Xylanase Hyper-Producer: Genome of the Thermophilic Fungus *Thermomyces lanuginosus*

Nokuthula Peace Mchunu

2014

Xylanase Hyper-Producer: The Genome of the Thermophilic Fungus *Thermomyces lanuginosus*

Nokuthula Peace Mchunu

2014

Submitted in complete fulfillment of the requirements for the Degree of Doctor of Technology (Biotechnology), Department of Biotechnology, Faculty of Applied Sciences, Durban University of Technology, Durban, South Africa.

Supervisors:

Prof. Kugen Permaul

Prof. Suren Singh

DECLARATION

I hereby declare that this dissertation is my own, unaided work. It is being submitted for the degree, Doctor of Technology, to the Durban University of Technology, Durban, South Africa. It has not been submitted before for any degree or dissertation to any other institution.

_____ _____

Nokuthula Peace Mchunu **Date**

_____ _____

Supervisor : Prof. Kugen Permaul **Date**

_____ _____

Supervisor : Prof. Suren Singh **Date**

ACKNOWLEDGEMENTS

I would like to acknowledge the following:

Prof. Kugen Permaul and Prof. Suren Singh (Durban University of Technology) for their guidance, patience, support and the many opportunities and doors they opened for me;

Prof. Maqsudul Alam (former Director of the Centre for Chemical Biology, Centre for Chemical Biology (CCB) and also a Faculty member at Hawaii University) and all the staff of CCB (Universiti Sains Malaysia) for their input into the study and allowing me to do a project of this nature;

The National Research Foundation, South Africa for the financial support;

Ms Aneliswa Makhathini, Mrs. Siphi Mbamali and Mrs Kameshnee Naidoo (Durban University of Technology), for their friendship and support;

My Mum, sister, Fikile, Zama Cele and family for their tremendous love, encouragement and support;

Zakhe Nxumalo, for his support;

Mostly, all praise must go to the Almighty **JEHOVAH** who has given me the strength and skill to complete my studies.

"You are worthy, Jehovah, even our God, to receive the glory and the honor and the power, because you created all things, and because of your will they existed and were created."
Revelation 4:11

CONTENTS

DECLARATION ..... **i**

ACKNOWLEDGEMENTS ..... **ii**

CONTENTS ..... **iii**

LIST OF FIGURES ..... **viii**

LIST OF TABLES ..... **xii**

ABSTRACT ..... **xiii**

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW **15**

1.1 INTRODUCTION 15

1.2 FUNGAL BIOLOGY 17

1.3 INDUSTRIAL APPLICATION OF FILAMENTOUS FUNGI 18

1.3.1 Food industry 18

1.3.2 Fermentation processes 19

1.3.3 Enzyme production 20

1.3.3.1 Enzymes in food and beverage industry 22

1.3.3.2 Biopulping 23

1.3.3.3 Animal feed 24

1.3.3.4. Recombinant enzymes 24

1.3.4 Biofuels 25

1.3.5 Filamentous fungi as plant pathogens 28

1.3.6 Pharmaceuticals production 29

1.4 THERMOPHILES 30

1.5 *THERMOMYCES LANUGINOSUS* 33

1.6 WHOLE GENOME SEQUENCING PLATFORMS 35

1.6.1 Roche genome sequencer 38
1.6.2 Illumina/Solexa genome analyzer 40
1.6.3 SOLiD sequencing platform 43
1.6.4 Sequence assembly 45
1.6.5 *De novo* assembly vs mapping assembly 47
1.7 FUNCTIONAL GENOMICS 49
1.7.1 Transcriptome 50
1.7.2 DNA microarrays 51
1.7.3 Metabolomics 52
1.7.4 Phenotypic microarray 52
1.8 METABOLIC RECONSTRUCTION 53
1.9 MOTIVATION FOR THE STUDY 54
CHAPTER 2: GENOME ANALYSIS OF *T. lanuginosus*..... 55
2.1 INTRODUCTION 58
2.2 MATERIALS AND METHODS 64
2.2.1 Growth and maintenance of *T. lanuginosus* 64
2.2.2 Genomic DNA isolation 64
2.2.3 Total RNA isolation 65
2.2.4 DNA and RNA quantification 65
2.2.5 Agarose gel electrophoresis 66
2.2.6 Total RNA integrity validation 66
2.2.7 Roche FLX Raw sequencing data generation 67
2.2.7.1 Paired-end library preparation 67
2.2.7.2 cDNA library preparation 68
2.2.7.3 emPCR 68

2.2.7.4 Roche Titanium sequencing 69
2.2.8 Solexa Illumina Sequencing 70
2.2.9 Genome assembly 70
2.2.10 Genome annotation 71
2.2.11 Detection of CAZy family proteins 71
2.2.12 Manual curation of data 72
2.2.13 Synteny analysis 72
2.2.14 Fungal proteome comparisons 72
2.3 RESULTS 73
2.3.1 Genomic DNA and total RNA isolation 73
2.3.1.1 Genomic DNA isolation 73
2.3.1.2 Total RNA isolation 73
2.3.2 Sequencing statistics 75
2.3.3 Genome assembly 76
2.3.3.1 Genome overview 76
2.3.3.2 Genome comparison with other fungi 80
2.3.4 General metabolism 83
2.3.4.1Cluster of Orthologous Groups (KOG/COG) analysis 83
2.3.4.2 Information, storage and processing 84
2.3.4.3 Cellular Processes and signaling 85
2.3.4.4 Metabolism 86
2.3.4.5 Poorly -haracterized genes 87
2.3.5 Detection of CAZy family protein 88
2.3.5.1 Comparison with other fungi 88
2.3.5.2 Cellulose-degrading genes 89

2.3.5.3 Xylan-degrading genes 92
2.3.6 Thermostability and thermo-tolerant features 93
2.3.6.1 DNA stability 93
2.3.6.2 Membrane stability and electron chain transport 94
2.3.6.3 Cell stability at high temperature 95
2.3.6.4 Heat shock proteins 97
2.4 DISCUSSION 99
CHAPTER 3: METABOLIC RECONSTRUCTION, PHENOTYPIC AND DNA MICROARRAY 114
3.1 INTRODUCTION 123
3.2 Materials and Methods 129
3.2.1 Pathway Reconstruction 129
3.2.2 Phenotypic MicroArray 129
3.2.3 DNA MicroArray 131
3.3 Results 132
3.3.1 Pathway reconstruction 132
3.3.2 Phenotypic MicroArray 139
3.3.2.1 Cluster analysis of carbon source utilization profiles 139
3.3.2.2 Hexose and pentose metabolism and growth induction 147
3.3.2.3 Oligosaccharide assimilation and growth 150
3.3.2.4 Amino acid assimilation and growth 153
3.3.3 DNA microarray 156
3.3.3.1 Xylose 156
3.3.3.2 Glucose 159
3.4 DISCUSSION 162

CHAPTER 4: GENERAL DISCUSSION 172

4.1 RESEARCH PERSPECTIVE 172

4.2 FUTURE PROSPECTS..... 180

5. REFERENCES 182

6. APPENDIX: 221

LIST OF FIGURES

Figure 1: A summary of global revenue of industrial enzymes market, 2009-2016 (BCC, 2012). 22
Figure 2: Major producers of biofuels (BP Statistical Review of World Energy, 2012). 26
Figure 3: World energy consumption (BP Statistical Review of World Energy, 2012). 27
Figure 4: Diagram showing the major components (cellulose, hemicellulose and lignin) found in wood fibres. 28
Figure 5: *T. lanuginosus* grown on potato dextrose agar after 7 days incubation at 50°C (Singh *et al.,* 2000a). 33
Figure 6: *T. lanuginosus*'s conidia under transmission electron microscopy (TEM) (Singh *et al.,* 2000a). 34
Figure 7: Overview of the Roche 454 sequencing system. DNA is sheared into small fragments to which adapters are ligated, fragments are attached to beads, mixed into an emulsion, amplified by emulsion PCR, deposited in wells of a picotitre plate, then sequences are determined by pyrosequencing (www.futuremedicine.com)..... 40
Figure 8: Overview of the Illumina/solexa platform strategy (Illumina Inc). 42
Figure 9: Schematic showing a basic genome assembly process (Wu *et al.,* 2012). 47
Figure 10: Schemtic showing *de no* assembly with overlapping regions (blue) and gaps (red dotted line where no ovelarpping sequences are found-www.nimr.mrc.ac.uk). 48
Figure 11: Schematic view of reference or mapping assembly method for genomes. In whole-genome assembly, the BAC fragments (red) and the reads (black) are combined to produce a contig and a consensus sequence (green). The contigs are connected into scaffolds, shown in red, by pairing end sequences, which are also called mates. If there is a gap between consecutive contigs, it has a known size. Next, the scaffolds are mapped to the genome (Venter *et al.,* 2001)..... 49
Figure 12: Diagram showing an example of workflow of genome annotation adapted from (Hull *et al.,* 2006). 60
Figure 13: Agarose gel electrophoresis of *T. lanuginosus* genomic DNA (Lanes 2 & 3) against high range molecular weight markers (Lane 1) and lambda/HindIII (lane 4). 73

Figure 14: Agarose gel electrophoresis of *T. lanuginosus* total RNA (Lanes 1-7). Lanes 1, 3, 5 and 7 show raw total RNA samples while Lanes 2, 4, and 6 show RNA samples that have been DNAse treated. 74
Figure 15: Bioanalyzer electropherogram of total RNA samples of *T. lanuginosus* confirming intact RNA as shown by the presence of the 18S and 28S peaks. 74
Figure 16: Gbrowser image showing typical information for *T. lanuginosus* genome. The Gbrowser uses contig information and all the evidence used for gene annotations (gene predictions from MAKER, Augustus, Genemark and ESTs). 79
Figure 17: Synteny analysis of *T. lanuginosus* vs *A. nidulans* genomes using WebACT software. 82
Figure 18: Synteny analysis of *T. lanuginosus* vs *A. fumigatus* genomes using WebACT software. 82
Figure 19: Synteny analysis of *T. lanuginosus* vs *T. reesei* genomes using WebACT software. 82
Figure 20: Annotation of *T. lanuginosus* genes into major groups represented in COG analysis. A total of 4480 proteins were annotated by KOG/COG. The proteins were assigned into different KOG/COG major functional categories. 83
Figure 21: *T. lanuginosus* genes that belong in information storage and processing group. The sub-groups are presented both in percentage and the actual number of genes. The sub- groups found in information storage and processing belong in general to DNA, RNA and protein synthesis. 84
Figure 22: *T. lanuginosus* genes that belong to the cellular processes and signaling COGs. The sub-groups are presented both in percentage and the actual number of genes. 85
Figure 23: *T. lanuginosus* genes that belong in metabolism. The sub-groups are presented both in percentage and the actual number of genes. 86
Figure 24: *T. lanuginosus* genes that belong in poorly characterized genes. The sub-groups are presented both in percentage and the actual number of genes. The sub- groups are genes with unknown function or general predictions. 87
Figure 25: Phylogeny of cellulase enzymes found in *T. lanuginosus* created with ClustalW. The cellulases (1-4) were compared with the cellulase gene from *T. reesei* downloaded from the NCBI database. 90

Figure 26: Alignment of putative *T. lanuginosus* cellulases with a cellulase *T. reesei*. 91
Figure 27: Alignment of a putative cellulase number 2 from *T. lanuginosus* compared with a cellulase gene from *A. kawachi*. 92
Figure 28: Reconstructed pathway for trehalose synthesis in *T. lanuginosus*. The pathways were constructed using Pathway Studio (Araidne Genomics). 97
Figure 29: Phylogeny of heat shock protein analogs found in *T. lanuginosus* created with ClustalW. 98
Figure 30: Carbon sources (95) found in the filamentous fungal (FF) MicroPlate from Biolog, Inc. 130
Figure 31: Predicted hexose transformation pathway in *T. lanuginosus*..... 134
Figure 32: Predicted arabinose degradation pathway in *T. lanuginosus*. 135
Figure 33: Predicted trehalose biosynthesis and conversion pathway in *T. lanuginosus*..... 136
Figure 34: Predicted cellulose degradation pathway in *T. lanuginosus*. 136
Figure 35: Predicted general sugar and sucrose transportation in *T. lanuginosus*. 137
Figure 36: Predicted amino-sugar biosynthesis pathway in *T. lanuginosus*. 138
Figure 37: Joining cluster analysis applied to 95 carbon sources based on their metabolism of carbon sources by *T. lanuginosus* measured at 490 nm using the Biolog system. 140
Figure 38: Joining cluster analysis applied to 95 carbon sources based on the growth of *T. lanuginosus* on these carbon sources. Growth was measured at 750 nm using the Biolog system. 141
Figure 39: Metabolism of monomeric sugars (hexose and pentose) by *T. lanuginosus* SSBP. Metabolism was measured at an absorbance of 490 nm for 68 hours at 2 h intervals using Phenotypic MicroArray. 148
Figure 40: Growth of *T. lanuginosus* SSBP on monomeric sugars (hexose and pentose). Growth was measured at an absorbance of 750 nm for 68 hours at 2 h intervals using Phenotypic MicroArray**.** 149
Figure 41: Metabolism of oligosaccharides by *T. lanuginosus* SSBP. Assimilation was measured at an absorbance of 490 nm for 68 hours at 2 h intervals using Phenotypic MicroArray. 151

Figure 42: Growth of *T. lanuginosus* SSBP on oligosaccharide compounds. Growth was measured at an absorbance of 750 nm for 68 hours at 2 h intervals using Phenotypic MicroArray. 152
Figure 43: Metabolism of amino acid based compounds by *T. lanuginosus* SSBP. Metabolism was measured at an absorbance of 490 nm for 68 hours at 2 h intervals using Phenotypic MicroArray. 154
Figure 44: Growth of *T. lanuginosus* SSBP on amino acid based compounds. Growth was measured at an absorbance of 750 nm for 68 hours at 2 h intervals using Phenotypic MicroArray. 155

LIST OF TABLES

Table 1: Enzymes use in the food and feed industries adapted from (Ghorai *et al.*, 2009)..... 21

Table 2: Recombinant enzymes used in the food and feed industries adapted from (Ghorai *et al.*, 2009)..... 25

Table 3: Examples of thermophilic fungi (Salar and Aneja, 2007; Singh *et al.*, 2003) 32

Table 4: Potential industrial proteins from *T. lanuginosus (*Puchart *et al.*, 1999; Singh *et al.*, 2000b)..... 35

Table 5: Platforms and the detailed information on different next generation sequencing (NGS) technologies (Zhang *et al.*, 2011) 37

Table 6: Selected fungal genomes found on the FUNGIpath database (Grossetête *et al.*, 2010) 59

Table 7: Summary of the sequencing data of *T. laniginosus* produced by Roche 454 and Solexa sequencing strategies 76

Table 8: General features of *Thermomyces lanuginosus* genome 77

Table 9: *T. lanuginosus* assembled scaffolds 78

Table 10: Comparative analysis of *T. lanuginosus* genome with other fungal species..... 81

Table 11: Comparison of CAZy enzyme families, by class, in the selected fungal genomes. 89

Table 12: Selected enzymes that can degrade various carbohydrate substrates..... 93

Table 13: Comparative analysis of methyltransferase genes found in the *T. lanuginosus* genome and other selected fungal species (MTs, methyltransferases; MK, menaquinone; ubiE, ubiquinone; PETMs, phosphatidylethanolamine methyltransferases) 95

Table 14: Dimutases found in *T. lanuginosus* 96

Table 15: Metabolism of the 95 carbon sources by *T. lanuginosus* measured at 490 nm using phenotypic microarray..... 142

Table 16: Growth analysis of *T. lanuginosus* in 95 different carbon sources measured at 750 nm using phenotypic microarray. 144

Table 17: DNA microarray of *T. lanuginosus* in xylose showing genes that were up-regulated 3-fold 157

Table 18: DNA microarray of *T. lanuginosus* in glucose showing genes that were up-regulated 3-fold..... 160

ABSTRACT

The global demand for green technology has created a need to search for microbes that can play an active role in advancing a greener and cleaner future. Microbial enzymes are nature's keys to life and their efficiency, specificity and environmental-friendliness has lead to their increased use in industrial processes. *Thermomyces lanuginosus* is a thermophilic fungus that can degrade plant biomass and produces a variety of enzymes that have industrial application. The fungus *T. lanuginosus* SSBP has been reported in literature to produce the highest level of xylanase among other *Thermomyces* strains and some of its enzyme s viz., amylase and lipase are already being used. Because of this ability, it has been identified as one of the organisms that can have various industrial applications. Although a few proteins from this fungus have been cloned and used commercially, the vast majority are still unknown. In order to identify new protein candidates and understand their biochemical interactions, the *T. lanuginosus* genome (DNA) and the transcriptome (mRNA) were sequenced using 454 Roche and Solexa sequencing platforms. Genome and transcriptome data was assembled using Newbler software forming a genome size of 23.3 Mb contained 30 scaffolds. Protein prediction identified 5105 candidates as protein-coding genes and these gene models were supported by expressed sequence tag and transcriptomic data. The annotated data was assembled into metabolic pathways in order to identify functional pathways and validate the accuracy of the annotation process. *T. lanuginosus* is usually found in composting plant material thus protein related to plant hydrolysis were analysed. The total number of plant biomass-degrading and related proteins that fall into the carbohydrate-active enzyme (CAZy) family was 224. Most of these proteins were similar to proteins found in other filamentous fungi. Surprisingly, *T. lanuginosus* contained a single gene coding for xylanase which hydrolyses xylan although this organism is well known for being among the highest producers of this enzyme. An important subset of the above group of proteins is the cellulose degrading-proteins as this can be used in biofuel production. Eight candidates belonging to this group were identified, making this fungus significant in the biofuels. Among the eight cellulase candidates, phylogenetic analysis revealed that three of them were closely related to *Trichoderma reesei*, a well known industrial cellulase-producer. Utilization of cellulase-related compounds was validated by phenotypic microarray experiments, with cellobiose having inducing biomass in *T. lanuginosus*. Proteins that are involved in high temperature survival are vital for the survival

of this thermophilic fungus. Interestingly, *T. lanuginosus* contains 19 heat shocking proteins which are responsible for thermostability. Another adaptation identified in this fungus is the accumulation of trehalose to combat heat stress. Furthermore, *T. lanuginosus* contains the highest reported number methyltransferases, which have been linked to producing thermostable proteins and higher energy production. Also because of this organism's ability to grow on composting environments, the assimilation and ability to produce biomass on different carbon sources were analysed using phenotypic microarray technique. The results showed that xylose was the best compound to induce biomass followed by trehalose, maltose and maltotriose. The genomic sequencing of this fungus has provided valuable information that can be used for various biotechnological applications, as well as providing greater insights into its thermostability. Understanding the metabolic pathways involved may allow for manipulation to increase production of these enzymes or cloning into other hosts. This can have an impact in the field of biofuel production and other plant biomass-related processes.

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

Fungi comprise one of the three major kingdoms that make up the eukaryotic organisms, the others being plants and animals. The fungal kingdom is made up of unicellular fungi (yeasts) and filamentous fungi (molds). These play a key role in economics, health, ecology and in our daily lives (Gazi *et al.*, 2007). Fungal organisms are found in diverse natural habitats living on degraded or decomposing organic material (Arantes *et al.*, 2011). Yeasts have been a part of human diet for millennia as they have been used in baking and wine making. While filamentous fungi have been used in cheese-making, more recently they have become widely known for their industrial applications, namely, the production of food processing enzymes and metabolites such as antibiotics and organic acids (Chiang *et al.*, 2009; Dodd and Cann, 2009; Marjamaa *et al.*, 2013; Naganagouda *et al.*, 2009; Peterson *et al.*, 2011; Puchart *et al.*, 1999; Shraddha *et al.*, 2011; Wiater *et al.*, 2008). Some of these biological molecules produced from fungi are secreted and induced by appropriate upstream molecules. The biological pathways used by these organisms can be manipulated further for human benefits, either economically or medically. Therefore, there is a need for research in fungal metabolism and on its secreted proteins.

