview of Bacterial Expression Systems for Heterologous Protein Production: From Molecular and Biochemical Fundamentals to Commercial Systems. *Applied Microbiology and Biotechnology*, **72**, pp.211–223.

Thim, L., Hansen, M. T., Norris, K., Hoegh, I., Boel, E., Forstrom, J., Ammerer, G. & Fiil, N. P., 1986. Secretion and Processing of Insulin Precursors in Yeast. *Proceedings of the National Academy of Sciences of the USA*, **83**, pp.6766–6770.

Tilburn, J., Scazzocchio, C., Taylor, G., Zabicky-Zissman, J. H., Lockington, R. A. & Davies, R. W., 1983. Transformation by Integration in *Aspergillus nidulans*. *Gene*, **26**, pp.205–221.

Toledo, A. V., Virla, E., Humber, R. A., Paradell, S. L. & Lastra, C. C. L., 2011. First Record of *Clonostachys rosea* (Ascomycota : Hypocreales) as an Entomopathogenic Fungus of *Oncometopia tucumana* and *Sonesimia grossa* (*Hemiptera : Cicadellidae*) in Argentina. *Journal of Invertebrate Pathology*, **92**, pp.7–10.

Tonooka, Y. & Fujishima, M., 2009. Comparison and Critical Evaluation of PCR-Mediated Methods to Walk Along the Sequence of Genomic DNA. *Applied Microbiology and Biotechnology*, **85**, pp.37–43.

Unkles, S. E., 1992. Gene Organization in Industrial Filamentous Fungi. In J. R. Kinghorn & G. Turner, eds. *Applied Molecular Genetics of Filamentous Fungi*. Glasgow: Blackie Academic and Professional, pp.28–53.

Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y. Y., Reading, C. A. & Chhiba, K. 2012. An Engineered Eukaryotic Protein Glycosylation Pathway in *Escherichia coli. Nature Chemical Biology*, **8**, pp.434–436.

Valencia, E. Y. & Chambergo, M. F. S., 2013a. Fungal Enzymes. In M. M. Polizeli & R. Mahendra, eds. *Genetic Tools for Production of Proteins in Fungi.* Boca Raton, FL, USA: CRC Press, p. 376.

Valencia, E. Y. & Chambergo, M. F. S., 2013b. Genetic Tools for Production of Proteins in Fungi. In M. Polizeli & M. Rai, eds. *Fungal Enzymes*. Boca Raton, FL, USA: CRC Press, p.376.

Vanhanen, S., Penttifä, M., Lehtovaara, P. & Knowles, J., 1989. Isolation and Characterization of the 3-Phosphoglycerate Kinase Gene (*pgk*) from the Filamentous Fungus *Trichoderma reesei*. *Current Genetics*, **15**, pp.181–186.

Vecht-Lifshitz, S., Magdassi, S. & Braun, S., 1990. Pellet Formation and Cellular Aggregation in *Streptomyces tendae*. *Biotechnology & Bioengeneering*, **35**, pp.890–896.

Visser, H., Joosten, V., Punt, P. J., Gusakov, A. V., Olson, P. T., Joosten, R., Bartels, J. *et al*., 2011. Development of a Mature Fungal Technology and Production Platform for Industrial Enzymes Based on a *Myceliophthora thermophila* Isolate, Previously Known as *Chrysosporium lucknowense* C1. *Industrial Biotechnology*, **7**, pp.214–224.

Voulgaridou, G. Mantso, T., Chlichlia, K. & Panayiotidis, M. I., 2013. Efficient *E. coli* Expression Strategies for Production of Soluble Human Crystallin ALDH3A1. *PloS one*, **8**, e56582.

Walsh, G., 2005. Therapeutic Insulins and Their Large-Scale Manufacture. *Applied Biochemisty and Biotechnology*, **67**, pp.151–159.

Wang, L., Ridgway, D., Gu, T. & Moo-Young, M., 2005. Bioprocessing Strategies to Improve Heterologous Protein Production in Filamentous Fungal Fermentations. *Biotechnology Advances*, **23**, pp.115–129.

Ward, O. P., 2012. Production of Recombinant Proteins by Filamentous Fungi. *Biotechnology Advances*, **30**, pp.1119–1139.

White, T. J., Bruns, T. D., Lee, S. & Taylor, J. W., 1990. Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. In Innis, M. A., Gelfand, D. H., Sninsky,

J. J. & White, T. J, eds. *PCR Protocols: a Guide to Methods and Applications*. New York, USA: Academic Press, pp.315–322.

Wilson, D. B., 2009. Cellulases and Biofuels. *Current Opinion in Biotechnology*, **20**, pp.295–299.

Wood, K. V., 1995. Marker Proteins for Gene Expression. *Current Opinion in Biotechnology*, **6**, pp.50–58.

Wucherpfennig, T., Kiep, K. A., Driouch, H., Wittmann, C. & Krull, R., 2010. Morphology and Rheology in Filamentous Cultivations. *Advances in Applied Microbiology*, **72**, pp.89–136.

Yamanishi, M., Ito, Y., Kintaka, R., Imamura, C., Katahira, S., Ikeuchi, A., Moriya, H. & Matsuyama, T., 2013. A Genome-Wide Activity Assessment of Terminator Regions in *Saccharomyces cerevisiae* provides a "terminatome" Toolbox. *ACS Synthetic Biology*, **2**, pp.337–347.

Yang, J., Tian, B., Liang, L. & Zhang, K., 2007. Extracellular Enzymes and the Pathogenesis of Nematophagous Fungi. *Applied Microbiology and Biotechnology*, **75**, pp.21–31.

Yin, J., Li, G., Ren, X. & Herrler, G., 2007. Select What You Need: A Comparative Evaluation of the Advantages and Limitations of Frequently Used Expression Systems for Foreign Genes. *Journal of Biotechnology*, **127**, pp.335–347.

Yu, M., Wen, S. & Tan, T., 2010. Enhancing Production of *Yarrowia lipolytica* lipase Lip2 in *Pichia pastoris*. *Engineering in Life Sciences*, **10**, pp.458–464.

Yuan, Z. L., Rao, L. B., Chen, Y. C., Zhang, C. L. & Wu, Y. G., 2011. From Pattern to Process: Species and Functional Diversity in Fungal Endophytes of *Abies beshanzuensis*. *Fungal Biology*, **115**, pp.197–213.

Zanphorlin, L. M., Cabral, H., Arantes, E., Assis, D., Juliano, L. & Juliano, M. A., 2011. Purification and Characterization of a New Alkaline Serine Protease from the Thermophilic Fungus *Myceliophthora sp*. *Process Biochemistry*, **46**, pp.2137–2143.

Zhong, Y., Liu, X., Xiao, P., Wei, S. & Wang, T., 2011. Expression and Secretion of the Human Erythropoietin using an Optimized cbh1 Promoter and the Native CBH I Signal Sequence in the Industrial Fungus *Trichoderma reesei*. *Applied Biochemistry and Biotechnology*, **165**, pp.1169–1177.