Thermomyces lanuginosus is a thermophilic filamentous fungus that can be found in most composting environments. It is known to produce a host of thermophilic enzymes which can be applied in industrial processes (Puchart *et al.*, 1999; Singh *et al.*, 2000b). These thermostable enzymes include: amylase, glucoamylase, xylanase, lipase, phytase, protease and chitinase, which can be applied in various industries. Amylases and glucoamylases can be used in the food industry for the production of sugar syrup, baking industry and animal feed(Kunamneni *et al.*, 2005). Xylanases have been used in the pulp and paper industry for bleaching of pulp instead of chlorine which produces toxic waste (Bissoon *et al.*, 2002; Christopher *et al.*, 2005; Kumar *et al.*, 2009c; Madlala *et al.*, 2001; Manimaran *et al.*, 2009). This enzyme can also be used in the pre-digestion of animal feed and in the food industry. Phytase is used in the pre-digestion of animal feed in order to release phosphate, making it bio-available (for animal usage) . Chitinases

have potential for application in bioremediation of seafood waste by degrading the chitin component which makes up shells in crustacean organisms (Hartl *et al.*, 2012; Samolski *et al.*, 2009). Some of these enzymes have been cloned and expressed using different hosts, but with many of them there has been limited success (Stephens *et al.*, 2007). There has also been an interest in some of the accessory enzymes produced by *T. lanuginosus* involved in xylan degradation but these enzymes are produced in very low quantities compared to the xylanase which cleaves the main backbone of xylan or the amylase enzyme involved in starch degradation (Singh *et al.*, 2000a).

Thermostable carbohydrate-degrading proteins can also be applied in one of the most promising and life-altering fields of biofuel production, using non-food crops (Dodd and Cann, 2009; Hrmova *et al.*, 1991; Mandels and Andreotti, 1978; Marjamaa *et al.*, 2013). The biggest problem for bioethanol production is finding a suitable organism that converts different carbohydrate compounds directly to ethanol with biological and physiological properties that fit into the process (Taherzadeh and Karimi, 2008). Recently, the total production of biofuels is around 80 million litrers (BP, 2012) and this can be increased if suitable microorganisms or enzymes for the conversion are identified in combination with advances in process design.

In order to fulfill this need, many have turned to genomics or genome sequencing. Genome sequencing enables the determination of the complete genetic information present in the genomes of various organisms. The relationship between genes can be deduced with some confidence. Sequencing provides insight into genome organization and evolution, and the mechanisms involved. The sequencing of the *T. lanuginosus* genome is important because this will provide clues on some of the existing pathways in this organism. It may also provide necessary information to understand a long-standing pecularity about this fungus, i.e., why it produces a large amount of the xylanase enzyme while other enzymes involved in the same pathway are produced in minor quantities. This project will also provide information on the genes that have been difficult to clone and this will be the first report on sequencing of the genome of this fungus. Genomic data will also reveal information on possible promoter sequences which may lead to over-expression of already cloned genes. In the future, the

information from this study may be useful in the production of biofuels as some of the enzymes produced by this fungus can be applied in this field.

1.2 FUNGAL BIOLOGY

Fungi are a diverse group of eukaryote microorganisms made up of unicellular yeasts and multicellular molds. The most studied fungal species is the unicellular yeast, *Saccharomyces cerevisiae*. This yeast is widely used in industry for alcohol production and baking. It has also become the model organism for the study of eukaryotic systems. Studies using yeast have been important in understanding eukaryotic biological systems including humans, while providing the ease of genetic manipulation because of being single-celled. Nevertheless, multicellular fungi still remain the most attractive candidate for industrial application and bio-prospecting for industrial enzymes (Blättel *et al.*, 2011; Ghorai *et al.*, 2009; Nevalainen *et al.*, 2005).

Multicellular fungal organisms or filamentous fungi which are made up of a mass of thread-like filaments called hyphae, which make up the fungal body called the mycelium. The hyphae itself is composed of fungal cells connected to each other in a chain formation. Filamentous fungal species contain two types of hyphae, (a) those containing hyphae with continuous cytoplasm with many nuclei or (b) those with individual cell containing septa which are barriers across the filament. Fungi unlike bacteria have a cell wall that is made up of a polysaccharide called chitin, not peptidoglycan (Hartl *et al.*, 2012).

The two large fungal divisions are the Ascomycetes and Basidiomycetes. Basidiomycetes externally produce specialized cells called basidia while ascomycetes produce spores internally inside a sac called the ascus (McLaughlin *et al.*, 2009; Pearce, 2012). In general, the group basidiomycetes include mushrooms, bracket fungi, and numerous mold-like fungi called rusts and smuts. These are usually involved in damaging food crops, grain and other plants. Ascomycota includes yeast, cup fungi, truffles, morels and mildews (e.g., *Saccharomyces cerevisiae*, *Aspergillus niger*, *Trichoderma reesei*). Filamentous fungi are unique in that they can produce a wide range of natural products including enzymes and secondary metabolites. (Hibbett *et al.*, 2007; McLaughlin *et al.*, 2009; Stajich *et al.*, 2009). Filamentous fungi can be found in

the both Basidiomycetes and Ascomycetes. Filamentous fungi have evolved to produce high amounts of industrially-important proteins or enzymes. For these reasons, filamentous fungi are being studied for application in a number of industrial processes (de Oliveira and de Graaff, 2011).

1.3 INDUSTRIAL APPLICATION OF FILAMENTOUS FUNGI

1.3.1 Food industry

Mushrooms have been used for consumption by humans for thousands of years. In recent years, the consumption of fungi has increased as people become more health conscious and the number of vegetarians is increasing. Among one of the earliest uses of filamentous fungi is cheese which is thought to have begun 4000 years ago but some archaeological evidence shows the use of edible wild mushroom in Chile, approximately 13 000 years ago (Rojas and Mansur, 1995; Toma *et al.*, 2013). In the production of cheese, the presence of visible fungal mycelium is sometimes a desirable characteristic. Some of the most famous cheeses, known for their moldy appearance, are Roquefort, also known as blue cheese and Camembert. The filamentous *Penicillium* species is used in both these cheeses, *P. camemberti* in Camembert cheese and *P. roqueforti* in Roquefort cheese (Ropars *et al.*, 2012). Fungal enzymes have also been used in cheese manufacturing by the addition of rennet, lipase and proteinases (Ghorai *et al.*, 2009).

Filamentous fungi have served as a direct food source for centuries, as in the case of mushrooms. Mushrooms have been consumed directly or processed and used as a delicacy world-wide. Fungi have numerous advantages compared to a general vegetarian and meat diet. They have good protein content (20-30%), and are thus capable of being used as a meat substitute (Braaksma and Schaap, 1996). The cells have a high vitamin B content, are low in fat and cholesterol-free and the chitin-containing cell wall acts as a source of dietary fibre (Mishra *et al.*, 2013; Reis *et al.*, 2012). Their dietary benefits and the low cost of cultivation have made mushrooms an attractive economic investment, since most of them can be cultivated on agricultural or other industrial waste products. All agricultural production generates enormous waste because so little of each crop is actually used (Zhang *et al.*, 2002). The waste produced in agricultural activities can be used in the cultivation of mushrooms. Oyster mushrooms (*Pleurotus* species) have been shown

to grow readily on cotton wastes. Similarly, although the straw mushroom is traditionally grown in South East-Asia on rice straw, it can also be grown on cotton waste (Zhang *et al.*, 2002). It is speculated that there are hundreds of species of edible mushrooms in the wild, but less than 20 species are used extensively as food and only 8–10 species are regularly cultivated to any significant extent (Ko *et al.*, 2005). Examples of these regularly-consumed mushrooms include button or Portobello mushrooms (*Agaricus bisporus*), straw mushrooms (*Volvariella volvacea*), oyster mushrooms (*Pleurotus ostreatus*), shiitakes (*Lentinula edodes*), and enokitake (*Flammulina* spp.). Levels of consumption vary, with South-East Asia being the largest consumers (Mishra *et al.*, 2013; Okhuoya and Etugo, 1993).

1.3.2 Fermentation processes

Citric acid production by *Aspergillus niger* is a good example of a fungus that can efficiently convert of simple sugars into a commodity product (Andersen *et al.*, 2011; Kubicek *et al.*, 2011). Fungal-based processes for manufacturing other organic acids such as gluconic, lactic, fumaric and malic acid are also in commercial use (Andersen *et al.*, 2011; Magnuson and Lasure., 2004). Many of the commercial production processes for organic acids are well established fungal biotechnology processes. Although the use of organic acids has less visible impact on society, these processes are not widely known but they are still very important. Commercial processes using *A. niger* can convert glucose to citric acid with greater than 80% efficiency and at final concentrations in hundreds of grams per litre (de Oliveira and de Graaff, 2011). Remarkably, organic acid production has received less research interest compared to other processes. Perhaps a greater understanding of this extraordinary capacity of filamentous fungi to produce organic acids in high concentrations will allow greater exploitation of these organisms via application of new knowledge derived from genomics-based biotechnology (Magnuson and Lasure., 2004).

The unicellular yeast, *S. cerevisiae* has dominated the use of fungi in the fermentation and the alcohol industry. This is due to the efficiency of this organism in converting inputs used in brewing, wine making and bread into desirable products. But the inhibition of this yeast by

alcohol concentrations of 10-12% and the inability to utilize starchy materials has led to the use of other organisms in biofuel production (Celestino *et al.*, 2006; Teh and Lutz, 2010).

In the fermentation processes, carbohydrates have to be converted into monomeric sugars. In the beer industry, malting under aerobic conditions in which typically barley grains are steeped in water for up to a day and then allowed to germinate under moist aerobic conditions. The germination process produces enzymes, which degrade the polysaccharides and proteins into simpler sugars and amino acids which is then fermented into alcohol (Spencer-Martins, 1994). In the production of 'sake', rice is used as starting material and *Aspergillus oryzae*, a filamentous fungus is inoculated on the surface of the rice. The fungus produces enzymes which are then secreted and convert the starch into simpler sugars, which then can be used by the yeast to produce sake beverage with a alcohol concentration of 18% which is difficult to produce with brewer's yeast alone (Yamane *et al.*, 2002).

1.3.3 Enzyme production

Industries and humankind have benefited enormously through the usage of enzymes as early as the start of cheese consumption, brewing and baking via yeasts. Enzymes are applied in various fields, including technical use, food manufacturing, animal nutrition, cosmetics, medication and as well as tools for research and development (Ghorai *et al.*, 2009; Hartl *et al.*, 2012; Li *et al.*, 2012; Maheshwari *et al.*, 2000; Marquez *et al.*, 2011; Niehaus *et al.*, 1999). More than 4000 enzymes are known and approximately 200 from microbial original types are being used commercially (Li *et al.*, 2012). The global markets for enzymes in industrial applications was valued at $3.1 billion in 2009 (BCC, 2012) and reached about $3.6 billion in 2010 (**Figure 1**). The estimated enzyme market for 2011 was about $3.9 billion. Enzymes used in the food industries comprise the largest segment of the industrial enzymes with $1.2 billion in 2010. The second-largest category is technical enzymes with $1.1 billion in 2010 and nearly $1.2 billion in 2011. It is projected that the enzyme market will grow at a compounded annual growth rate of 9.1% to reach $6 billion by 2016, while other independent research reports even higher growth estimations of $8 billion by 2015 (Li *et al.*, 2012). **Table 1** summarises most of the reported enzymes that have been applied in industry.

Table 1: Enzymes use in the food and feed industries adapted from (Ghorai *et al.*, 2009)**.**

Industries	Enzymes	References
Butter and butter oils	Catalase, glucose oxidase, lipase	Gupta et al. (2003)
Cheese	Rennet, lipase, proteinase	Freitas and Malcata (2000)
Animal Feed	Amylase, glucoamylases, glucanase, cellulase, pentosanase, xylanase, proteinases, phytase	Wang et. al. (2006)
Alcohol	Amylase, amyloglucosidase, glucanase, cellulase, cellobiase, pectinase, proteinase	Sharma et. al. (2002)
Biscuits	Amylase, cellulase, hemicellulase, proteinase, pentosanase	Taniwaki et al. (2001)
Breads	Amylase, amyloglucosidase, cellulase, glucanase, glucose oxidase, hemicellulase, lipase, pentosanase, proteinase	Taniwaki et al. (2001)
Brewing	Acetolactase, decarboxylase, amylase, amyloglucosidase, cellulase, glucanase, lipase, pentosanase, proteinase, xylanase	Okamura et al. (2001)
Coffee	Cellulase, hemicellulases, galactomannanase, pectinase	Soccol et al. (2008)
Confectionery	Amylase, invertase, pectinase, proteinase	Stroh (1998)
Egg processing	Proteinase, lipase phospholipase, catalase, glucose oxidase	Singh et al. (2007)
Fats	Esterase, glucose oxidase, lipase	Bobek et. al. (1994)
Fish	Proteinase	Prasad (2001)
Dairy products	Lactase, proteinase, sulphydryl oxidase, lactoperoxidase, lysozyme, peroxidase, catalase	Archer (2000), Beauchemin et. al (1999), Rode et al.(1999)
Dibittering	Peptidase, naringinase	Knoss et al. (1998)
Flavors	Glucanase, peptidase, proteinase, esterase, lipase, amylase	Shahani et al.(1976)
Fructose	Glucose isomerase, inulinase, amylase, amyloglucosidase, cellulase, glucanase, hemicellulases, isomerase, lipase, phospholipase, pectinases, protease	Sørensen et al. (2004)
Fruit, cloudy juices	Amylases, pectinase, cellulase, proteinase	Brandelli et al. (2005)
Vegetable and fruit	Cellulase, macerating enzyme, pectinases	Brandelli et al. (2005)
Tea	Cellulase, glucanase, pectinase, tannase	Pasha, and Reddy (2005)
Wine	Amylase, amyloglucosidase, cellulase, glucanase, hemicellulases, pectinase, proteass, glucose oxidase, catalase, pentosanase, anthocyanase	Okamura et al. (2001)
Malt extract	Amylase, amyloglucosidase, cellulase, glucanase, proteinase, xylanase	Feng, Larsen, and Schnürer (2007)
Animal oil/fats	Esterase, lipase, proteinase	Beldman,et. al. (1984)

Industy	Enzymes	References
Protein	Amylase, cellulase, glucanase, hemicellulases, pectinase, protease	Semenova et al. (2006)
Starch	Amylase, amyloglucosidase, cellulase, glucanase, hemicellulase, isomerase, lipase, phospholipase, pectinase, protease	Albershein (1966),
Fruit extraction	Amylase, amyloglucosidase, cellulase, pectinase, pentosanase, limonoate, dehydrogenase, naringinase, Anthocyanase	Albershein (1966), Kashyap et. al. (2001)

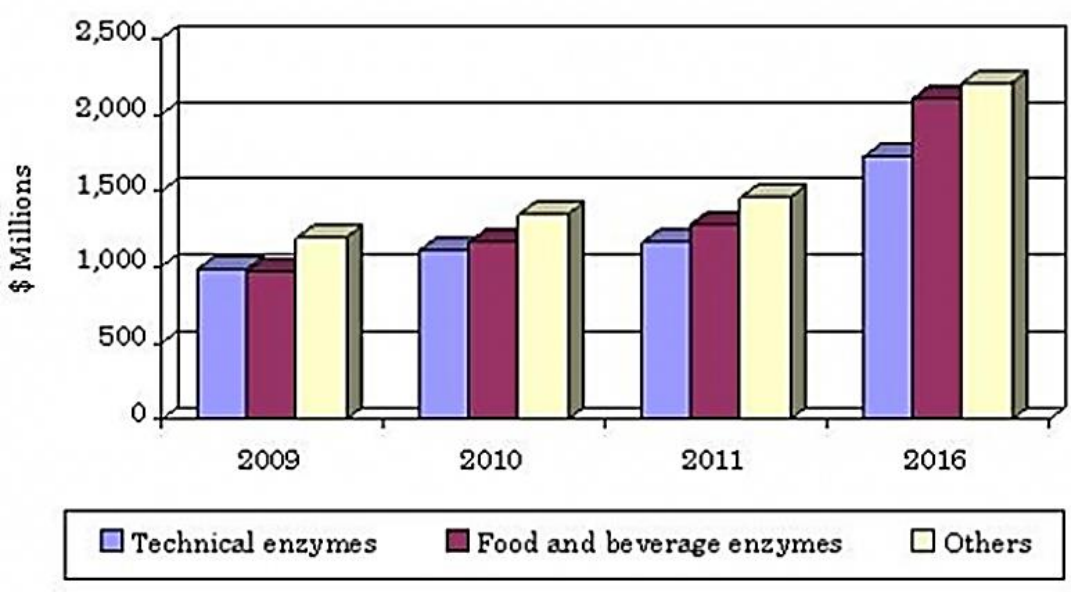

Figure 1: A summary of global revenue of industrial enzymes market, 2009-2016 (BCC, 2012)**.**

1.3.3.1 Enzymes in the food and beverage industry

Enzymes have many applications in the beverage industry *viz.*, pectinase. which is used in juice clarification (Brandelli *et al.*, 2005). Lipases are used in the production of fruit juices, baked foods, desirable flavours in cheeses, and in the esterification of fats and oils to produce modified acylglycerols (Saxena *et al.*, 1999). The starch industry became one of the major uses of microbial enzymes, invariably fungal enzymes. The chemical acid hydrolysis of starch was substituted by α-amylases and glucoamylases, which converts starch with over 90% yield to

glucose (Kirk *et al.*, 2002). Later, the starch industry was overtaken by the detergent industry in the total value of enzymes used. Rennin is used in cheese-making to coagulate milk proteins. β-galactosidase or lactase, which splits the milk-sugar lactose into glucose and galactose is used to produce milk products that are consumed by lactose-intolerant individuals. Other enzymes are also used in fruit juice manufacturing, in addition to pectinases. Xylanase and cellulase enhances the removal of pulp from the juice. Also, xylanase have been used for clarifying juices and for liquefying fruits and vegetables. Xylanases, have also been used in baking industry to improve the quality of bread and increase bread volume in combination with other enzyme *viz*., pectinases and amylases (Al-Reedy *et al.*, 2012; Amaro-Reyes *et al.*, 2010; Archer, 2000; Kirk *et al.*, 2002; la Grange *et al.*, 2010; Lange, 2010; Maheshwari *et al.*, 2000; Niehaus *et al.*, 1999; Puchart *et al.*, 1999; Salar and Aneja, 2007).

1.3.3.2 Biopulping

The current established wood pulping processes are relatively inefficient and environmentally costly due to the toxic by-products generated during chemical bleaching with chlorine. Thus, new pulping processes are needed. Enzymes continue to play a role in such developments. Bio-bleaching involves using microbial enzymes to bleach pulp (Bissoon *et al*., 2002; Madlala *et al*., 2001; Roncero *et al*., 2000). It relies on the ability of some microorganisms to depolymerize lignin directly and uses microorganisms or enzymes that attack hemicellulose, which facilitates subsequent de-polymerization. Enzymes, specifically xylanases, are used in the pre-bleaching of kraft pulps. This is due to xylanases attacking the hemicellulose component of pulp, thus facilitating the removal of lignin-associated hemicellulosic parts, without damaging the cellulose which makes up the paper structures. This process is not as harsh as the chemical process and more cost effective (Manimaran *et al.*, 2009; Purkarthofer *et al.*, 1993). But the most attractive feature of this enzyme-catalyzed process is that, it is less toxic than conventional chemical treatment. Besides just bleaching the pulp, the enzyme application also increases pulp fibrillation, reduces beating times in the original pulp and increasing the elasticity in recycled fibres. Xylanase treatment leads to a significant reduction in the amount of chemicals required

for bleaching and, subsequently, in the decreased quantities of toxic chlorine compounds released into the environment (Roncero *et al.*, 2000).