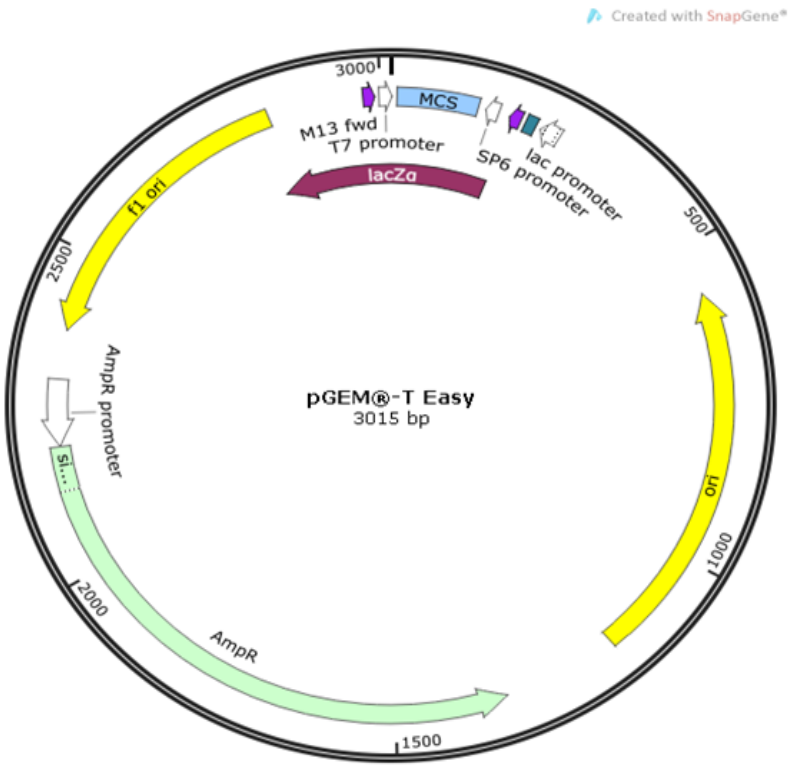
Zhu, J., 2012. Mammalian Cell Protein Expression for Biopharmaceutical Production. *Biotechnology Advances*, **30**, pp.1158–1170.

Žnidaršič, P. & Pavko, A., 2001. The Morphology of Filamentous Fungi in Submerged Cultivations as a Bioprocess Parameter. *Food Technology and Biotechnology*, **39**, pp.237–252.

Zorzano, A., Sebastián, D., Sanchez-Wandelmer, J., Miret, L. & Albericio, F., 2012. Gene Promoters and Transcription Control Regions as Therapeutic Targets. In L. M. Botana & M. Loza, eds. *Therapeutic Targets: Modulation, Inhibition, and Activation*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp.327–350.

APPENDICES

APPENDIX A: The pGEM® -T Easy vector map



APPENDIX B: Table of isolates used in the study

Taxon	Culture number	Host/Source	Collector
Alternaria solani	CMW 5971	*Solanum tubersum* (Potato)	Smith J
Alternaria solani	CMW 9515	*Solanum tubersum* (Potato)	Van der Waals J
Ambrosiella sp.	CMW 374	*Occotea bullata*	Wingfield MJ
Ambrosiella sp.	CMW 375	*Occotea bullata*	Wingfield MJ
Ambrosiella sp.	CMW 29834	*Tabernaemontana ventricosa*	Roux L
Ambrosiella sp.	CMW 29835	*Tabernaemontana ventricosa*	Roux L
Anthostomella clypeata	CMW 22204	*Restio egregious*	Lee S
Anthostomella clypeata	CMW 22250	*Restionaceae sp.*	Lee S
Anthostomella conorum	CMW 20398	*Protea magnifica*	Lee S
Anthostomella conorum	CMW 20403	*Protea nerifolia*	Lee S
Armillaria fuscipes	CMW 3950	*Litchi chenensis*	Coetzee MPA
Armillaria mellea	CMW 3788	Basidiocarp	Coetzee MPA
Armillaria mellea	CMW 3975	*Quercus spp.*	Wingfield MJ/Coetzee MPA
Armillaria mellea	CMW 3787	*Hydrangium sp.*	Coetzee MPA
Armillaria mellea	CMW 3340	Josia tree	Coetzee MPA
Armillaria mellea	CMW 3328	Parte House Chestnut	Coetzee MPA
Armillaria mellea	CMW 3341	*Quercus sp.*	Coetzee MPA
Armillaria sp.	CMW 11168	*Pinus taeda*	Roux J