1.3.3.3 Animal feed

Enzymes have been used to treat wheat-based diets of animals in order to improve nutrient absorption. Xylanase enzymes have become one of the leading enzymes for this application and have become important in feed formulation. Xylanases are enzymes that hydrolyze the backbone of the plant polymer xylan which is second only to cellulose in abundance in nature (Singh *et al.*, 2000a). Usually a feed enzyme preparation is a multi-enzyme cocktail containing glucanases, xylanases, proteinases and amylases (Kim *et al.*, 2008b; Kunamneni *et al.*, 2005). The benefits of using enzymes in feed are (i) lowering feed viscosity; (ii) enhancing absorption of nutrients by liberating nutrients after hydrolysis of non-degradable fibres or blocked by these fibres; and (iii) reducing the amount of fecal waste produced by the animals. Treating feeds with enzymes may also improve digestibility by direct hydrolysis, improvement in palatability, change gut viscosity, and complementary actions of ruminal enzymes. Most animals that benefit from enzyme usage are non-ruminants, e.g., pigs, chickens and turkey (Pariza and Cook, 2010). An example of a recent addition of enzymes in the animal feed is the use of a phytase in monogastrics to increase the utilization of dietary phosphorus, which reduces the need for phosphorus supplements. In ruminants, the addition of phytase has improved the quality of the feed, with studies demonstrating that supplementing cattle diets with fibre-degrading enzymes can improve feed utilization and animal performance. Enzyme suppplements used for animal feed are largely produced through fungal fermentations that have these specific enzymatic activities (Beauchemin *et al.*, 1996).

1.3.3.4. Recombinant enzymes

Recombinant technology has influenced the field of biotechnology and all related processess profoundly. Recombinant fungal proteins have become the main source of enzymes for industrial applications some of which are listed in **Table 2**. It is estimated that over 60% of the enzymes used in the detergent, food, starch and other processing industries are recombinant products due

to the low yields obtained with natural hosts. Many recombinant food-grade proteins are of fungal origin, although the actual number of recombinant enzymes approved for food applications is not very large due to food safety issues (Archer, 2000; Rao *et al.*, 1998; Ward, 2012). Another recombinant enzyme that used is lipase with three fungal recombinant lipases currently being used in the food industry, produced from *Rhizomucor miehei*, *T. lanuginosus* and *Fusarium oxysporum*, all expressed in *A. oryzae* (Kirk *et al.*, 2002). The application of hydrolytic lipases in laundry detergents forms a major fraction of commercial enzymes sales, with of over 30% of the total sales being lipases (Ghorai *et al.*, 2009).

Table 2: Recombinant enzymes used in the food and feed industries adapted from (Ghorai *et al.*, 2009)

Enzyme	Expression host	Donor
Catalase	*A. niger*	*Aspergillus* sp.
Cellulase	*A. oryzae*	*Humicola* sp.
Cellulase	*T. reesei*	*Trichoderma* sp.
β-galactosidase	*A. oryzae*	*Aspergillus* sp.
β-glucanase	*T. reesei*	*Trichoderma* sp.
Glucose oxidase	*A. niger*	*Aspergillus* sp.
Lipase	*A. oryzae*	*Candida* sp., *Rhizomucor* sp., *Thermomyces* sp.
Phytase	*A. niger, A. oryzae*	*Aspergillus* sp.
	A. niger, A. oryzae, T. reesei	*Aspergillus* sp., *Aspergillus sp., Thermomyces* sp., *Trichoderma* sp.
Chymosin	*A. niger*	Calf
Protease	*A. oryzae*	*Rhizomucor* sp.

1.3.4 Biofuels

Depletion of easily accessible fossil energy resources in combination with recent concerns of global climate change and other environmental problems have led to efforts to decrease the use of fossil fuel and increase renewable energy sources such as bioethanol and biodiesel. Currently, biofuels are mostly produced from food crops which triggers an increase in the production cost as these feedstocks are expensive, with corn and sugarcane being the most prominent feed stocks used (Dodd and Cann, 2009). The BP Statistical Review of World Energy (2012) reported that the cost of petroleum based-fuel is likely to continue increasing. This also has an adverse effect on food prices and the fight on poverty, as food crops must now be shared between people and

biofuel production (Heaton *et al.*, 2008). Therefore, biofuel production needs to be increased and alternative feedstocks for biofuel production needs to be explored. There are very few countries that contribute significantly to the production of biofuel with seven countries making up of 90% of the world supply (**Figure 2**). While the demand for energy in the world is measured around 40 billion litres for 2011, biofuel contributes just over 50 million litres (**Figure 3**). Thus, alternative and new approaches are necessary for improvement of production rates.

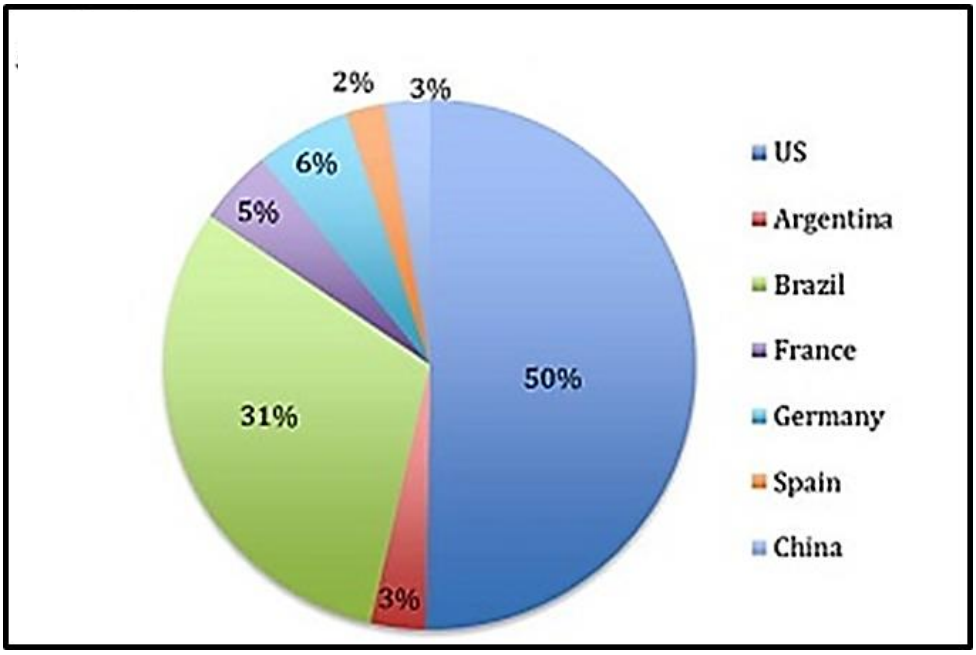

Figure 2: Major producers of biofuels (BP Statistical Review of World Energy, 2012)**.**

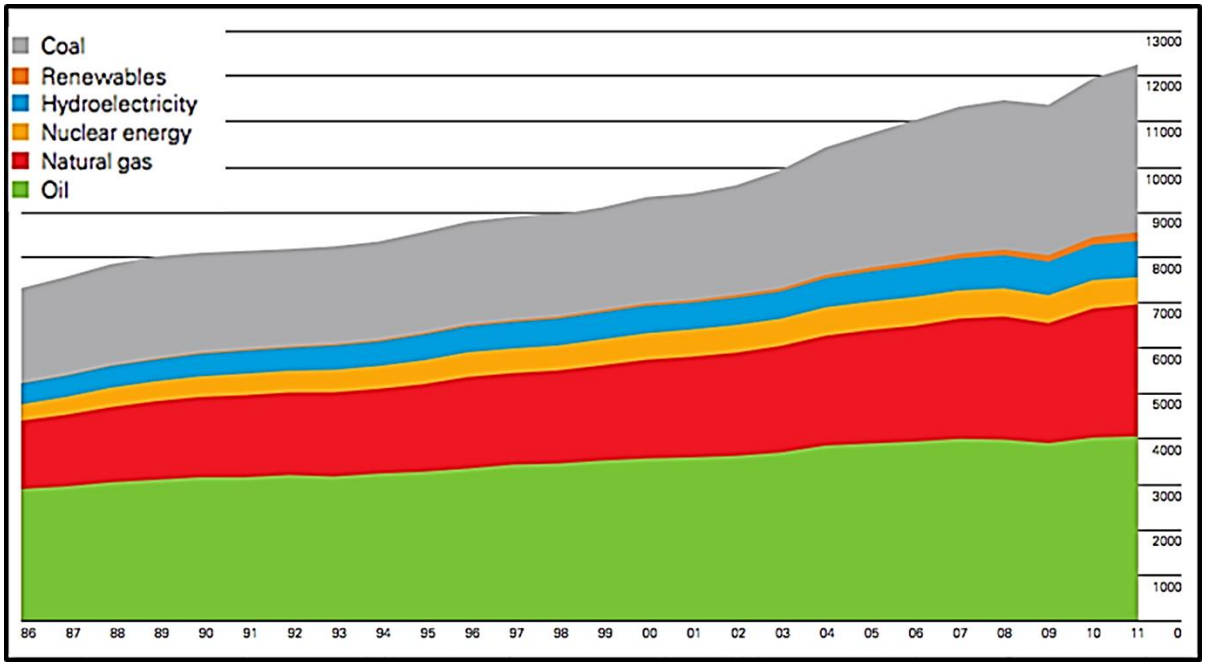

Figure 3: World energy consumption (BP Statistical Review of World Energy, 2012).

Lignocellulosic biomass is one renewable energy resource that can provide long term sustainability in the biofuel industry and is the most under-utilized (Taherzadeh and Karimi, 2008). The biggest difficulty in application of this biomass is the degradation of the plant polymers xylan and cellulose to monosaccharides which are required by the yeast for conversion to ethanol. In biofuel production, the degradation of complex biomass polymers into simple sugars is accomplished by the addition of enzymes (Banerjee *et al.*, 2010; de Oliveira and de Graaff, 2011; Mach and Zeilinger, 2003). These enzymes are produced from different microorganisms and the most widely used are enzymes from filamentous fungi. In their natural environment, fungi encounter a wide variety of polysaccharides, such as starch and plant cell wall polysaccharides like cellulose, hemicellulose and pectin. This requires an adaptation of these organisms, which is shown by filamentous fungi, to have a wide range of degradative enzymes which they use in their natural environment. The most widely used filamentous fungus or its cellulases in the biofuel industry is a mesophilic ascomycete called *Trichoderma reesei* (Martinez *et al.*, 2008). However, the production of cellulases is expensive making its use in the bioethanol production relatively more expensive than petroleum-based fuel. Therefore,

production processes can be improved to reduce cost but other enzymes or organisms can be investigated for ethanol production.

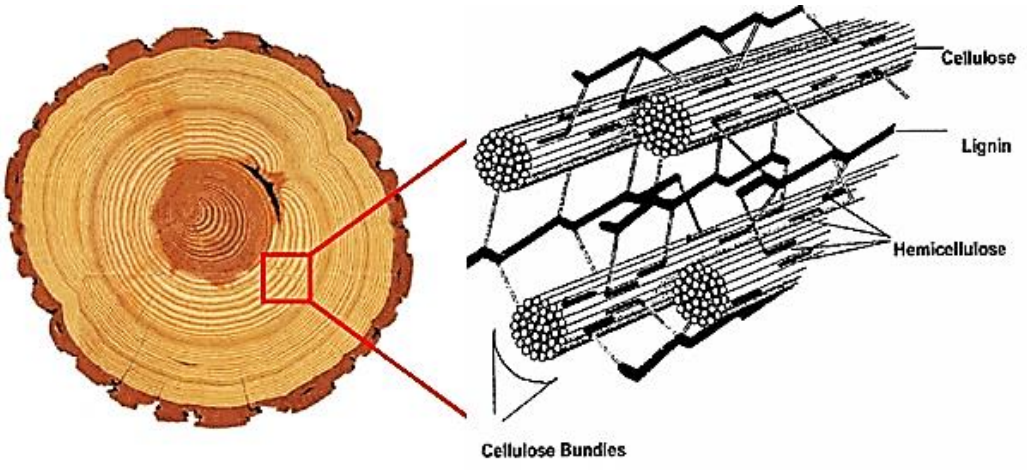

Figure 4: Diagram showing the major components (cellulose, hemicellulose and lignin) found in wood fibres (adapted from Stephens *et al.*, 2009).

Xylanase in synergy with several other enzymes, such as mannanase, ligninase, xylosidase, glucanase, glucosidase, and other enzymes can be used for the generation of biological fuels, such as ethanol from lignocellulosic biomass (Chen *et al.*, 2010; Seiboth *et al.*, 2003). The biological process of ethanol fuel production requires the removal of lignin to release cellulose and hemicellulose, followed by hydrolysis of the carbohydrate polymers (cellulose and hemicellulose) to produce free sugars, and finally, fermentation of mixed pentose and hexose sugars to produce ethanol. Thermophilic microorganisms and their degrative enzymes are ideal candidates for the development of more active, cost-effective enzymes for lignocellulose processing. Elevated operating temperatures would also be beneficial in fermentations to produce biofuels. In addition to lower risk of microbial contamination, a higher temperature would reduce cooling costs and facilitate ethanol removal and recovery (Nigam and Singh, 2011; Puri *et al.*, 2012).

1.3.5 Filamentous fungi as plant pathogens

Most fungal organisms are generally non-pathogenic, but a few are parasitic and can cause devastating plant infections, human infection and some produce deadly and carcinogenic toxins

(Dufresne *et al.*, 2011; James, 2006; Joubert *et al.*, 2011; Kronstad *et al.*, 2011; Ma *et al.*, 2010; Nierman *et al.*, 2005; Rep and Kistler, 2010). Plant pathogens, which are usually filamentous fungi such as rusts and smuts can destroy crops, especially cereals such as wheat and corn (Cristiano *et al.*, 2011; Goodwin *et al.*, 2011). Feeding the 7 billion people that inhabit the planet is the most important of all the global challenges that is increased by plant pathogens. This is because fungal plant pathogens increase post-harvest loss thus making food supply difficult and leading to an increase in food prices (Agrios, 2009; Fox, 1997). Filamentous fungi also affect food crops through production of mycotoxins making the harvest unfit for both human food and animal feed (Xu *et al.*, 2007).

1.3.6 Pharmaceuticals production

Fungi are also beneficial in medical applications. This is because filamentous fungi have the ability to produce antibiotics in large quantities. The first among these antibiotics is penicillin, discovered by Alexander Fleming in 1928 (Sternbach and Varon, 1992). This was isolated from the ascomycete fungus *Penicillium notatum*. Penicillin was later mass-produced commercially for medical use and its use was encouraged by the necessity to cure wounded soldiers of infections during the second world war. Streptomycin was isolated from an actinomycete, by Selman Waksman in 1944, when the era of antimicrobial compounds really began, and during the next two decades, more than 1000 antibiotics were discovered, many of them from actinomycetes and other fungi (Doull and Vining, 1990; Sternbach and Varon, 1992).

Today, important antibiotics derived from fungi, other than penicillin, include cephalosporins from *Cephalosporium* species, griseofulvin from *Penicillium griseofulvum,* lentinan from *Lentinus* species and schizophyllan from *Schizophyllum commune*. Penicillin and cephalosporin are known antibiotics used for the treatment of Gram-positive bacterial infections. Griseofulvin is an antifungal antibiotic useful in treating dermatophyte infections. Lentinan is used against *Mycobacterium tuberculosis, Listeria* species and Herpes Simplex Virus-1 (HSV-1). Schizophyllan has both antibacterial and antifungal activity and is used against *Candida albicans* and *Staphylococcus aureus* infections (Wainwright, 2008). But an increase in the number of immuno-compromised people has given rise to more human infections by opportunistic

pathogens. This not only increases the frequency of patients attacked by fungal infections, but also result in a broader spectrum of infectious fungal species. However, the biggest problem with fungal diseases is not the infections but the treatment because fungal infections are hard to cure because fungal cells are closely related to human cells. These similarities and relatedness between human beings and fungi make antifungal compounds, which have potential to eradicate the fungal pathogens, also toxic to human cells thus making their application difficult and allowing the fungal diseases to persist (Latgé and Calderone, 2002).

There are other important therapeutic compounds which are not antimicrobial agents obtained from fungi that have been used for medical purposes. Cyclosporin produced by *Tolypocladium inflatum,* is used in organ transplantation surgery as a immunosuppressant (Avramoff *et al.*, 2012). Lovastatin is a cholesterol biosynthesis inhibitor isolated from *Aspergillus terreus* used as a cholesterol reducing agent (Pecyna and Bizukojc, 2011). A similar cholesterol reducing drug is produced from *Penicillium citrinum*, and it is called Pravastatin (Kazumi *et al.*, 1995). Fungi are also a source of vitamin B12 (*S. cerevisiae*) and other vitamins (*S. cerevisiae, Ashbya* sp., *Blakeslea* sp.) and steroids used in fertility regulation (*Rhizopus* spp.) (Lamoolphak *et al.*, 2006). About 20% of the drugs produced by pharmaceutical industry in 2009 were derived from fungi (Masurekar, 2009).

1.4 THERMOPHILES

The evolution theory suggests that prokaryotic extremophiles were the first form of life on earth and that all life evolved from these organisms. Their ability to adapt is said to have produced a wide variety of microorganisms due to different extreme environments. Thermophilic organisms (bacteria, archaea, and eukarya) are found in various kinds of environments which were control by several stress-related factors *viz.*, temperature, pH level, oxygen level, osmolarity, ion balance, water activity, pressure, biological variables, relationship with other microorganisms (competition, predation, parasite, nutrient availability), presence of toxic metals, and radiation levels (Kirk *et al.*, 2002; Niehaus *et al.*, 1999). These variables usually work in combination with each other, exposing the organisms to stressful environments which produced different sub-groups of extremophiles, e.g., halophiles, thermophiles, psychrophiles, acidophiles, alkalophiles,

barophiles etc. In order to survive these harsh environments, these microorganisms must possess mechanisms or strategies for survival. These mechanisms can either be resistance or capacity adaptation or both based on avoiding or surviving the stress until conditions return to normal. On the other hand, extremophiles have adapted themselves not only to survive to but to thrive in harsh conditions by adapting cell components such as enzymes and membranes to work in these extreme condition (Cowan, 2004).

A group of extremophiles that has been attracting a huge interest, due to its proteins being seen as a viable substitute in most industrial processes are thermophiles. Like other extreme environments, only microorganisms are found living in extreme high temperature environments. An example of such environments are deep sea hydrothermal vents, where microorganisms thrive in temperature above 100°C at high pressure. Thermophilic organisms are divided into three groups, classified by using the range of temperature where active growth can be detected. The first group is made up of moderately thermophilic microorganisms with optimum growth at temperatures between 45°C to 60°C. The second group, classified as thermophilic has their optimum growth temperature between 60°C and 70°C while the last group has optimum growth temperature above 80°C called hyperthermophiles. Although research into extremophiles is thought to have begun in the late 19th century, the discovery of *Thermus aquaticus* in 1969 put new emphasis in this field (Olędzka *et al.*, 2003). This has led to the discovery and the identification of a significant amount of thermophilic organisms including some eukaryotic species. In the thermophile group, *Georgemma barossii* has the highest recorded maximum temperature for growth at 121°C and can survive 130°C (Kashefi, 2003) .

In general, prokaryotic organisms have been shown to have a large number of population that can survive in extreme environments compared to eukaryotes due to their cell structure. This could be due to the fact that prokaryotes have the capacity and rate of adaptation by alteration to membrane, protein, DNA, and repair and protection mechanism of the cell thus survive better in thermophilic environments (Vieille and Zeikus, 2001). Nevertheless, there are eukaryotic organisms that are thermophilic and all belong to the fungal kingdom (**Table 3**). The most thermophilic fungal organisms to have been reported grow in a range of between 50-60°C (Luke

et al., 2007; Maheshwari *et al.*, 2000; Morgenstern *et al.*, 2012; Salar and Aneja, 2007; Singh *et al.*, 2003; Vieille and Zeikus, 2001).

Table 3: Examples of thermophilic fungi (Salar and Aneja, 2007; Singh *et al.*, 2003)

Fungus	**Optimum growth temperature (°C)**	**Maximum growth temperature (°C)**
Canariomyces thermophila	45	50
Chaetomium mesopotamicum	45	52
Chaetomium thermophilum	45–55	58–61
Coonemeria aegyptiaca	40	55
Coonemeria crustacea	40	60
Coonemeria verrucosa	30–40	55
Corynascus thermophilus	50	60
Dactylomyces thermophilus	40–45	50
Malbranchea cinnamomea	45	57
Melanocarpus albomyces	45	57
Melanocarpus thermophilus	35	50
Myceliophthora hinnulea	40–45	50
Myceliophthora thermophila	45–50	55
Myriococcum thermophilum	45 50	53
Rhizomucor miehei	35–45	57
Rhizomucor pusillus	35–45	55
Scytalidium thermophilum	40	58
Stilbella thermophila	35–50	55
Talaromyces byssochlamydioides	40–45	>50
Talaromyces emersonii	40–45	55
Talaromyces thermophilus	45–50	60
Thermoascus aurantiacus	49–52	61
Thermomyces ibadanensis	42–47	61
Thermomyces lanuginosus	45–50	60
Thermomyces stellatus	40	50
Thielavia australiensis	35–40	50
Thielavia pingtungia	40	50
Thielavia terrestris	40–45	52

Most of these thermophilic fungi are found in different types of composting plant materials (Gazi *et al.*, 2007). In these environments, fungi can produce a variety of hydrolyzing enzymes which make them a potential source of thermophilic enzymes with industrial potential (Maheshwari *et al.*, 2000).

1.5 *THERMOMYCES LANUGINOSUS*

T. lanuginosus is a thermophilic fungus (**Figure 5 and 6**) that thrives at temperatures of up to 60°C, the upper temperature limit for growth of true fungi and fails to grow below 20°C. This fungus belongs to a group of filamentous fungi known as deuteromycetes. *T. lanuginosus* is found almost around the world due to the abundance of its natural habitat which is self-heating plant waste materials (Singh *et al.*, 2003). *T. lanuginosus* has an exceptional ability to produce extremely high levels of thermostable cellulase-free hemicellulases, most notably of which is xylanase (Singh et al., 2000a). Although differences in the production of xylanase by *T. lanuginosus* strains have been observed, it still remains the best producer of thermostable xylanase reported in literature (Bennett *et al.*, 1998; Puchart *et al.*, 1999; Singh *et al.*, 2000b; Singh *et al.*, 2000a). The strain that has been reported to produce the xylanase highest activity is *T. lanuginosus* SSBP isolated by Singh et al. (2000a) and has been used in different application including pulp and paper industry (Bissoon *et al.*, 2002; Christopher *et al.*, 2005; Kunamneni *et al.*, 2005; Mchunu *et al.*, 2009). But besides the xylanase, *T. lanuginosus* produces a host of other thermophilic enzymes, i.e., amylase, glucoamylase, xylanase, lipase, phytase, protease and chitinase (**Table 4**).

Figure 5: *T. lanuginosus* grown on potato dextrose agar after 7 days at 50°C (Singh *et al.*, 2000a).

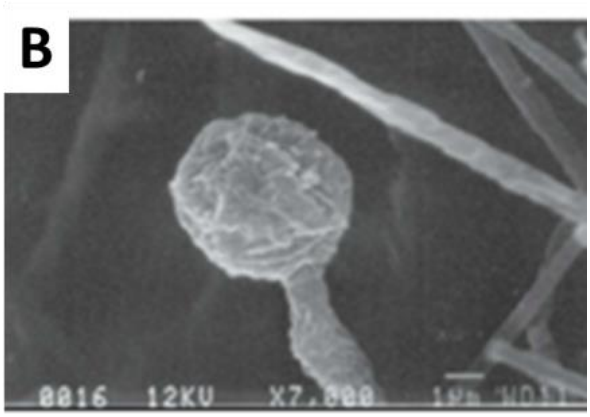

Figure 6: *T. lanuginosus*'s conidia under transmission electron microscopy (TEM) (Singh *et al.*, 2000a).

Thermophilic fungi are a rich source of thermostable industrial enzymes (Maheshwari et al. 2000) due to their ability to mass produce and secrete these proteins. Thermal stability is an attractive feature for many biotechnological applications including biomass degradation. Furthermore, fungi have a more complex system for degradation of lignocellulosic biomass which contains more than one type of substrate thus making them more attractive than bacteria. Due to the ability *T. lanuginosus* to produce thermostable proteins/enzymes it has become a target for industrially-applicable enzymes.

Table 4: Potential industrial proteins from *T. lanuginosus (*Puchart *et al.*, 1999; Singh *et al.*, 2000b)

Enzyme	**Temperature optimum (°C)**
Xylanase	70
Xylosidase	50
Acetyl xylan esterase	60
Acetyl esterase	60
Arabinofuranosidase	50
Coumaroyl esterase	40
Feruloyl esterase	50
Glucosidase	60
Lipase	50
Mannanase	60
Mannosidase	80

1.6 WHOLE GENOME SEQUENCING PLATFORMS

Advances in DNA sequencing have revolutionized the biological sciences, ushering the new field of genomics. Although data processing and sample numbers were bottlenecks previously, it is now possible to generate large amounts of sequence data very rapidly and at low cost (Baker *et al.*, 2008). Using current high-throughput sequencing instruments, the amount of sequencing data produced by the human genome project over 13 years for example, can now be produced within weeks and for a price of a few tens of thousands of dollars instead of millions of dollars (Bonetta, 2010). Reducing sequencing costs means DNA sequencing is now available to many more researchers and projects. Sequencing applications are wide and varied and include whole genome sequencing, measuring population and species variation, determining transcriptome organization, DNA microarrays, determining DNA methylation and unveiling DNA-protein interactions (Baker, 2012; de Oliveira and de Graaff, 2011; Galagan *et al.*, 2005b).

With genomics, it is now possible to apply these different sequencing-based approaches for comparative genomic studies. The high resolution of sequencing approaches allows studies of species and populations that have not been accessible so far due to the small evolutionary differences to be analyzed (Francis *et al.*, 2010; Hane *et al.*, 2011; Littlejohn *et al.*, 2012). Furthermore, these high-throughput sequencing technologies allow studies of ancient DNA samples with a low fraction of endogenous DNA to be completed. Where previously these studies would have been too expensive when using traditional sequencing approaches (Kirsanow and Burger, 2012). This technology also provides insight into the basis of diseases or developmental problems that affect uniquely human traits, such as speech and mental disorders (Dorrell *et al.*, 2002).

But besides human genome sequencing, the application of this technology in microbes has provided important information on how microorganisms function and are organized. There are many strategies used in genome sequencing and the most widely used are listed **Table 5**. These genomic sequencing platforms include, (i) pyrosequencing approach developed by Roche known as 454 Genome Sequencing, (ii) the reversible terminator technology used by Illumina Genome Technologies, and (iii) SOLiD sequencing platform using ligation-based principle (DiGuistini *et al.*, 2009; Kumar and Blaxter, 2010; Xuan *et al.*, 2012). These three platforms the Roche/454 FLX, the Illumina/Solexa Genome Analyzer, and the Applied Biosystems (ABI) SOLiD Analyzer are currently the most widely used technologies. Although the Polonator G 007 and the Helicos HeliScope have recently been introduced, but they have not become popular among the research community (Zhang *et al.*, 2011).

Table 5: Platforms and the detailed information on different next generation sequencing (NGS) technologies (Zhang *et al.*, 2011)

Technology	Read length	Throughput	Sequence by synthesis
Roche/GS FLX Titanium	400-600 bp	500 Mbp/run	Pyrosequencing
Illumina	2 x 100 bp	200 Gbp/run	Reversible terminators
ABI/SOLiD	50-100 bp	>100 Gbp/run	Sequencing-by-ligation (octamers)
Polonator/G.007	26 bp	8-10 Gbp/run	Sequencing-by-ligation (monomers)
Helicos/Heliscope	35 (25-55) bp	21-37 Gbp/run	True single-molecule sequencing
In development			
Pacific BioSciences/RS	1000 bp	N/A	Single-molecule real time (SMRT)
Visigen Biotechnologies	>100 bp	N/A	Base-specific FRET
U.S. Genomics	N/A	N/A	Single-molecule mapping
Genovoxx	N/A	N/A	Single-molecule sequencing by synthesis
Oxford Nanopore Technologies	35 bp	N/A	Nanopores/exonuclease-coupled
NABsys	N/A	N/A	Nanopores
Electronic BioSciences	N/A	N/A	Nanopores
BioNanomatrix/NanoAnalyzer	400 bp	N/A	Nanochannel arrays
GE Global Research	N/A	N/A	Closed complex/nanoparticle
IBM	N/A	N/A	Nanopores
LingVitae	N/A	N/A	Nanopores
Complete Genomics	70 bp	N/A	DNA nanoball arrays
base4innovation	N/A	N/A	Nanostructure arrays
CrackerBio	N/A	N/A	Nanowells
Reveo	N/A	N/A	Nano-knife edge
Intelligent BioSystems	N/A	N/A	Electronics
LightSpeed Genomics	N/A	N/A	Direct-read sequencing by EM
Halcyon Molecular	N/A	N/A	Direct-read sequencing by EM
ZS Genetics	N/A	N/A	Direct-read sequencing by TEM
Ion Torrent/PostLight	N/A	N/A	Semiconductor-based pH sequencing
Genizon BioSciences/CGA	N/A	N/A	Sequencing-by-hybridization

*N/A: Not available

1.6.1 Roche genome sequencer

The 454 Genome Sequencer (GS) platform was the first of the new high-throughput sequencing platforms on the market when released in 2005 (**Figure 7**). It is based on the pyrosequencing approach developed at the Royal Institute of Technology, Stockholm in 1996. In contrast to the Sanger sequencing technology which has been around longer, pyrosequencing is based on iteratively complementing single strands and simultaneously reading out the signal emitted from the nucleotide being incorporated also called sequencing by synthesis or sequencing during extension (Zhang *et al.*, 2011). Electrophoresis is therefore no longer required to generate an ordered read-out of the nucleotides as in the Sanger sequencing, as the read-out is done simultaneously with the sequence extension. In pyrosequencing, one nucleotide at a time is washed over several copies of the sequence to be determined, causing polymerases to incorporate the nucleotide if it is complementary to the template strand. The incorporation stops if the longest possible stretch of complementary nucleotides has been synthesized by the polymerase. In the process of incorporation, one pyrophosphate per nucleotide is released and converted to adenosine triphosphate (ATP) by an ATP sulfurylase. The ATP drives the light reaction of luciferases present and the emitted light signal is measured. To prevent the deoxyadenosine triphosphate (dATP) provided in a typical sequencing reaction from being used directly in the light reaction, deoxy-adenosine-5'-(alpha-thio)-triphosphate (dATPs), which is not a substrate of the luciferase, is used for the base incorporation reaction of adenine. Standard deoxyribose nucleotides are used for all other nucleotides. After capturing the light intensity, the remaining unincorporated nucleotides are washed away and the next nucleotide is provided (Gharizadeh *et al.*, 2004).

In 2005, pyrosequencing technology was developed on a picotitre plate by 454 Life Sciences (later bought by Roche Diagnostics) to allow high-throughput sequencing. The sequencing plate has about two million wells. The beads found in the wells are incubated with a polymerase and single-strand binding proteins, ATP sulfurylases and luciferases. Free nucleotides are then washed over the sequencing plate and the light emitted during the incorporation is captured for all wells in parallel using a high resolution CCD camera, exploiting the light-transporting features of the plate used. One of the adapters added is complementary to oligonucleotides on the

sequencing beads and thus allows molecules to be bound to the beads by hybridization. Low molecule-to-bead ratios and amplification from the hybridized double-stranded sequence on the beads makes it possible to grow beads with thousands of bound copies of a single starting molecule (Gharizadeh *et al.*, 2004). Using the second adapter, beads covered with molecules can be separated from empty beads. In the case of 454 sequencing errors can be caused by the following factors (Gharizadeh *et al.*, 2006; Gilles *et al.*, 2011);

(i) A reduction in enzyme efficiency or loss of enzymes which results in a reduction of the signal intensities,
(ii) Some molecules on the beads no longer being elongated and,
(iii) An increase phasing. Phasing is observed when a population of DNA molecules amplified from the same starting molecule (ensemble) is sequenced, and describes the process whereby not all molecules in the ensemble are extended in every cycle.

The current 454 Roche GS FLX Titanium platforms makes it possible to sequence about 1.5 million beads in a single experiment and to determine sequences of length between 300-500 nucleotides. The length of the reads is determined by the number of cycles, i.e. the number of times all four nucleotides have been washed over the plate, as well as by the base composition and the order of the bases in the sequence to be determined. Currently, 454 Roche limits this number to 200 cycles, resulting in an expected average read length of about 400 nucleotides. This is largely due to limitations imposed by the polymerases and luciferases which decrease over the sequencing run. In 2011, Roche released a new version of their sequencing chemistry allowing for sequencing of about two times longer reads. The platform allows the creation of about 750 Mb of DNA sequence per day at a cost of about $20/Mb (Zhang *et al.*, 2011). This has become the method of choice for non-model organisms that require *de novo* assembly because longer reads are more suitable to *de novo* assembling and annotation (DiGuistini *et al.*, 2009; Haridas *et al.*, 2011; Kumar and Blaxter, 2010; Links *et al.*, 2011; Nowrousian *et al.*, 2010; Scott *et al.*, 2009).

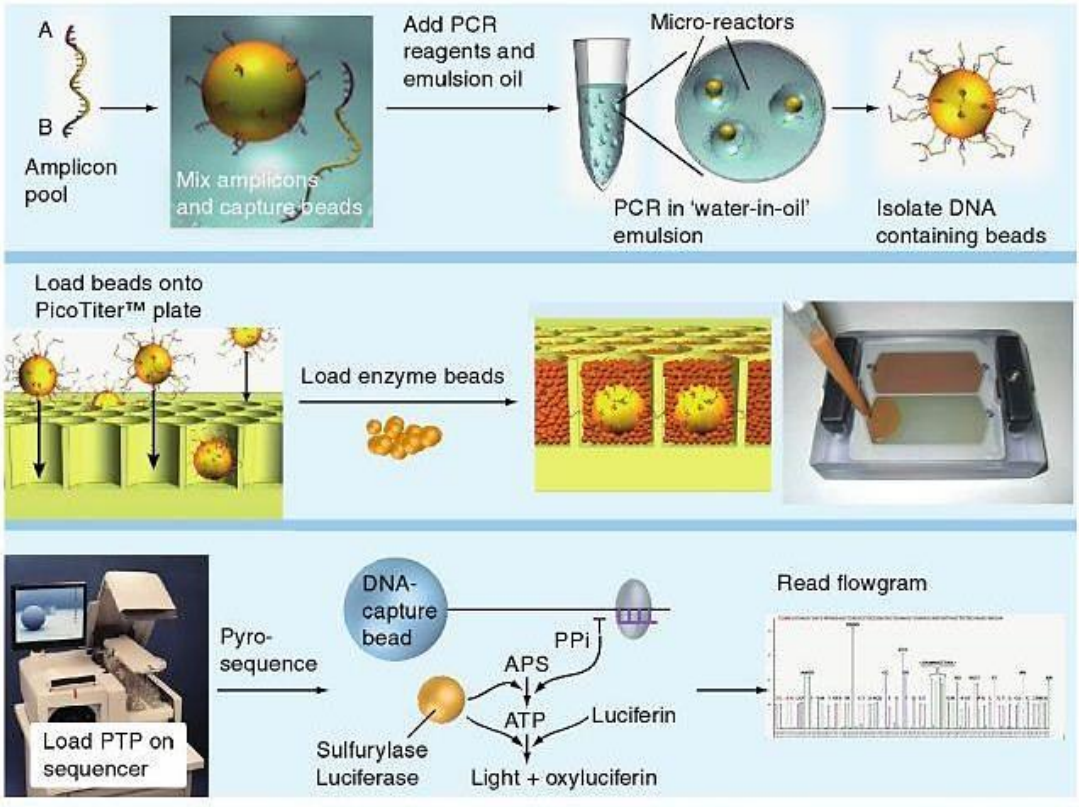

Figure 7: Overview of the Roche 454 sequencing system. DNA is sheared into small fragments to which adapters are ligated, fragments are attached to beads, mixed into an emulsion, amplified by emulsion PCR, deposited in wells of a picotitre plate, then sequences are determined by pyrosequencing (www.futuremedicine.com).

1.6.2 Illumina/Solexa genome analyzer

The reversible terminator technology used by the Illumina Genome Analyzer employs the sequencing by synthesis concept that is more similar to that used in Sanger sequencing than the 454 Roche technology, i.e. the incorporation reaction is stopped after each base, the label of the base incorporated is read out with fluorescent dyes and the sequencing reaction is then continued with the incorporation of the next base (**Figure 8**) (Zhang *et al.*, 2011). Similar to 454 Roche technologies, the Illumina sequencing process requires that samples for sequencing are converted into a sequencing library, which allows them to be amplified and immobilized for sequencing. For this purpose two different adapters are added to the 5' and 3' ends of all molecules using

ligation of so-called forked adapters. The library is then amplified using longer primer sequences which extend and further diversify the adapters, i.e., add further unique nucleotides at both adapter ends, to create the signal sequence needed in subsequent steps.

This double-stranded library is melted using sodium hydroxide to obtain single stranded DNAs, which are then pumped at a very low concentration through the channels of a flow cell. This flow cell has on its surface two populations of immobilized oligonucleotides complementary to the two different single stranded adapter ends of the sequencing library. These oligonucleotides are hybridized to the single stranded library molecules. By reverse strand synthesis starting from the hybridized (double-stranded) part, the new strand being created is covalently bound to the flow cell. If this new strand bends over and attaches to another oligonucleotide complementary to the second adapter sequence to the free end of the strand, it can be used to synthesize a second covalently-bound reverse strand. This process of bending and reverse strand synthesis, called bridge amplification, is repeated several times and creates what are termed clusters, the accumulation of several thousand copies of the original sequence in very close proximity to each other on the flow cell (Liu *et al.*, 2010).

These randomly distributed clusters contain molecules that represent the forward as well as reverse strands of the original sequences. Before determining the sequence, one of the strands has to be removed to prevent it from hindering the extension reaction sterically or by complementary base-pairing. Selective strand removal targets base modifications of the oligonucleotide populations on the flow cell. Following strand removal, each cluster on the flow cell consists of single stranded, identically oriented copies of the same sequence; which can be sequenced by hybridizing the sequencing primer onto the adapter sequences and starting the reversible terminator chemistry. Solexa sequencing was introduced in 2007, initially allowed for the simultaneous sequencing of several million very short sequences of about 20 nucleotides in a single experiment. In recent years, there have been several technical, chemical and software updates. The product, which is now called the Illumina Genome Analyzer (GA), has increased flow cell cluster densities to about 300 million clusters per run (Liu *et al.*, 2010) .

A paired-end module was also introduced enabling the sequencing of the reverse strand of each molecule (Gao *et al.*, 2013; Lassen *et al.*, 2012). This is achieved by chemical melting and washing away the synthesized sequence, repeating a few bridge amplification cycles for reverse strand synthesis and then selectively removing the starting strand before blocking 3' ends and annealing another sequencing primer for the second read (**Figure 8**). This paired-end sequencing approach produces double the amount of data can be generated by 454 sequencing. But the drawbacks are that the Illumina library and flow cell preparation includes several *in vitro* amplification steps which cause a high background error. Further, the flow cell preparation creates a fraction of ordinary looking clusters which are initiated from more than one individual sequence. This results in mixed signals and mostly low quality sequences for these clusters.