Taxon	Culture number	Host/Source	Collector
Arthrographis cuboidea	CMW 5621	Red/pink stained pine floor plank	De Beer W
Beauvaria bassiana	CMW 393	*Hylurgus ligniperda*	Tribe G
Beauvaria bassiana	CMW 394	*Orthotomicus erosus*	Strauss L
Botryosphaeria dothidea	CMW 775	*Pinus halepensis*	Burns M
Botryosphaeria dothidea	CMW 892	*Cupressis horizontalis*	Unknown
Botryosphaeria dothidea	CMW 1177	*Protea grandiceps*	Wingfield MJ
Botryosphaeria dothidea	CMW 1178	*Protea grandiceps*	Wingfield MJ
Botryosphaeria sp.	CMW 207	*Mangifera indica* (Mango)	Jacobs R
Botryosphaeria sp.	CMW 242	*Malus pumila* (golden delicious/ royal gala apples)	De Smit A
Camarospoirum brabeji	CMW 22165	*Protea sp.*	Lee S
Candida sp.	CMW 19090	Pterocarpus angolensis	Mehl J
Candida sp.	CMW 19091	*Pterocarpus angolensis*	Mehl J
Canninghamella Echinulata	CMW 1893	Dung	Unknown
Capronia kleinmondensis	CMW 22136	*Protea cynaroides*	Lee S
Capronia leucadendri	CMW 19995	*Leucadendron sp.*	Lee S
Celoporthe dispersa	CMW 29381	*Syzygium legatti*	Vermeulen M / Roux J
Celoporthe dispersa	CMW 29898	*Syzygium legatti*	Vermeulen M / Roux J
Ceratocystiopsis sp.	CMW 27016	*Terminalia catappa*	Begoude D / Roux J
Ceratocystiopsis sp.	CMW 27183	*Eucalyptus saligna* sumps	Kamgan NG
Ceratocystis adiposa	CMW 2573	Unknown	Unknown
Ceratocystis albifundus	CMW 4083	*Acacia mearnsii*	Roux L
Ceratocystis albifundus	CMW 4068	*Acacia mearnsii*	Roux L
Ceratocystis albifundus	CMW13980	Unknown	Unknown
Ceratocystis albifundus	CMW 17620	*Terminalia serecia*	Roux L
Ceratocystis albifundus	CMW 24685	Unknown	Unknown
Ceratocystis albifundus	CMW 24860	Unknown	Unknown
Ceratocystis albifundus	CMW 8217	Nitulid beetle	Heath RN
Ceratocystis fagacearum	CMW 2656	Unknown	Unknown
Ceratocystis moniliformis	CMW 10134	*Eucalyptus grandis*	Van Wyk M
Ceratocystis paradoxa	CMW 1546	Unknown	Unknown
Ceratocystis populicola	CMW 14789	Unknown	Unknown
Ceratocystis radicicola	CMW 1032	Unknown	Unknown
Ceratocystis virescens	CMW 1339	Unknown	Unknown
Cercospora piaropi	CMW 2043	*Eichornia sp.*	Morris MJ
Cercospora zeina	CMW 25441	*Zea Mays*	Ros B
Cercospora zeina	CMW 25443	*Zea Mays*	Ros B
Ceuthospora sp.	CMW 18003	*Podocarpus elongatus*	Wood A
Ceuthospora sp.	CMW 18035	*Elegia equisetacea*	Lee S
Ceuthospora sp.	CMW 18367	*Podocarpus henckellii*	Lee S