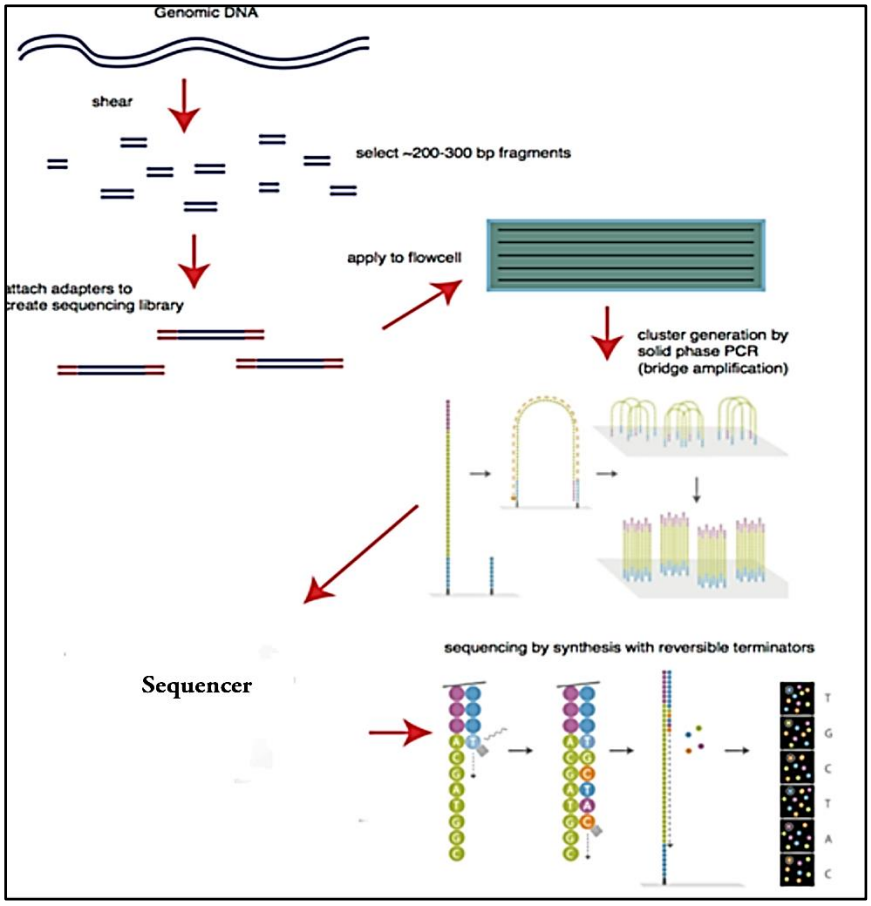

Figure 8: Overview of the Illumina/solexa platform strategy (Illumina Inc).

As is the case for the other platforms, the error rate increases with increasing position in the determined sequence (Minoche *et al.*, 2012). This is mainly due to phasing, which increases the background noise as sequencing progresses. While the ensemble sequencing process for pyrosequencing creates unidirectional phasing from lagging, non-extended molecules, reversible terminator sequencing creates bi-directional phasing as some incorporated nucleotides may also fail to be correctly terminated allowing the extension of the sequence by another nucleotide in the same cycle. The increasing cycle numbers causes the intensities extracted from the clusters to decline. This may be due to fewer molecules participating in the extension reaction as a result of non-reversible termination/DNA degradation, due to dimming effects of the sequencing fluorophores, or due to an increase in background noise by the accumulation of the sequencing fluorophores in the flow cell (Abnizova *et al.*, 2012). Even though the Illumina Genome Analyzer reads show a higher average error rate and are considerably shorter than 454 Roche reads, this technology output is more than 10,000 Mb per day with an estimated price of about $0.50/Mb. This is more than 12-fold higher daily throughput than 454 Roche and for a considerably 40-fold lower price per Mb (Bonetta, 2010; Zhang *et al.*, 2011). This has presumably led to the shutdown of 454 sequencing platform (www.genomeweb.com).

1.6.3 SOLiD sequencing platform

Life Technologies SOLiD was the prototype developed by George Church's laboratory at Harvard Medical School and the Howard Hughes Medical Institute in 2005 and was further developed and later sold by Applied Biosystems (ABI) and later by Life Technologies as the SOLiD sequencing platform, (Zhang *et al.*, 2011). After its release commercially in 2007, SOLiD was the third new high-throughput system entering a highly competitive market but the Church lab at Harvard Medical School continued the development of the system driving the cost down with a cheaper version of the system called Polonator in collaboration with Dover System. By 2008, companies like a biotechnology company from Mountain View, California, named Complete Genomics started to offer a human genome sequencing service. This technology is also based on the SOLiD concept, but combined with a new strategy of sequencing library construction and sequence immobilization using rolling circle amplification (DiGuistini *et al.*, 2010).

The SOLiD system is based on the principle of sequencing-by-ligation which is very different from 454 Roche and Illumina technologies. The sequence extension reaction is not carried out by polymerases but rather by ligases. In sequencing-by-ligation process, a sequencing primer is hybridized to single stranded copies of the library molecules to be sequenced. A mixture of probes carrying four distinct fluorescent labels that compete for ligation to the sequencing primer. The fluorophore encoding for the nucleotide, which is based on the two 3' most nucleotides of the probe, is read. Three nucleotide bases including the dye are cleaved from the 5' end of the probe, leaving a free 5'phosphate on the extended primer, which is then available for further ligation. After multiple ligations, the synthesized strands are melted and the ligation product is washed away before a new sequencing primer is annealed. Starting from the new sequencing primer, the ligation reaction is repeated. The same process is followed for three other primers, facilitating the read out of the dinucleotide encoding for each start position in the sequence. Using a specific fluorescent label encoding for specific nucleotide, the dye read outs (i.e. colours) can be converted to a sequence. This conversion from colour space to sequence requires a known first base, which is the last base of the library adapter sequence. Given a reference sequence, this encoding system allows for the detection of machine errors and the application of an error correction to reduce the average error rate. In the absence of a reference sequence, however, colour conversion fails with an error in the dye read out and causes the sequence downstream of the error to be incorrect. For parallelization, the sequencing process uses beads covered with multiple copies of the sequence to be determined. These beads are created in a similar fashion as in the 454 Roche platform. In contrast to the 454/Roche technology, the SOLiD system does not use a picotitre plate for fixation of the beads in the sequencing process; instead, the 3' ends of the sequences on the beads are covalently bound onto a glass slide. As for the Illumina system, this creates a random dispersion of the beads in the sequencing chamber and allows for higher loading densities (Lapidus, 2009).

However, random dispersion complicates the identification of bead positions from images and results in the possibility that chemical crystals, dust and lint particles can be misidentified as clusters. Sequencing errors are (Liu *et al.*, 2012; Miller *et al.*, 2013):

i) *In vitro* amplification steps cause higher background error rate than *in vivo* amplification used in the Sanger method,

ii) Beads carrying a mixture of sequences and beads in close proximity to one another create false reads and low quality bases,

ii) Signal decline and incomplete dye removal result in increasing error as the ligation cycles progress,

iv) Incomplete cleavage of the dyes may allow cleavage in the next ligation reaction, which then allows for the extension in the next reaction by one cycle. This causes a different phasing effect and additional noise from the previous cycle's dyes and in the dye identification process.

The SOLiD system currently allows sequencing of more than 300 million beads in parallel, with a typical read length of between 25 to 75 nucleotides. The ABI SOLiD system's throughput is comparable to the Illumina Genome Analyzer system, in terms of data produced and cost. Average error rates are dependent on the availability of a reference genome for error correction. In the absence of a reference genome, assembly and consensus calling may be performed based on dye read outs (so called colour space sequences) to reduce the errors before the conversion to the nucleotide sequence. If no reference genome is available for error corrections and no assembly and consensus calling is performed, the average error rate is higher than for the Illumina (Liu *et al.*, 2012). Illumina and SOLiD technologies are the methods of choice for model organism where a wealth of genomic data is available and new data can be mapped to reference assemblies.

1.6.4 Sequence assembly

In bioinformatics, sequence assembly refers to aligning and merging fragments or read sequences to a much longer DNA sequence, in order to reconstruct the original sequence. This is required as genome sequencing technology cannot read whole genomes, but rather reads small pieces of between 20 and 1000 bases, depending on the technology used (454 Roche about 300-400 nucleotides, Illumina about 100 nucleotide and SOLiD about 25-75 nucleotides). Typically, the short fragments, called reads, result from shotgun sequencing genomic DNA, or gene

transcripts (Expressed Sequence Tags). One of the most famous and effective comparisons to describe genome assembly is to compare the process to a book. Sequence assembly can be compared to taking many copies of a book, shredding them, then putting the book back together just by looking at the shredded pieces. Besides the obvious difficulty of this task, some extra problems can complicate the matter. The original may have many repeated paragraphs, "the book" may contain some typographical errors and also parts from other books may also be added thus, and some shreds may be completely unrecognizable.

To assemble a genome, assembly programs use data made of single and paired reads. Single reads are simply the short sequenced fragments themselves; they can be joined up through overlapping regions into a continuous sequence known as a "contig" (Gilles *et al.*, 2011). Repetitive sequences, missing data and mistakes limit the length of the contigs that assemblers can build. Paired reads typically are about the same length as single reads, but they come from either end of DNA fragments that are too long to be sequenced straight through (Fullwood *et al.*, 2009; Gao *et al.*, 2013). Depending on the preparation technique, that distance can be as short as 200 base pairs or as large as several tens of kilobases. Knowing that paired reads were generated from the same piece of DNA can help link contigs into 'scaffolds', ordered assemblies of contigs with gaps in between (**Figure 9**). Paired-read data can also indicate the size of repetitive regions and how far apart contigs are.

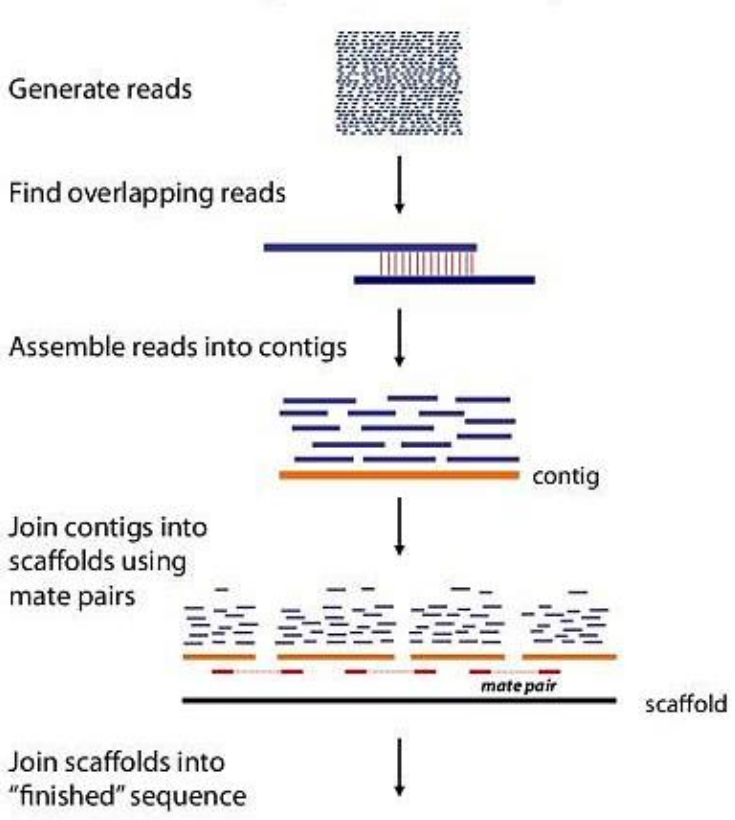

Figure 9: Schematic showing a basic genome assembly process (Wu *et al.*, 2012).

1.6.5 *De novo* assembly vs mapping assembly

The *de novo* assembly method attempts to re-create the genome without a reference genome. Tens of millions of DNA fragments are sequenced from both ends, and then assembled by computers looking for an overlapping matching sequence from different fragments. Large fragments and a range of different sizes, enable assembly to take place despite the presence of confusing repetitive sequences. Random positioning of fragments means that there are still gaps where the sequence is unknown, although the order of the sequenced sections relative to each other is known. The efficiency of *de novo* genome sequence assembly processes depends heavily on the length, fold coverage and per base accuracy of the sequence data (**Figure 10**).

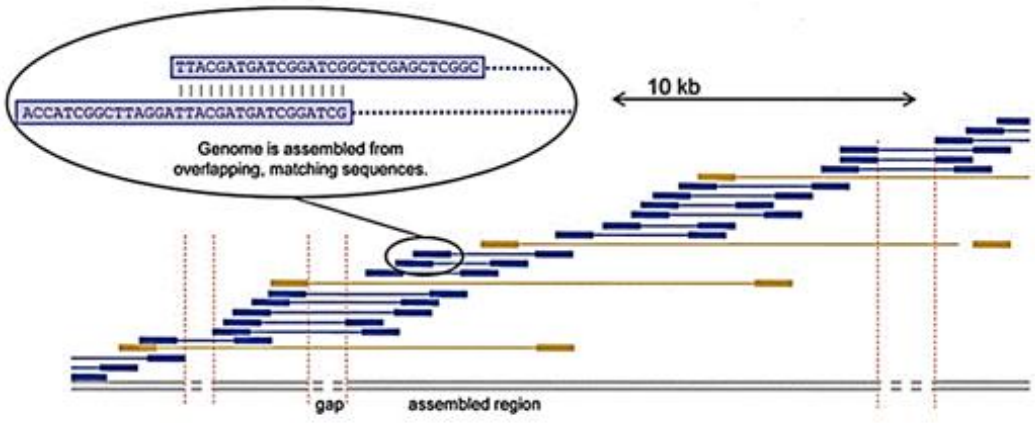

Figure 10: Schemtic showing *de no* assembly with overlapping regions (blue) and gaps (red dotted line where no ovelarpping sequences are found-www.nimr.mrc.ac.uk).

Mapping assembles rely on the existence reference or previous assembly of the same organism. The mapping is the process of comparing each one of the reads with the reference genome. The new assembly is based on alignments obtained between each read and the genome (**Figure 11**).

In terms of complexity and time requirements, *de novo* assemblies are orders of magnitude slower and more memory intensive than mapping assemblies. This is mostly due to the fact that the assembly algorithm needs to compare every read with every other read with reads measuring millions depending on the sequencing platform used. While for mapping assemblies, one would have a very similar book as a template, the *de novo* assemblies are more 'hardcore' in a sense as one would not know in advance whether this would become a science book, or a novel, or a catalogue etc.

Assessing quality of the assembly is made more difficult because sequencing technology changes so quickly. And the absence of a high quality reference genome, new genome assemblies are often evaluated on the basis of (i) the number of scaffolds and contigs required to represent the genome, (ii) the proportion of reads that can be assembled, (iii) the absolute length of contigs and scaffolds, and (iv) the length of contigs and scaffolds relative to the size of the genome. The

most commonly used metric is N50, the smallest scaffold or contig above which 50% of an assembly would be represented. But this metric may not accurately reflect the quality of an assembly.

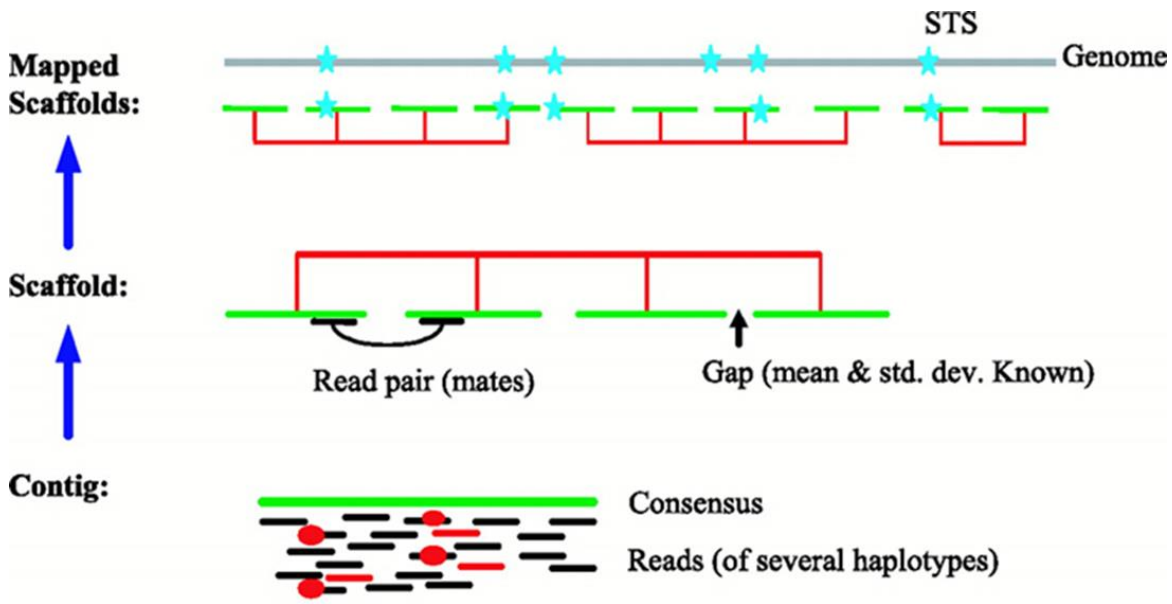

Figure 11: Schematic view of reference or mapping assembly method for genomes. In whole-genome assembly, the BAC fragments (red) and the reads (black) are combined to produce a contig and a consensus sequence (green). The contigs are connected into scaffolds, shown in red, by pairing end sequences, which are also called mates. If there is a gap between consecutive contigs, it has a known size. Next, the scaffolds are mapped to the genome (Venter *et al.*, 2001).

1.7 FUNCTIONAL GENOMICS

Functional genomics is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomic sequencing and annotation to describe gene function(s) and interactions. Unlike genomics, functional genomics focuses on the dynamic aspects such as gene transcription, translation, and protein–protein interactions, as opposed to the static aspects of the genomic information such as DNA sequence or structures. This technique attempts to answer questions about the function of DNA at the levels of genes, RNA transcripts, and protein

products (Doerks *et al.*, 2012; Giaever *et al.*, 2002; Morozova and Marra, 2008; Nikitin *et al.*, 2003; Stajich *et al.*, 2012; Tatusov *et al.*, 2000).

Functional genomics aim to understand the relationship between a gene and its phenotype. This can be done using many approaches to understand function and interaction of the entirety of an organism's genes and products including biochemical, cellular, physiological properties of each and every gene product. It also involves studies of natural variation in genes, RNA, and proteins over time as well as studies of natural or experimental functional disruptions (mutants) affecting genes, chromosomes, RNAs, or proteins (Dorsett and Tuschl, 2004). It enhances understanding of the synthesized genomic and proteomic knowledge into an understanding of the dynamic properties of an organism at cellular level. This would provide a more complete understanding of how biological function arises from the information encoded in an organism's genome (Conesa *et al.*, 2005). The possibility of understanding how a particular mutation leads to a given phenotype has important implications for human genetic diseases, as answering these questions provided base for treatment or cure. It can also improve our understanding of biological processes and improve our understanding of gene and protein functions and interactions thus providing more efficient industrial applications. But a key characteristic of functional genomics studies is their genome-wide approach to these questions, generally involving high-throughput methods rather than a more traditional single gene approach (Conesa *et al.*, 2005).

1.7.1 Transcriptome

The transcriptome is made of all RNA molecules, including mRNA, rRNA, tRNA, and other non-coding RNA produced by an organism or a population of cells. It can also be described as the total set of transcripts in an organism, or to the specific subset of transcripts present in a particular cell type (Morozova and Marra, 2008). Usually, the transcriptome analysis can also include the amount or concentration of each RNA molecule in addition to the molecular identification, unlike genome analysis which is just qualitative. Also, the main difference is that the genome may have minor changes due to mutations, whreas the transcriptome can vary depending on the environmental conditions that the organism is exposed to. The transcriptome represents the genes that are being expressed at that particular time (Baumann *et al.*, 2011;

DiGuistini *et al.*, 2010; Gao *et al.*, 2013; Wang *et al.*, 2010; Zuccaro *et al.*). The use of next-generation sequencing technology to study the transcriptomes at the nucleotide level is known as RNA-Seq (Abnizova *et al.*, 2012; Dufresne *et al.*, 2011; Kubicek *et al.*, 2011; Maccallum *et al.*, 2009). Like with genomes, transcriptomes can be assembled using a reference genome, either of the organism itself or a closely related species. A second method is *de novo* transcriptome assembly, using software to directly assemble short sequence reads obtained from all the mRNA, rRNA, tRNA, and other non-coding RNA (Gao *et al.*, 2013). The transcriptome can be used as the first step for proteome analysis, that is, the entire set of proteins expressed by a genome.

But the disadvantage for transcriptome analysis is that relatively small changes in mRNA expression can produce large changes in the total amount of the corresponding protein present in the cell. Another difficulty for RNA-based procedures is the emergence of small RNAs, which are challenging to analyze (Liu, 2005).