Taxon	Culture number	Host/Source	Collector
cf. Diaporthe sp.	CMW 6255	*Musa sp*. (Williams banana)	Viljoen A
cf. Drechslera sp.	CMW 6192	Unknown	Unknown
cf. Selenophoma asterina	CMW 6259	*Musa sp.* (Williams banana)	Viljoen A
cf. Selenophoma asterina	CMW 6260	*Musa sp.* (Williams banana)	Viljoen A
cf. Selenophoma asterina	CMW 6261	*Musa sp*. (Williams banana)	Viljoen A
Chaetochalara hughesii	CMW 18307	*Elegia capensis*	Lee S
Chaetomium erectum	CMW 222	*Pinus taeda*	Lee S
Chaetomium funicola	CMW 22191	*Leucadendron sp.*	Lee S
Chaetopsina sp.	CMW 592	*Pinus ocarpa*	Wingfield MJ
Chaetosbolisia sp.	CMW 18009	*Ischyolepsis cf. gaudichaudiana*	Lee S
Chaetosphaeria sp.	CMW 18758	*Ischyolepsis cf. gaudichaudiana*	Lee S
Chaetosticta sp.	CMW 18000	*Pinus ocarpa*	Wingfield MJ
Chalara elegans	CMW 4683	*Cinnamomam camphorum*	Wingfield MJ
Chalara elegans	CMW 4690	Acacia mearnsii	Wingfield MJ
Chalara elegans	CMW 5463	Ground nuts	Geldenhuis N
Chrysoporthe austroafricana	CMW 8755	*Eucalyptus grandis*	Vd Westhuizen I
Cladosporium cladosporioides	CMW 23422	*Tipuana tipu*	Mehl J
Coelomycete sp.	CMW 330	*Windringtonica cedarbergensis*	Wingfield MJ
Coniodictyum chevalieri	CMW 23046	*Zizyphus mucronanta*	Maier W
Coniodictyum chevalieri	CMW 23047	*Zizyphus mucronanta*	Maier W
Coniothyrium zuluense	CMW 5690	*Eucalyptus grandis* clone & clone hybrid	Van Zyl L
Coniothyrium zuluense	CMW 5691	*Eucalyptus grandis* clone & clone hybrid	Van Zyl L
Conithyrium zuluense	CMW 11220	*Eucalyptus grandis* clone and clone hybrid	Van Zyl L
Cryphonectria cubensis	CMW 9342	*Tibouchina sp.*	Roux L
Cryphonectria cubensis	CMW 9344	*Tibouchina sp.*	Roux L
Cryphonectria cubensis	CMW 10060	*Syzygium cordatum*	Heath R, Roux L
Cryphonectria cubensis	CMW 10061	*Syzygium cordatum*	Heath R, Roux L
Curvularia unata	CMW 6139	*Musa sp*. (Banana)	Viljoen A
Curvularia pallescens	CMW 6176	Banana	Viljoen A
Cylindrocarpon sp.	CMW24211	*Ptetocarpus angolensis*	Mehl J
Cylindrocarpon sp.	CMW 24217	*Ptetocarpus angolensis*	Mehl J
Cylindrocladium candelabrum	CMW3912	*Acacia mearnsii*	Roux L
Cylindrocladium macroconidiale	CMW 28527	*Eucalyptus grandis*	Crous PW
Cylindrocladium scoparium	CMW1166	*Eucalyptus nitens*	Swart W
Cylindrocladium scoparium	CMW 2151	*Eucalyptus nitens*	Wingfield MJ
Cylindrocladium sp.	CMW 5636	*Podocarpus sp.*	Unknown
Cylindrocladium sp.	CMW 9151	*Acacia mearnsii*	Lombard L

Taxon	Culture number	Host/Source	Collector
Cylindrotrichum oligospermum	CMW 16695	Unknown	Lee S
Cytagonospora sp.	CMW 18247	*Ischyrolepis cf. gaudichaudiana*	Lee S
Cytospora eucalypticola	CMW 226	*Eucalyptus grandis*	Unknown
Cytospora eucalypticola	CMW 914	*Eucalyptus grandis*	Wingfield MJ
Cytospora sp.	CMW 1505	*Eucalyptus nitens*	Wingfield MJ
Cytospora sp.	CMW 1506	*Eucalyptus nitens*	Wingfield MJ
Cytospora sp.	CMW 5442	*Eucalyptus* hybrid (GC550)	Smith H
Cytospora sp.	CMW 18278	*Elegia equisetacea*	Lee S
Cytospora sp.	CMW 20895	*Pterocarpus augolensis*	Mehl J
Cytospora sp.	CMW 20902	*Tipuana tipu*	Mehl J
Davidiella impi	CMW 37667	*Eucalyptus* hybrid clone (GC540)	Gryzenhout M
Decaisnella sp.	CMW 17973	*Cannomois virgata*	Lee S
Diaporthe ambigua	CMW 5288	Apple	Smit A
Dictyochaeta simplex	CMW 18037	*Elegia equisetacea*	Lee S
Epicoccum nigrum	CMW 6140	*Musa sp*. (Banana)	Viljoen A
Erythricium salmonicolor	CMW 7130	*Eucalyptus macarthurrii*	Roux J
Erythrogleum sp.	CMW 17969	*Restio filiformis Poir.*	Lee S
Erythrogleum sp.	CMW 17972	*Restio filiformis Poir.*	Lee S
Erythrogleum sp.	CMW 18015	*Ischyrolepis cf. gaudichaudiana*	Lee S
Erythrogleum sp.	CMW 18301	*Elegia capensis*	Lee S
Erythrogloeum sp.	CMW 17970	*Restio filiformis Poir.*	Lee S
Erythrogloeum sp.	CMW 18301	*Elegia capensis*	Lee S
Eutypella sp.	CMW 18256	*Brabejum stellatifolium*	Lee S
Eutypella sp.	CMW 22119	*Protea nitida*	Lee S
Fairmaniella sp.	CMW 17999	*Ischyrolepsis cf gaudichaudiana*	Lee S
Fusarium annanarum sp. nov.	CMW 18685	Unknown	Van Wyk PS
Fusarium circinatum	CMW 13229	*Pinus sp.*	Viljoen A
Fusarium circinatum	CMW 13232	Lab strain	Venter E
Fusarium circinatum	CMW 13233	Lab strain	Venter E
Fusarium crookwellense	CMW 7013	Potato	Marasas WFO
Fusarium crookwellense	CMW 7195	Wheat	Marasa WFO
Fusarium graminearum	CMW 6424	*Eucalyptus grandis*	Unknown
Fusarium graminearum	CMW 6425	*Acacia mearnsii*	Roux J
Fusicoccum atrovirens sp. nov.	CMW 22674	*Pterocarpus angolensis*	Mehl J
Fusicoccum atrovirens sp. nov.	CMW 22682	*Pterocarpus angolensis*	Mehl J
Fusicoccum sp.	CMW 794	*Cryptomeria lawsonana*	Unknown
Fusicoccum sp.	CMW 795	*Cryptomeria lawsonana*	Unknown
Fusicoccum sp.	CMW 22171	*Erica plukenetii*	Lee S
Fusicoccum sp.	CMW 22251	*Restionaceae sp.*	Lee S

Taxon	Culture number	Host/Source	Collector
Fusicoccum sp.	CMW 35612	*Apodytes dimidiata*	Marincowitz S
Fusicoccum sp.	CMW 35721	*Protea simplex*	Marincowitz S
Fusicoccum sp.	CMW 3910	*Acacia mearnsii*	Roux L
Fusicoccum sp.	CMW 18288	*Restio quadratus*	Lee S
Ganoderma sp.	CMW25886	*Acacia mellifera*	Roux J
Ganoderma sp.	CMW 25887	*Rhus leptodictya*	Roux J
Ganoderma sp.	CMW 25891	Unknown	Roux J
Ganoderma sp.	CMW 25892	*Terminalia sp.*	Roux J
Ganoderma sp.	CMW 25893	*Rhus sp.*	Wingfield MJ
Ganoderma sp.	CMW 25894	*Jacaranda mimosifolia*	Wingfield MJ
Ganoderma sp.	CMW 25895	*Acacia tortilus*	Slippers B
Ganoderma sp.	CMW 25896	*Rhus lancia*	Slippers B
Ganoderma sp.	CMW 25897	*Salix bakylonica*	Heath R
Ganoderma sp.	CMW 25898	*Acacia kaffera*	Coetzee M
Ganoderma sp.	CMW 25899	*Acacia sp.*	Coetzee M
Ganoderma sp.	CMW 25900	*Acacia sp.*	Coetzee M
Ganoderma sp.	CMW 25901	*Acacia mearnsii*	Roux J
Ganoderma sp.	CMW 25902	*Acacia xantophloea*	Roux J
Ganoderma sp.	CMW 25903	*Quercus sp.*	Greyling I
Ganoderma sp.	CMW 29957	*Acacia sp.*	Coetzee M
Ganoderma sp.	CMW 25883	African tulip tree	Roux J
Leptosphaeria sp.	CMW 16675	*Thamnochortus cf. insignis*	Lee S