1.7.2 DNA microarrays

Microarrays measure the amount of mRNA in a sample that corresponds to a given gene or probe DNA sequence. Probe sequences are immobilized on a solid surface and allowed to hybridize with fluorescently-labeled mRNA. The intensity of fluorescence of a specific spot is proportional to the amount of target sequence that has hybridized to that spot, and therefore to the abundance of that mRNA sequence in the sample. Microarrays allow for identification of candidate genes involved in a given process based on variation between transcript levels for different conditions and shared expression patterns with genes of known function (Dorrell *et al.*, 2002).

Filamentous fungi are extensively used in industrial processes. Microarray analysis has been used to examine the expression of industrially important enzymes. An example was the analysis of genes responsible for biosynthesis of koji in *A. oryzae* (Dorrell *et al.*, 2002). Also, microarrays have been used to understand heterologous protein expression in filamentous fungi, where low yields are obtained compared to native protein. Thus, investigating the transcriptome (Carvalho *et al.*, 2011; de Oliveira and de Graaff, 2011; Prosad *et al.*, 2011) is important for improving efficiency of recombinant protein secretion (Ward, 2012).

Although microarray studies can reveal the relative amounts of different mRNAs in the cell, levels of mRNA are not directly proportional to the expression level of the proteins they code for. The number of protein molecules synthesized using a given mRNA molecule as a template is highly dependent on translation-initiation features of the mRNA sequence. In particular, the ability of the translation-initiation sequence is a key determinant in the recruiting of ribosomes for protein translation (Sonenberg and Hinnebusch, 2009).

1.7.3 Metabolomics

Metabolomics has emerged as another option to complement existing techniques and provide detailed information on metabolic regulation and secondary metabolism. Metabolomics is the scientific study of chemical processes involving metabolites or systematic study of the unique chemical fingerprints that specific cellular processes leave behind, the study of their small-molecule metabolite profiles (Allwood *et al.*, 2011). The metabolome represents the collection of all metabolites in a biological cell, tissue, organ or organism, which are the end products of cellular processes. Thus, while mRNA gene expression data and proteomic analyses do not tell the whole story of what might be happening in a cell, metabolic profiling can give an instantaneous view of the physiology of those cells. One of the challenges of systems biology and functional genomics is to integrate proteomic, transcriptomic, and metabolomic information to give a more complete picture of living organisms (Bino *et al.*, 2004).

1.7.4 Phenotypic microarray

Phenotype microarrays (Biolog, Inc.) represent the third major technology, alongside DNA microarrays and proteomic technologies that is needed in the genomic era of research and drug development. Phenotype microarray technology tests cellular phenotypes. Included in the tests are assays of basic cellular nutritional pathways for C, N, P, and S metabolism, osmotic and pH and sensitivity, and sensitivity to chemical agents (Bochner *et al.*, 2001).

The most common application of phenotypic microarrays is to assess the phenotypic effects of mutations. A change in genotype of a cell should lead to one or more changes in phenotype. Phenotypic microarray allows testing of deletion or insertion mutants to understand the biological changes that occur consequent to genetic changes. Another common application is phenotypic characterization of a collection of related strains. For example, it is possible to determine the phenotype relatedness of a collection of isolates of a given species. Phenotypic microarray analysis has been successfully implemented for a wide variety of model microbial cells including bacteria and fungi. Phenotype microarray analysis has been successfully implemented for a wide variety of model microbial cells including *Escherichia coli*, *Salmonella enterica*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, *Sinorhizobium meliloti*, *Saccharomyces cerevisiae*, *Bacillus subtilis*, *Bacillus cereus*, *Staphylococcus aureus*, *Proteus mirabilis*, and many other species (Ito *et al.*, 2005; Lea and Irina, 2010; von Eiff *et al.*, 2006).

1.8 METABOLIC RECONSTRUCTION

In response to the growing amount of sequence data, computational tools for genome analysis have been developed and merged into shared analytical environments, such as, KEGG enhancing cross-genome analysis tools. Computational approaches are based on various types of clustering of potential genes, such as clusters of orthologous genes or position on the chromosome, such as in operons. An important stage of genome analysis is the integration of gene annotation into functional pathways, which is the conceptual assembly of metabolic pathways, transport units and signal transduction pathways, thus leading to a more accurate assembly and gene annotation (Overbeek *et al.*, 2000). Metabolic reconstructions from genomes are already in place or under development for a growing number of organisms including eukaryotic, prokaryotic and archaeal species. However, although there is large number of genomes that have been fully sequenced, there very few specific genome that have their metabolic pathway constructed. Thus the metabolic reconstruction of T. *lanuginosus* will add invaluable information to the understanding of fungal and thermophilic organisms (Suthers *et al.*, 2009).

1.9 MOTIVATION FOR THE STUDY

The importance of filamentous fungi formimng part of a more sustainable solution to important global problems is apparent. Therefore, fungal research has to expand from the more traditional role of just identification and protein application. Thus, it has been rewarding to observe large-scale sequencing projects that have given priority to fungi; not only within genomics but also transcriptomics and proteomics. Many model fungi have been sequenced, including the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, the filamentous fungi, *Neurospora crassa* and *Aspergillus nidulans*, plant pathogens, *Magnaporthe grisea* and *Fusarium graminearum* and the industrialy- important filamentous fungus *Trichoderma reesei* to name a few (Grossetête *et al.*, 2010). These eukaryotes have been studied using tools of modern cell biology, genetics and biochemistry to elucidate not only details of their unique lifestyles, but also fundamental aspects of biology that they share with other organisms. However, numerous non-model organisms that belong to filamentous fungal group have not yet been analyzed, and only just a handful of genomes of thermphilic fungi including *Myceliophthora thermophila* and *Thielavia terrestris* have been sequenced (Berka *et al*., 2011). The number of filamentous fungal genome sequencing projects is considerably lower than that of bacterial genome projects but given their economic and scientific importance, fungal genomes deserve more attention.

T. lanuginosus is an organism with a potential for wide application in different industries. The ability of this organism to produce numerous amount of carbohydrate degrading-enzymes coupled with its thermophilicity makes it a perfect candidate for whole genome sequencing. Although there has been one online report of genome sequencing of this organism the data is not available for public usage. In addition, the strain (SSBP) that will be used in this study is the highest producer of xylanase protein which could suggest possible difference in the genetic make-up of these two strains (Singh *et al.*, 2000b). Also, this study will attempt to reconstruct the biological pathways that may exist in *T. lanuginosus*, based on the genome data. Some of the pathways related to carbohydrate-utilisation will be validated using phenotypic microarray which is designed specifically for fungi, a first for *T. lanuginosus*. This study aims to provide more understanding of thermostability in eukaryotes, identify novel genes and enhance understanding of metabolic pathways in this organism.

CHAPTER 2: GENOME ANALYSIS OF *T. lanuginosus*

This chapter was published in: Mchunu, Nokuthula Peace, Kugen Permaul, Ahmad Yamin Abdul Rahman, Jennifer A. Saito, Suren Singh, and Maqsudul Alam. (2013). Xylanase Superproducer: Genome Sequence of a Compost-Loving Thermophilic Fungus, *Thermomyces lanuginosus* Strain SSBP. Genome Announcements 1, no. 3.

genomeA
Journals.ASM.org

Xylanase Superproducer: Genome Sequence of a Compost-Loving Thermophilic Fungus, *Thermomyces lanuginosus* Strain SSBP

Nokuthula Peace Mchunu,[a,c] Kugen Permaul,[a] Ahmad Yamin Abdul Rahman,[c] Jennifer A. Saito,[b,c] Suren Singh,[a] Maqsudul Alam[b,c]

Durban University of Technology, Department of Biotechnology and Food Technology, Durban, South Africa[a]; Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii, Honolulu, Hawaii, USA[b]; Centre for Chemical Biology, Universiti Sains Malaysia, Bayan Lepas, Penang, Malaysia[c]

We report here the draft genome sequence of *Thermomyces lanuginosus* strain SSBP, which was isolated from soil in South Africa. This fungus produces the largest amount of xylanase ever reported in the literature.

Received 8 May 2013 **Accepted** 13 May 2013 **Published** 20 June 2013

Citation Mchunu NP, Permaul K, Abdul Rahman AY, Saito JA, Singh S, Alam M. 2013. Xylanase superproducer: genome sequence of a compost-loving thermophilic fungus, *Thermomyces lanuginosus* strain SSBP. Genome Announc. 1(3):e00388-13. doi:10.1128/genomeA.00388-13.

Address correspondence to Nokuthula Peace Mchunu, nokuthula@dut.ac.za.

Thermomyces lanuginosus is a thermophilic fungus that has the ability to degrade plant biomass and produces the largest amounts of hydrolyzing enzymes (1). Because of this ability, this fungus has been identified as one of the organisms that can have various industrial applications. Thus, sequencing of the *T. lanuginosus* genome is important because this will provide necessary information needed for better industrial usability of this fungus.

Whole-genome sequencing was performed using Roche 454 and Illumina paired-end sequencing strategies. The Roche 454 sequencing strategy with single ends and paired ends was used to construct a genomic library, and a Solexa sequencing strategy was used to construct a paired-end and mate-paired genomic library. We are reporting a 23.3-Mb *T. lanuginosus* genome sequence that was created by *de novo* assembling of 98% of the sequencing data generated using next-generation sequencing technology. The repeat reads were <0.5% of the total reads generated. The assembly has long-range continuity, with an N_{50} scaffold size of >4 Mb. Most of the assembly (90%) is contained in the 6 largest scaffolds and the rest in 24 smaller scaffolds with lengths of <10 kb.

The proteome that was predicted from this assembly had a total of 5,105 genes, which is less than that of other filamentous fungi, with 83 tRNA genes (2–6). The genome was annotated using a modified version of Maker (7). The GC content of the whole genome was calculated to be 52.14%, while coding regions had a significantly higher GC content, at 55.6%. As this thermophilic fungus is a wood-degrading fungus, the CAZy family protein was analyzed in depth. Carbohydrate-active enzymes (CAZy) have the ability to cleave or add monomers to polysaccharides, and because of this characteristic many of these fungi have major importance in the biotechnology industry. A total of 224 predicted proteins that were identified using the CAZy database belong to this group of proteins (8).

Mechanisms of thermal adaptation by this fungus, especially in relation to the genomic material, were suggested by the presence of several DNA-related pathways. *T. lanuginosus* has a ubiquitin degradation pathway which plays a crucial role in the responses to various stressors, such as nutrient limitation, heat shock, and heavy metal exposure (9). *T. lanuginosus* is a thermophilic organisms that generally grows on dead woody material. The ubiquitin degradation system may be essential for adaptation during rising temperatures in composting materials. *T. lanuginosus* is also capable of histone acetylation/deacetylation and poly ADP-ribosylation and contains high numbers of methylases. Histone acetylation/deacetylation and methylation play important roles in packing and condensation of DNA (10). Poly ADP-ribosylation is the addition of one or more ADP-ribose moieties to a protein (11). It plays an important role in cell signaling and the control of many cell processes, including DNA repair and apoptosis (12, 13). All of these DNA condensation machinery and repairing mechanisms make this fungus well adapted to living and thriving at high temperatures.

Nucleotide sequence accession number. This Whole-Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number ANHP00000000. The version described in this paper is the first version.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation (South Africa) and APEX funding (Malaysia Ministry of Higher Education) to the Centre for Chemical Biology, Universiti Sains Malaysia.

REFERENCES

1. **Singh S, Madlala AM, Prior BA.** 2003. *Thermomyces lanuginosus*: properties of strains and their hemicellulases. FEMS Microbiol. Rev. **27**:3–16.
2. **Andersen MR, Salazar MP, Schaap PJ, van de Vondervoort P, Culley D, Thykaer J, Frisvad CJ, Nielsen KF, Albang R, Albermann K, Berka RM, Braus GH, Braus-Stromeyer SA, Corrochano LM, Dai Z, van Dijck PW, Hofmann G, Lasure LL, Magnuson JK, Menke H, Meijer M, Meijer SL, Nielsen JB, Nielsen Ml, van Ooyen AJ, Pel HJ, Poulsen L, Samson RA, Stam H, Tsang A, van den Brink JM, Atkins A, Aerts A, Shapiro A, Pangilinan H, Salamov A, Lou Y, Lindquist E, Grimwood J, Grigoriev IV, Kubicek CP, Martinez D, van Peij NN, Roubos JA, Nielsen J, Baker SE.** 2011. Comparative genomics of citric-acid-producing *Aspergillus niger* ATCC 1015 versus enzyme-producing CBS 513.88. Genome Res. **21**: 885–897.
3. **Bartoszewska M, Kiel JA.** 2011. The role of macroautophagy in development of filamentous fungi. Antioxid. Redox Signal. **14**:2271–2287.
4. **DiGuistini S, Liao NY, Platt D, Robertson G, Seidel M, Chan SK, Docking TR, Birol I, Holt RA, Hirst M, Mardis E, Marra MA, Hamelin**

RC, Bohlmann J, Breuil C, Jones SJM. 2009. De novo genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. Genome Biol. **10:**R94.

5. **Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EG, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Barbote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS.** 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat. Biotechnol. **26:**553–560.
6. **van den Berg MA, Albang R, Albermann K, Badger JH, Daran JM, Driessen AJ, Garcia-Estrada C, Fedorova ND, Harris DM, Heijne WH, Joardar V, Kiel JA, Kovalchuk A, Martin JF, Nierman WC, Nijland JG, Pronk JT, Roubos JA, van der Klei IJ, van Peij NN, Veenhuis M, von Döhren H, Wagner C, Wortman J, Bovenberg RA.** 2008. Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. Nat. Biotechnol. **26:**1161–1168.
7. **Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M.** 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. **18:**188–196.
8. **Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B.** 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. **37:**D233–D238.
9. **Staszczak M.** 2008. The role of the ubiquitin-proteasome system in the response of the ligninolytic fungus *Trametes versicolor* to nitrogen deprivation. Fungal Genet. Biol. **45:**328–337.
10. **Nowak SJ, Corces VG.** 2004. Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. Trends Genet. **20:**271–285.
11. **Belenky P, Bogan KL, Brenner C.** 2007. NAD^+ metabolism in health and disease. Trends Biochem. Sci. **32:**9–12.
12. **Berger F, Ramírez-Hernández MH, Ziegler M.** 2004. The new life of a centenarian: signalling functions of NAD(P). Trends Biochem. Sci. **29:**111–118.
13. **Corda D, Di Girolamo M.** 2003. Functional aspects of protein mono-ADP-ribosylation. EMBO J. **22:**1953–1958.

2.1 INTRODUCTION

Genomics is a recent research field which has taken off in the late 90s and early 2000s reaching its first high with the completion of the human genome. The ability to study a whole genome and transcriptome has given scientists the ability to generate a high degree of intricate information about any living organism. The total number of genome and transcriptome projects that have been undertaken stands at 24608 with more than 4325 completed projects (Genomes OnLine database, GOLD, http://www.genomesonline.org). According to Genomics OnLine database, currently, 3956 bacterial, 186 archeal and 183 eukaryotic projects have been completed. Eukaryotic projects are the lowest which leads to the reasonable conclusion that very few fungal genomes have been completed, especially thermophilic filamentous fungi (Genomes OnLine database, GOLD, http://www.genomesonline.org). Thus, any filamentous fungus genome project will greatly contribute to the expanding knowledge. In addition to genomes, transcriptomes also have become an important source of information about an organism, which complements genome sequencing.

The brewer's yeast, *S. cerevisiae* was the first eukaryote to have its genome sequenced in 1996, revealing only 43.3% of the genes have known biochemical functions (Goffeau *et al.,* 1996). This created a base for other studies of fungi and other eukaryotic organisms to follow. Numerous filamentous fungal genome projects have followed with many showing some similarities and differences. For example, the genome of *N. crassa* assembled five years after *S. cerevisiae* from genomic data of more than 20-fold sequence coverage of the genome and has the largest genome size among all published fungal genomes to date (Galagan *et al.,* 2003). For a number of *Neurospora* genes the only known homologs are found in prokaryotes, indicating that occupation of similar ecological niches has resulted in conservation of genes for substrate degradation and secondary metabolism (Galagan *et al.*, 2003).

Magnaporthe grisea, one of the most problematic pathogens causing blast disease in rice, destroys enough rice crops to feed 60 million people annually. The genome of this pathogen contained an extended family of G-protein- coupled receptors, several new virulence-associated genes and large collections of enzymes involved in secondary metabolism, which may play a role in resistance and pathogenicity (Dean *et al.*, 2005). This was followed by the sequencing of

both *C. albicans* and *C. neoformans* enabling a comparison between these divergent fungi. Moreover, high-quality draft sequences and completed genomes of *A. niger*, *A. flavus*, *T. reesei*, *A. nidulans* and *A. fumigatus* are already in the public domain and selected genomes can be found in **Table 6**, the rest in the FUNGIpath database (http://embg.igmors.u-psud.fr/fungipath/).

Table 6: Selected fungal genomes found on the FUNGIpath database (Grossetête *et al.*, 2010)

Genome Name	Source
Aspergillus nidulans	Broad Institute
Aspergillus oryzae	NITE
Batrachochytrium dendrobatidis	Broad Institute
Chaetomium globosum	Broad Institute
Coprinus cinereus	Broad Institute
Fusarium graminearum	Broad Institute
Laccaria bicolor	JGI
Magnaporthe grisea	Broad Institute
Mycosphaerella graminicolla	JGI
Neurospora crassa	Broad Institute
Phycomyces blakesleeanus	JGI
Podospora anserina	IGM
Puccinia graminis	Broad Institute
Saccharomyces cerevisiae	Stanford
Schizosaccharomyces pombe	SANGER
Sclerotinia sclerotiorum	Broad Institute
Stagonospora nodorum	Broad Institute
Trichoderma reesei	JGI
Ustilago maydis	Broad Institute
Yarrowia lipolytica	Genolevures

The increasing amount of genome information needs to be properly analysed and assessed. This includes assembly, annotation and gene predictions. Genome annotation is the process of assigning function to predicted gene sequences. It consists of three main steps, (i) identifying parts of the genome that do not code for proteins, such as repeat regions, promoters, small RNA etc., (ii) gene prediction/calling, and (iii) assigning biological information to the predicted genes. Computational systems and algorithms provide tools for data handling and processing of data produced during genome sequencing. For easy application of these systems, data must be easy to access and to analyze in feedback mechanism with analyzed data being able to inform a new

process loop and new hypotheses to be established. Prediction methods (gene/protein and function) can be divided into two major groups: homology-based methods and non-homology based methods (**Figure 12**).

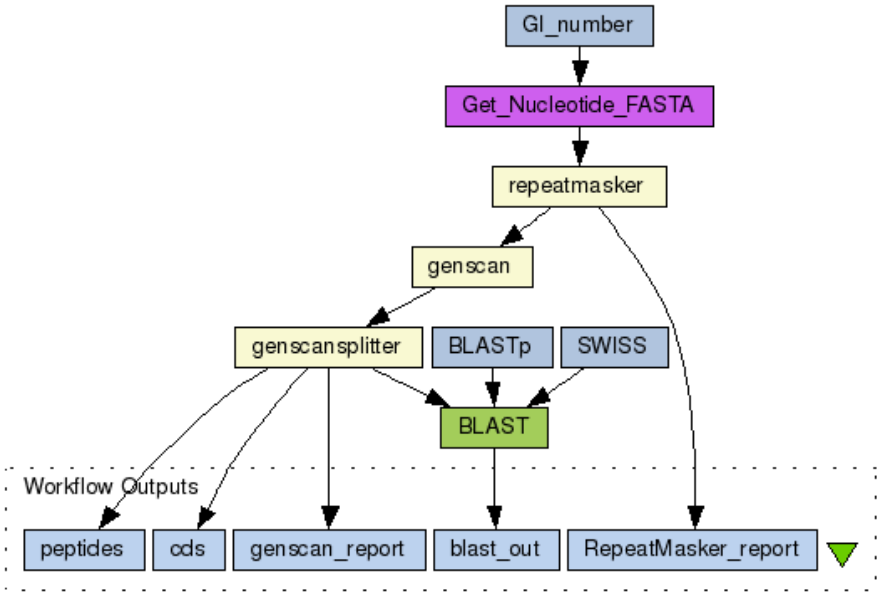

Figure 12: Diagram showing an example of workflow of genome annotation adapted from (Hull *et al.*, 2006).