*CMW 17970 (denoted in bold) is used in this study

APPENDIX C: ITS sequence of CMW 17970

Isolate source: obtained from the FABI culture collection (FCC); previously identified as *Erythrogloeum sp.* (referred to as CMW 17970/*Clonostachys rosea* 17970/*C. rosea* 17970 in this study)

```
  1 GGAAGTAAAA GTCGTAACAA GGTCTCCGTT GGTGAACCAG CGGAGGGATC ATTACCGAGT
 61 TTACAACTCC CAAACCCATG TGAACATACC TACTGTTGCT TCGGCGGGAT TGCCCCGGGC
121 GCCTCGTGTG CCCCGGATCA GGCGCCCGCC TAGGAAACTT AACTCTTGTT TTATTTTGGA
181 ATCTTCTGAG TAGTTTTTAC AAATAAATAA AAACTTTCAA CAACGGATCT CTTGGTTCTG
241 GCATCGATGA AGAACGCAGC GAAATGCGAT AAGTAATGTG AATTGCAGAA TTCAGTGAAT
301 CATCGAATCT TTGAACGCAC ATTGCGCCCG CCAGTATTCT GGCGGGCATG CCTGTCTGAG
361 CGTCATTTCA ACCCTCATGC CCCTAGGGCG TGGTGTTGGG GATCGGCCAA AGCCCGCGAG
421 GGACGGCCGG CCCCTAAATC TAGTGGCGGA CCCGTCGTGG CCTCCTCTGC GAAGTAGTGA
481 TATTCCGCAT CGGAGAGCGA CGAGCCCCTG CCGTTAAACC CCCAACTTTC CAAGGTTGAC
541 CTCAGATCAG GTAGGAATAC CCGCTGAACT TAAGCATATC AATAAGCGGA GGA
```

APPENDIX C: Primer design

```
PrC MRVSALLSILPLVAAAPAKREEVAPLHVPRDVEVIPGKYIVKLKEGVVSISSTISSIEAK 60
Sep MRVSALLSILPLVAAAPAKREEVAPLHVPRDVEVIPGKYIVKLKEGVVSISSTISSIEAK 60
```

CPR-For: 5'-**GGTCGAGTACGTCGAGCAGGATGC**-3'

```
PrC PDFEYEGGFQGFVRIYTSLVSQTRNMQITDNSSHRLVEYVEQDAIVSISATQTGAPWGIA 120
Sep PDFEYEGGFQGF---AGALT--EAEVQ--ALRESPEVEYVEQDAIVSISATQTGAPWGIA 113

PrC RLSNTNTGSTTYTYDDSAGDGTCAFIIDTGIYTSHSDFGGRASFAANYVDSSSTDGNGHG 180
Sep RLSNTNTGSTTYTYDDSAGDGTCAFIIDTGIYTSHSDFGGRASFAANYVDSSSTDGNGHG 173

PrC THVAGTVGGTTYGVAKKTKLYAVKVLDSSGSGTTSGVIAGMNYVTNSAGTYSCPKGVVVN 240
Sep THVAGTVGGTTYGVAKKTKLYAVKVLDSSGSGTTSGVIAGMNYVTNSAGTYSCPKGVVVN 233

PrC MSLGGGYSASLNTAANNIVSAGYFLAVAAGNSAANAANYSPASAASACTVGATTSSDALA 300
Sep MSLGGGYSASLNTAANNIVSAGYFLAVAAGNSAANAANYSPASAASACTVGATTSSDALA 293

PrC SYSNYGSIVDILAPGSSVLSAYNNGGTATLSGTSMASPHVAGLGAYYLGLGRASASGLCS 360
Sep SYSNYGSIVDILAPGSSVLSAYNNGGTATLSGTSMASPHVAGLGAYYLGLGRASASGLCS 353
```

CPR-Rev: 5'-GGGGAGGCCATGGAGGTACCAGAG -3'

```
PrC YIVSTGLTGRITSVPSGTPNVLAHLV 386

Sep YIVSTGLTGRITSVPSGTPNVLAHLV 379
```