Homology-based annotation methods rely on sequence similarity between the query gene and a known gene (Daniels *et al.*, 2013). This is based usually proteins that are highly similar in sequence, might share a similar function**.** The reasoning behind this is that similarity in sequence may suggest functional similarity (Descorps-Declère *et al.*, 2008). This is reasonable but not always true, as some proteins share numerous domains which may yield misleading function and incorrectly annotated database entries.

Regardless of such complications, homology-based methods are still the most widely used methods. To determine similarities with sequences of known proteins, the similarities are computed using a rigorous dynamic programming algorithm such as BLAST (Martin *et al.*, 2002). Further predicted protein similarity and comparison is analyzed by identifying motifs or domains which will provide additional information on possible function. A motif is a simple

combination of a few consecutive secondary structural elements with a specific geometric arrangement. But not all motifs, however, are associated with a specific biological function. A domain is the fundamental unit of structure folding and evolution. It may combine several secondary elements and motifs, not necessarily contiguous. A domain can fold independently into a stable 3D structure, and has a specific function. A variety of mathematical representations of protein motif or domains, have been developed using hidden Markov models and sparse Markov transducers as examples (Abad *et al.*, 2008; Fridlyand *et al.*, 2004; Gough *et al.*, 2001; Wei and Li, 2007).

Although homology-based annotation has been used successfully in annotation of proteins found in genome sequencing projects, a serious problem is that well characterized reference proteins must be found based on sequence similarity; otherwise, one cannot assign the putative function to the unknown protein. According to most genome data that is available, 30-40% of proteins have no clear sequence homology to proteins, found in protein or gene databases (Giaever *et al.*, 2002; Krause *et al.*, 2010; Venter *et al.*, 2004). Non-homology-based method; also called context-based function prediction, is complementary to homology-based function prediction. Phylogenetic profiles, domain fusion and gene neighbouring are examples of these methods. If two proteins tend to be either preserved or eliminated together in a new species, it is assumed that they are functionally linked. If the function of one of the protein is known, the prediction of function of the unknown protein is based on this relationship. The method of phylogenetic profiling could be useful in predicting the function of uncharacterized proteins in genomes, especially, when more and more fungal species are sequenced. But this method has to be performed manually because there is no free software available in assisting automation of the analysis, which can be time-consuming when working with thousands of genes (Baxevanis, 2002).

Even with profiling, it is not easy to assign a function to unknown or uncharacterized proteins. Another approach to use protein classification to identify the function of the new protein. Comparing a protein sequence with a database of protein families is more effective than a standard database search. In general, conserved proteins are classified according to their homologous relationships (Fleming *et al.*, 2006). Each protein group is composed of a set of proteins, which is represented as multiple alignments, regular expression profiles or hidden

Markov models. Protein classification is useful in structure and function prediction, and especially important in large-scale annotation efforts. Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families. For each protein family, Pfam allows investigating of multiple alignments, viewing protein domain architectures, examining species distribution, etc. Pfam complements releases of Swiss-Prot and TrEMBL (Finn *et al.*, 2010)with at least 70% of protein sequences in Swiss-Prot and TrEMBL have at least one match to Pfam. InterPro is a database of protein families, domains and functional sites in which identifiable features found in known proteins which can be applied to unknown protein sequences. It provides an integrated view of the commonly used signature databases like ProDom, PROSITE, PRINTS, SMART and Pfam. InterProScan combines different protein signature recognition methods native to the InterPro member databases into one resource to look up of corresponding InterPro and GO annotation (Daniels *et al.*, 2013).

Clusters of Orthologous Groups of proteins (COGs) were compiled by comparing protein sequences encoded in 66 complete genomes, representing 30 major phylogenetic lineages (Tatusov *et al.*, 2000). Currently, it is being further updated by including additional complete genomes representing broader families. Each COG consists of individual proteins or groups of paralogs from at least three lineages and thus corresponds to an ancient conserved domain. The problem with COGs is that the system only accessible within NCBI making annotation difficulty as it must done online. An additional problem is that COGs do not distinguish paralog (genes from the same genome which are related by duplication) from ortholog (genes in different species that evolved from the same ancestral protein). Orthologs typically have the same function, allowing transfer of functional information from one member to an entire COG. In contrast, paralogs are functionally diverse proteins whose genes duplicated after speciation, although high sequence similarity is normally preserved in paralogs. A system like COGs can only be used as a classifying system for automatically yielding a number of functional predictions for uncharacterized sets of genes or proteins (Ekman and Elofsson, 2010).

This chapter focuses on sequencing, assembly and features found in *T. lanuginosus* using sequencing data obtained from genomic DNA data (genome) and total RNA (transcriptome). In

addition a comparison with other selected fungal genomes (yeast and filamentous fungi) will be presented.

2.2 MATERIALS AND METHODS

2.2.1 Growth and maintenance of *T. lanuginosus*

T. lanuginosus SSPB (isolated in South Africa) was grown on potato dextrose agar at 50°C for 5 days before inoculation into liquid medium containing 0.1% KH_2PO_4, 2% glucose, 3% yeast extract at pH 6.5 (Singh et. al., 2000). *T. lanuginosus* was maintained in potato dextrose agar slants and sub-cultured after two months. A one cm^2 plug of the culture was excised from a plate inoculated into liquid medium and grown for 2 days at 50°C with shaking at 150 rpm. Twenty millilitres of culture was harvested by centrifugation at 2000 *g* for 10 min. Pellets were snap frozen in liquid nitrogen then stored at -80°C for genomic DNA and total RNA isolation.

2.2.2 Genomic DNA isolation

DNA extraction was done using a protocol by Melo *et al*. (2006), with minor modifications. The frozen *T. lanuginosus* pellet was resuspended in 5 ml extraction buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and glass beads (0.2 mm) were added. An equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) was added to the react tube and the suspension was vortexed at high speed for 5 min and then centrifuged for 10 min at 10 000 *g*. The supernatant was transferred to a microcentrifuge tube, and DNA was precipitated by adding 2 volumes of ice-cold absolute ethanol. The tube contents were mixed gently and incubated for 20 min at -70°C to precipitate nucleic acids. The sample was centrifuged at 4°C for 10 min and the pellet was resuspended in 1 ml of Tris-EDTA, pH 8 with 10 µl RNAse (20 mg/mL) and incubated for 1 h at 37°C without shaking to degrade contaminating RNA. The DNA was precipitated by adding 2 volumes of absolute ethanol and gentle mixing followed by re-incubation at -70°C. After centrifugation for 15 min, the DNA pellet was washed 3 times with 75% ethanol, then air-dried in a laminar flow cabinet and finally resuspended in 500 µl of HPLC-quality sterile water. The genomic DNA was quantified using a NanoDrop 2000 analyzer (Thermoscientific) and analysed by agarose gel electrophoresis before proceeding to whole genome sequencing.

2.2.3 Total RNA isolation

Total RNA from *T. lanuginosus* was isolated using a combination of trizol reagent (Sigma) and Plant Total RNA Isolation kit from Qiagen. This was done by adding 5 ml of trizol reagent to the frozen pellet of *T. lanuginosus,* followed by vortexing until the pellet thawed and was resuspended in trizol reagent. Thereafter 200 µl phenol-chloroform-isoamyl alcohol (25:24:1) of was added to the sample for every 1 ml of trizol. The tube was then vortexed for 15 s and incubated at room temperature for 5 min. This was followed by centrifugation at 4°C for 10 min at 10 000 *g*. The aqueous top layer of the mixture was transferred to a new microcentrifuge tube, followed by addition of an equal volume of ice0cold ethanol. The resulting mixture was transferred to a spin column in 700 µl aliquots and centrifuged at 5000 g for 15 s. The columns were washed with 500 µl of ethanol-containing RPE wash buffer. To remove contaminating DNA, on-column digestion was performed by adding 100 µl DNase supplied by Qiagen on to the column and incubating for 30 min at room temperature. The columns were washed again with 500 µl of RPE wash buffer followed by centrifugation at 5000 *g* for 2 min. The columns were dried by centrifuging them at 10 000 *g* for 1 min. The total RNA was eluted with 30 µl of RNase free water. Thereafter, the samples were kept on ice or stored at -70°C for further analysis. Total RNA quantification and integrity verification was done using a NanoDrop 2000 analyzer (Thermo Scientific), a Bioanalyzer (Agilent Technologies, USA) and by 1.2 % agarose gel electrophoresis, before proceeding with sequencing.

2.2.4 DNA and RNA quantification

Quantification of nucleic acids was done using the NanoDrop 2000 analyzer. The quantification was based on the assumption that an absorbance of 1 at optical density (OD) of 260 nm corresponds to 50 ng DNA/µl and 40 ng RNA/µl (Sambrook and Russell, 2006). Purity of the nucleic acid samples was estimated using the absorbance ratio at 260 nm and 280 nm. DNA samples with a ratio of greater than 1.8 were considered good quality samples and for RNA a ratio of more than 2 was ideal for sequencing.

2.2.5 Agarose gel electrophoresis

The quality of nucleic acid samples were also analysed using agarose gel electrophoresis. Genomic DNA samples were analyzed using 0.8% agarose while for total RNA 1.2% was used. The gels were prepared by mixing appropriate amounts of agarose and 1x TAE buffer, which was prepared from a 50x TAE stock (242 g Tris, 57.1 ml acetic acid, 100 ml of 0.5 M EDTA, pH 8) and microwaved for 45 s. The gel was allowed to cool to less than 50°C then SYBR Green I (Invitrogen) was added to the gel. This was poured into a casting tray with well combs and allowed to solidify at room temperature. Gel loading buffer (Fermentas) was added to the DNA samples in a ratio of 1:5 which was then loaded into the agarose gel wells. The RNA samples containing loading buffer were heated for 5 min at 60°C to break RNA self-priming then loaded into the gel. An aliquote of a 1000bp DNA Molecular Weight Marker (Fermentas) was loaded next to the samples and the gel was electrophoresed at 120 V for approximately 1 h (Bio-Rad Gel Electrophoresis). Gels were then viewed under UV and images captured and stored using a Gel Doc XR System (Bio-Rad) and the band sizes estimated by comparison to the DNA marker.

2.2.6 Total RNA integrity validation

The quality integrity of RNA samples was assesses using the Agilent Bioanalyzer 2100 and the RNA 6000 LabChip kit . Sample preparation was according to the manufacturer's instructions (Agilent). An amount of 550 µl gel matrix was added to a spin filter and centrifuged for 10 min at 1500 *g*. Aliquots of 65 µl were then prepared from the filtered gel and stored at 4°C. Gel matrix was prepared by mixing 1 µl of dye (provided in the kit) was added to the gel reagent and vortexed thoroughly for 10 s then centrifuged at 13 000 *g* for 10 minutes. A Nano Chip was placed on a provided priming station and loaded with 9 µl of gel-dye matrix onto a well marked G in bold. The chip priming station was closed and the plunger pressed for 30 seconds then released. Then 9 µl of the gel-dye matrix was added to the two additional wells marked G followed by the addition of 5 µl marker into the wells marked ladder and 12 RNA wells. One microlitre of high range RNA ladder (Fermentas) was loaded to ladder well and 1 µl of total RNA (50-500 ng) into each of the RNA wells. The chip was mixed for 1 minute at maximum speed using the chip mixer (Agilent).

The Agilent Bioanalyzer 2100 was washed with RNase-free water and RNase Away (Invtrogen) for 1 min each and the machine was left to dry for 10 seconds. The loaded Nano Chip was placed in the Bioanalyzer and the instrument set to perform RNA assay for eukaryotic samples. Samples having intact 18S and 28S ribosomal RNA and lacking any DNA contamination were used for further analysis.

2.2.7 Roche FLX Raw sequencing data generation

2.2.7.1 Paired-end library preparation

Sequencing libraries were prepared using the manufacturer's instructions for paired-end (PE) library preparation, Rapid Library preparation, emPCR, and sequencing. All sequencing reactions were performed at the Advanced Studies in Genomics, Proteomics and Bioinformatics (ASGPB), University of Hawaii. Roche 454 sequencing strategy with single-ends and paired-ends to construct a genomic library containing 8-kb, 15-kb and 20-kb inserts. With the Solexa sequencing strategy, a paired-end (PE) and mate-paired (MP) genomic libraries were constructed.

The double-stranded DNA sample was fragmented using Hydroshear DNA shearing system (Digilab) equipped with the Standard Assembly for the setting of 20 cycles of speed code 16. The frayed ends of the DNA fragments were blunted using klenow fragment (Roche) and 454 circularization adaptors ligated to the ends of the fragments. DNA fragments were separated by electrophoresis, and fragments with the correct size (400-1000bp) were electroeluted using Elutrap (Whatman). Nicked DNA was repaired with the *Bst* DNA polymerase (Roche) by strand displacement, and the DNA was quantified using the Nanodrop. The DNA was then circularized by a recombination at the *lox*P sites present within the circularization adaptors using *Cre* recombinase (Roche). The resulting circularized DNA molecules were fragmented by nebulization and frayed ends blunted and immobilized on streptavidin-coated magnetic beads (Lifetechnologies). After immobilization, library adaptors were ligated to the ends. This was then amplified by PCR and size-selected using AMPure size exclusion beads. The double-stranded DNA (dsDNA) amplified molecules were immobilized once more on streptavidin-coated

magnetic beads, and the single stranded paired-end DNA library was released from the magnetic beads by alkaline treatment, amplified by emulsion-based clonal amplification (emPCR), and then sequenced on the Genome Sequencer FLX Instrument.

2.2.7.2 cDNA library preparation

For cDNA library preparation, mRNA was isolated from total RNA using the Oligotex mRNA Mini Kit protocol (Qiagen). cDNA was synthesized from mRNA using the SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen) with 100 µM random hexamer primers (Fermentas, USA). After the second strand synthesis, the cDNA was purified using Qiagen quick PCR MinElute column (Qiagen) and eluted in 50ul elution buffer and then nebulized with the Nebulization Kit using the GS Titanium Library Preparation Kit (Roche) following their recommendations (200 Kpa for 1 minute) and re-purified with a Qiagen quick PCR MinElute column and eluted in 50ul elution buffer. The cDNA was then blunt-ended, dA-overhangs added at 3' ends and Titanium adaptors added to the ends. The cDNA was size-selected with AMPure beads (Agencourt), amplified using PCR and purified with a Qiagen quick MinElute PCR column. The cDNA library was normalized according to the protocol described in the Trimmer Direct Kit (Evrogen, Russia). The normalized cDNA was then amplified and gel-purified for the fragment size of 400-1000 bp in size on a 2% agarose Ex-GEL (Invitrogen). Normalized cDNA was quantified and the average size was determined on a Bioanalyzer chip (Agilent).

2.2.7.3 emPCR

Emulsion PCR (emPCR) amplifies the library of DNA template fragments from single, functionally clonal, bead-bound copy, to tens of millions of copies per bead for sequencing with the genome sequencer system. The DNA library prepared in section 2.2.7.1 was flanked with sequencing adaptors, then amplified using the GS FLX Titanium SDeries emPCR Kits. The cDNA library was immobilized onto DNA capture beads to allow for their segregation in the emulsion step. During emulsification, the captured DNA library was resuspended in the amplification mix and oil to form a water-in-oil mixture. The emulsified beads were subjected to PCR to clonally amplify each template DNA molecule. The DNA templates were hybridized to

bead-bound oligonucleotide primers or the capture primers. These capture primers acted also as PCR primers, anchoring the newly synthesized complementary strands to the beads. After amplification, typical immobilized template copy number ranges from 10 to 50 copies per bead. After amplification, the emulsion was broken, and the beads carrying the amplified DNA library (double-stranded at this point) were recovered and washed. The amplified DNA was enriched by first hybridizing biotinylated Enrichment Primer (supplied with the kit) to the adaptor of the amplified template and binding to streptavidin-coated magnetic beads. The final step in the emPCR amplification process was the annealing of the sequencing primer to the immobilized, amplified DNA templates. The DNA library beads are then ready for loading onto a Pico TiterPlate device (Roche) and sequencing on the Genome sequencer instrument with the GS FLX titanium sequencing kit XLR70 (Roche GS FLX Titanium Sequencing Method Manual-revised June 2010)

2.2.7.4 Roche Titanium sequencing

The Roche Titanium FLX System was used to generate shotgun and pair-end sequencing data according to the manufacturer's instructions. DNA library beads were added to the DNA bead incubation mix (containing DNA polymerase supplied with the kit) and layered with enzyme beads (containing sulfurylase and luciferase) onto the Pico TiterPlate device. This was centrifuged to deposit the beads into the wells. The loaded Pico TiterPlate was placed into the Genome Sequencer FLX Instrument. The four DNA nucleotides were added sequentially in a fixed order across the Pico TiterPlate during sequencing and the millions of copies of DNA bound to each of the beads were sequenced in parallel. Addition of each nucleotide generated a light signal that was recorded by the CCD camera in the instrument. This technique is based on sequencing-by-synthesis and is called pyrosequencing. Data was stored in standard flow gram format (SFF) files for assembly and further analysis.

2.2.8 Solexa Illumina Sequencing

Sequencing libraries were prepared by using Illumina Paired-End DNA Sample Prep Kit from the supplier. DNA (5 μg) was sheared by nebulization using 220 Kpa for 8 min, and the sheared fragments were purified and concentrated using the QIAquick PCR purification Kit (QIAgen). DNA fragments were treated with T4 DNA polymerase, T4 phosphonucleotide kinase and the Klenow fragment to fill 5′ overhangs and polish 3′ overhangs in 1.0 mM ATP, 0.5 mM dNTPs, 0.1 mg/ml bovine serum albumin, 20 U T4 DNA polymerase (NEB), 50 U T4 polynucleotide kinase (NEB) reactionand was incubated for 15 minutes at 10°C followed by 15 minutes at 25°C. Standard Illumina adapters were then ligated to DNA fragments. Adapter-ligated DNAs in the range of 300-400 bp pair-end and 1500-3000 bp for pair-mate, were selected using agarose gel electrophoresis. The products were purified using the QIAgen MiniElute gel Extraction Kit. The insert libraries were then amplified independently using 20-cycle PCR amplification and the standard Illumina primers and protocol. Amplified libraries were again size-selected using gel agarose electrophoresis, to remove PCR products below 200 bp.

Libraries (10 pmol) were denatured and processed with the Illumina Cluster Generation Station, following the manufacturer's recommendations. Sequencing was performed with an Illumina Genome Analyzer and was programmed to run for 40 cycles. After the run, sequence analysis, base calling and error estimation were performed using Illumina/Solexa Pipeline version 0.2.2.6

2.2.9 Genome assembly

The *T. lanuginosus* genome assembly was generated by *de novo* assembly of next-generation sequencing data using the CLC Genomoics Workbench assembler and the Newbler assembler (version 2.3) and data from Roche 454 and Illumina reads generated in section 2.2.7. Assembling parameters for the Newbler assembly were selected for *de novo* assembly with reads limited to one contig, minimum overlap length 50 bp, minimum overlap identity 95%. The Newbler assembler was used to generate the final assembly. RepeatMasker (www.repeatmasker.org) was used to identify and mask repeat regions in the genome assembly. TopHat (version 1.1.4-http://tophat.cbcb.umd.edu/) was used to align the transcriptome sequencing reads produced by

the Roche 454 sequencing platform to the assembly, for assessing the coverage of genome assembly and of coding regions.

2.2.10 Genome annotation

For genome sequence annotation and gene calling, a modified version of MAKER was used (Cantarel *et al.*, 2011). The assembly repeats were masked using RepeatMasker (http://repeatmasker.org). Annotation of the masked assembly was conducted through EVidenceModeler (EVM; version r03062010) using transcriptome, protein alignments and *ab initio* gene predictions. The transcriptome assembly was aligned to the genome using Exonerate (version 2.2.0) and was used as training set for the following *ab initio* gene prediction softwares: Augustus (version 2.5), GeneMark and SNAP. Protein sequences of the final gene sets were functionally annotated through BLASTP searches with an E-value cut-off of 10^{-5} against protein databases (Swiss-Prot, UniProt, NCBI, KEGG,and InterProScan). COG and EC assignments were extracted from BLASTP hits of STRING and GENES. GO assignment was extracted from searching results of the InterPro database. The gene predictions in the genome were then aligned with ESTs to inform MAKER's gene annotation process. Transcriptome, Swiss-Prot and Exonerate were used to realign matches and highly similar ESTs, mRNAs and proteins to the genomic input sequence. Based upon information obtained using the above validations, the predictions were then modified using internal Hidden Markov Models to final annotations. Also the Gbrowser was create for *T. lanuginosus* based on sequencing data, annotation, transcriptome, and ESTs.

2.2.11 Detection of CAZy family proteins

The detection of carbohydrate-active modules (glycosylhydrolases, glycosyltransferases, polysaccharide lyases and carbohydrate esterases) and their associated carbohydrate-binding modules (CBMs) in *T. lanuginosus* was performed using the carbohydrate-active enzyme database (http://afmb.cnrs-mrs.fr/CAZY/). All the predicted protein sequences were loaded into the CAZy family database whereby genes were annotated based on the similarity and identity

with known genes and the presence of carbohydrate domains (Byung *et al.*, 2010; Cantarel *et al.*, 2009).

2.2.12 Manual curation of data

The data was manually curated using gene function assignments from InterPro, Refseq and Swiss-Prot. If the best sequence similarity was above 80% identity and 80% coverage with respect to a protein for which there was published experimental evidence of its function, the gene sequence was assumed to represent that protein. If the sequence similarity was above 70% identity and 70% coverage, yet lower than 80% identity and 80% coverage, with respect to a protein with published evidence of function, it was described as a ‘‘candidate’’ for that function. If the sequence similarity was above 50% identity and 50% coverage with respect to a protein with published evidence of function, the function was described as ‘‘related to’’. Below this last threshold, all proteins were assigned as ‘‘hypothetical.’’ Unknown and hypothetical proteins with hits to only other unknown hypothetical proteins were assigned as ‘‘conserved hypothetical.’’

2.2.13 Synteny analysis

Syntheny analysis of *T. lanuginosus* was done against selected genome assemblies of *A. nudilans*, *A. fumigatus* and *T. reesei*. *T. lanuginosus* scaffolds were analyzed using the Artemis Comparison Tool (ACT; http://www.sanger.ac.uk/resources/software/act/, (Carver *et al.*, 2008)) and WebACT (Abbott *et al.*, 2005). WebACT was used to generate the comparison file and ACT was used to visualize the synteny between these filamentous fungal organisms.

2.2.14 Fungal proteome comparisons

The fungal organisms used for comparison in this study included *A. fumigatus*, *A.nidulans*, *F. graminearum*, *T. reesei*, *M. grisea*, *N. crassa*, *Ashbya gossypii*, *Candida albicans*, *Kluyveromyces lactis*, *Yarrowia lipolytica* and *S. cerevisiae* (Grossetête *et al.*, 2010).

2.3 RESULTS

2.3.1 Genomic DNA and total RNA isolation

2.3.1.1 Genomic DNA isolation

The frozen pellet of *T. lanuginosus* was used to isolate genomic DNA and total RNA for sequencing using next-generation platforms. DNA isolation done using the phenol-chloroform-isoamyl alcohol method yielded a high concentration of genomic DNA. A 100-fold dilution of the isolated DNA run on an agarose gel showed that intact DNA was isolated (**Figure 13**). *T. lanuginosus* genomic DNA was successfully isolated and the DNA samples showed very little shearing (lane 2 and 3). The size of the DNA samples was above 10 kb which was suitable for sequencing. NanoDrop measurements revealed that the concentration was above 10 µg/µl with an absorbance ratio above 1.8 at 260/280 nm.

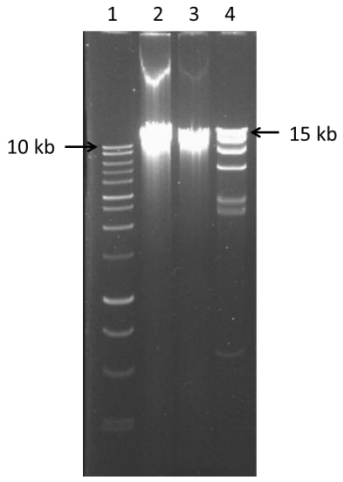

Figure 13: Agarose gel electrophoresis of *T. lanuginosus* genomic DNA (Lanes 2 & 3) against high range molecular weight markers (Lane 1) and lambda/HindIII (lane 4).

2.3.1.2 Total RNA isolation

From RNA sample two types of samples were analysed, one that were untreated and samples that were DNase treated. Agarose gel electrophoresis revealed that the integrity of the DNAse treated RNA sample was acceptable which was identified by sharp bands which represented 18S and 28S ribosomal RNA (**Figure 14,** lanes 2, 4 and 6), in comparison with the untreated RNA

samples contained a band above the 18S and 28S bands (lanes 1, 3, 5 and 7) which was assumed to be contaminating DNA. The contaminating band was not visible or less intense in samples that were DNAse treated (Lane 2, 4 and 6) pointing to the degradation of DNA. The Bioanalyzer results also supported agarose gel electrophoresis data showing that the 18S and 28S rRNAs were not degraded (**Figure 15**). The first peak in the Bioanalyser electrograph represents the RNA marker.

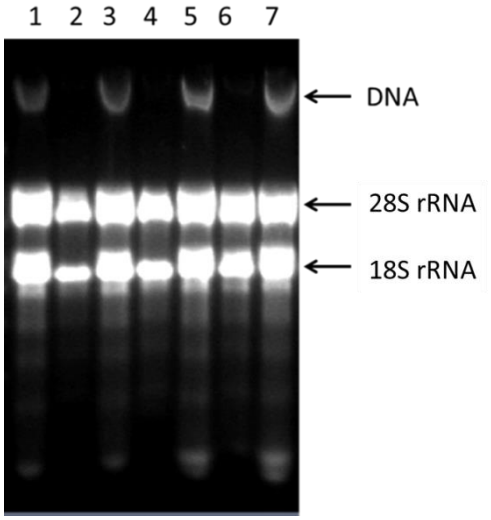

Figure 14: Agarose gel electrophoresis of *T. lanuginosus* total RNA (Lanes 1-7). Lanes 1, 3, 5 and 7 show raw total RNA samples while Lanes 2, 4, and 6 show RNA samples that have been DNAse treated**.**

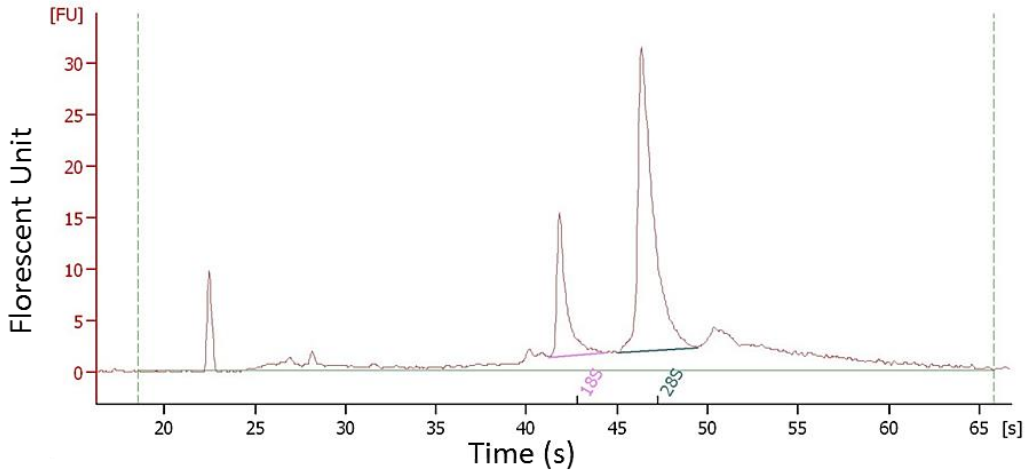

Figure 15: Bioanalyzer electropherogram of total RNA samples of *T. lanuginosus* confirming intact RNA as shown by the presence of the 18S and 28S peaks.

2.3.2 Sequencing statistics

A combination of Roche 454 and Illumina Solexa strategies were adopted to sequence the whole genome of *T. lanuginosus*. In Roche 454 sequencing, single-end and paired-end genomic libraries were constructed containing 8-kb, 15-kb and 20-kb inserts. In total, 83 409 shotgun reads and 812 861 paired reads were generated with the Roche 454 GS FLX System and the total peak depth reached 48.0 fold. With the Solexa sequencing strategy, paired-end and mate-paired genomic libraries were constructed, and the average length of inserts were about 365 and 28 680 bp, respectively. In total, 32 055 256 reads were generated to reach about 137.6-fold coverage of the genome (**Table 7**). For the assembly, more than 98% of the total reads produced were used to generate the genome assembly.

For the transcriptome, 1 484 071 reads were obtained using the Solexa sequencing strategy (**Table 7**). More than 96% of the reads were mapped to the genome assembly. When the read data was assembled using the *de novo* method with the Newbler assembler, 97% of the reads were assembled with a coverage of 48-fold. The transcriptome data was used only as part of the evidence to support gene annotation for the genome assembly, therefore in this study, there was no separate assembly of the transcriptome (**Figure 16**).

Table 7: Summary of the sequencing data of *T. laniginosus* produced by Roche 454 and Solexa sequencing strategies

Read type and strategy	Insert size (kb)	Total reads (bp)	Total bases (bp)	Fold coverage (X)
Roche 454 single-end	ND*	830 409	358 946 381	15.4
	ND*	814 105	350 927 338	15.1
Roche 454 paired-end	7031.3	232 700	91 176 092	4.0
	13964.3	300 281	114 459 285	5.0
	13960.9	279 880	100 853 056	4.4
	18078.4	232 700	91 176 092	3.9
Solexa paired-end	364.5	25 187 456	2 518 745 600	108.1
Solexa mate-paired	2868	6 867 800	686 780 000	29.5
Transcriptome solexa	ND*	1 484 071	585 383 245	12.0

*ND: not determined

2.3.3 Genome assembly

2.3.3.1 Genome overview

The data produced by the two sequencing strategies was used to assemble the genome, using newbler assembler (version 2.3). A 23.3 Mb *T. lanuginosus* genome sequence was created by *de novo* assembling 98% of the sequencing data generated using next-generation sequencing technology (Roche and Solexa data). The repeat reads were less than 0.5% of the total reads

generated. The assembly contained 30 scaffolds with 1125 contigs (**Table 8**). The N50 Scaffold size and N50 Contig size reached 4.26 Mb and 1.12 Mb, respectively. Most of the assembly (90%) is contained in the 6 largest scaffolds and the rest in 24 smaller scaffolds with 22 of the less than 10 kb (**Table 9**). The genome was annotated using a modified version of MAKER (Cantarel et al, 2007). The nucleotide composition of the chromosome was analyzed and GC content of the whole genome was calculated to be 52.14%, while coding regions had a GC content significantly higher at 55.6% **(Table 8).** Gene modelling performed using a combination of homology and *ab initio* strategies yielded 5105 ORFs with an average gene length of 1 615 base pairs (bp) containing 3.7 exons and 3 introns per gene. From the 5105 ORFs evaluated, 83 were identified as possible candidates for tRNA genes.

Table 8: General features of *Thermomyces lanuginosus* genome

Feature	**Value**
Genome size	23.3 Mb
Number of contigs	1125
N50 contig	1.12 Mb
Number of scaffolds	30
N50 scaffold	4260952 bp
Coding regions	35.40%
Predicted genes encoding proteins	5105
GC content	52.14%
GC content for coding genes	55.60%
Average gene size	1615 bp
Average number exons per gene	3.7
Average number of introns	3
tRNA genes	83

Table 9: *T. lanuginosus* assembled scaffolds

Scaffold	Length (bp)
scaffold00001	6976347
scaffold00002	4260952
scaffold00003	3532058
scaffold00004	2453013
scaffold00005	1593813
scaffold00006	1146835
scaffold00007	29449
scaffold00008	12760
scaffold00009	8040
scaffold00010	6909
scaffold00011	5263
scaffold00012	5010
scaffold00013	4809
scaffold00014	4185
scaffold00015	2789
scaffold00016	2657
scaffold00017	2587
scaffold00018	2548
scaffold00019	2388
scaffold00020	2385
scaffold00021	2330
scaffold00022	2261
scaffold00023	2248
scaffold00024	2238
scaffold00025	2146
scaffold00026	2145
scaffold00027	2134
scaffold00028	2115
scaffold00029	2092
scaffold00030	2041

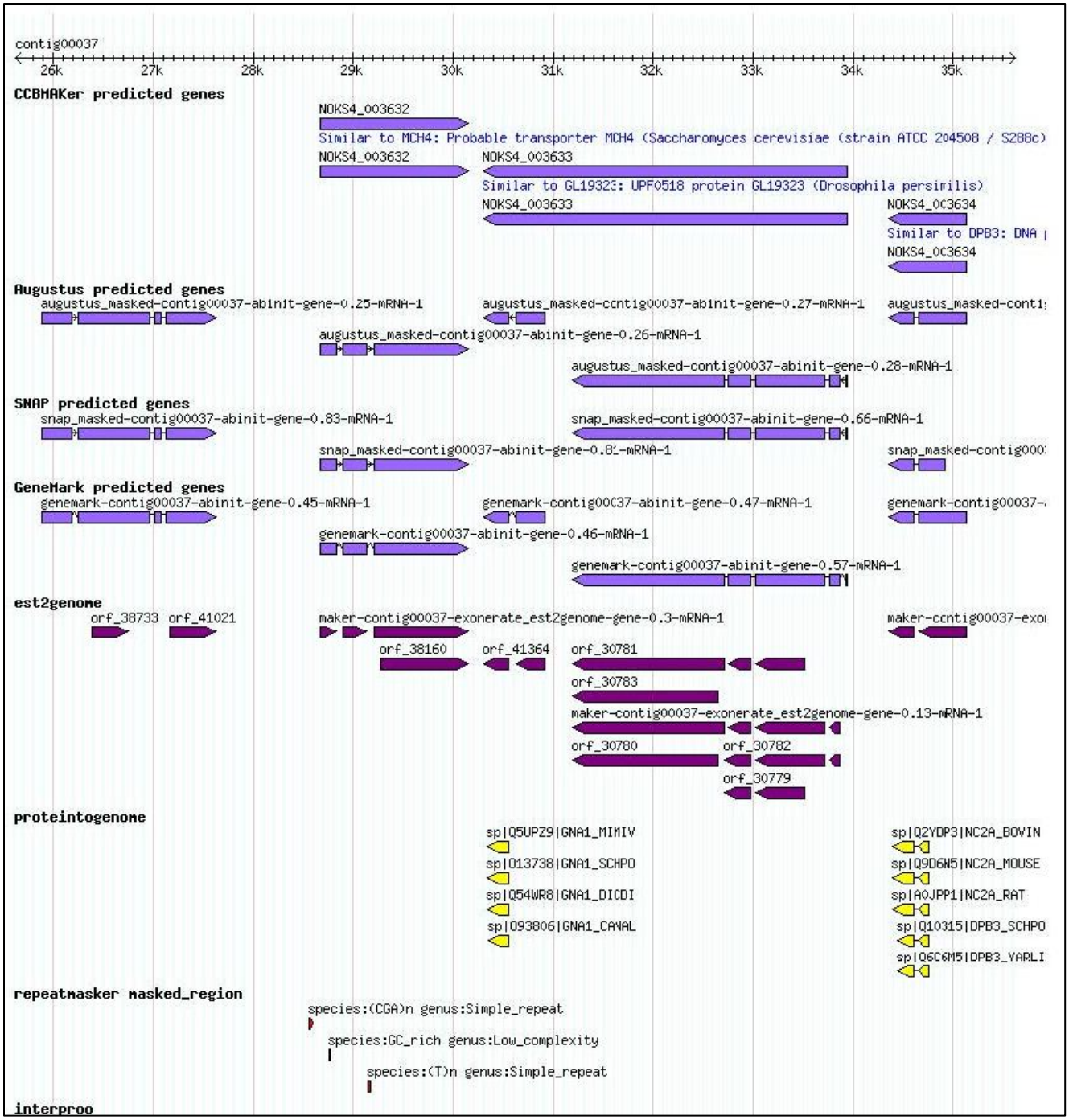

Figure 16: Gbrowser image showing typical information for *T. lanuginosus* genome. The Gbrowser uses contig information and all the evidence used for gene annotations (gene predictions from MAKER, Augustus, Genemark and ESTs).

2.3.3.2 Genome comparison with other fungi

The number of gene encoding proteins or predicted open reading frames (5105) obtained using homology and *ab initio* strategies were lower compared to the other filamentous fungi listed in **Table 10** (Druzhinina et al., 2006; Galagan et al., 2005, Galagan et al., 2003; Ma et al., 2010; Machida et al., 2005, Martin et al., 2010; Martinez et al., 2009; Nierman et al., 2005). The low number of predicted ORFs is surprising as the average number of ORFs reported in other filamentous fungi ranged between 9000-13000 (**Table 10**). However, the number of predicted genes was similar to that reported for yeasts, *K. lactis*, *K thermotolerans* and *S. cerevisiae*. The GC content of *T. lanuginosus* was generally similar to other filamentous fungal species which was calculated to be around 52% and above.

In order to analyze divergence of *T. lanuginosus*, analysis of shared synteny was performed. Synteny is the conservation of genetic order within two sets of chromosomes that are being compared with each other. Low synteny was evident with most of the fungal genomes analysed as shown from the comparison with *A. nidulans* and *A. fumigatus* (**Figure 17 & 18**) whereas moderate synteny was observed with *T. reesei* (**Figure 19**).

Table 10: Comparative analysis of *T. lanuginosus* genome with other fungal species

Organism	Genome size	Predicted genes	% GC
T. lanuginosus	**23.3 Mb**	**5 105**	**52.1%**
F. graminearum	36.1 Mb	11 640	48.3%
N. crassa	38.7 Mb	10 620	49.6%
M. grisea	39.4 Mb	12 841	52.0%
A. fumigatus	29.4 MB	9 926	49.9%
A. nidulans	30.1 Mb	10 701	50.3%
K. thermotolerans	10.4 Mb	5 350	47.0%
K. lactis	10.6 Mb	5 300	N/A*
S. cerevisiae	12.0 Mb	5 885	38.3%
P. chrysosporium	34.5 Mb	10 048	56.8%
T. reesei	33.9 Mb	9 129	52.0%

N/A*: Not available

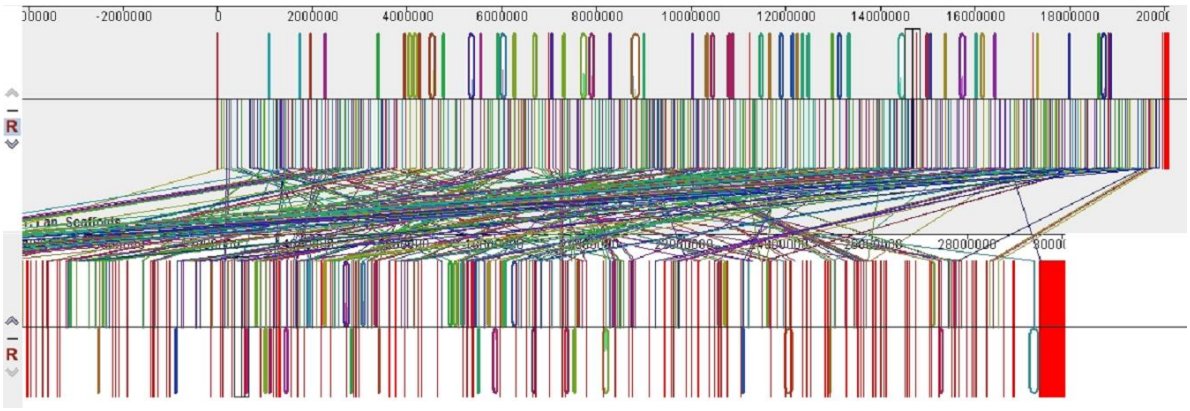

Figure 17: Synteny analysis of *T. lanuginosus* vs *A. nidulans* genomes using WebACT software.

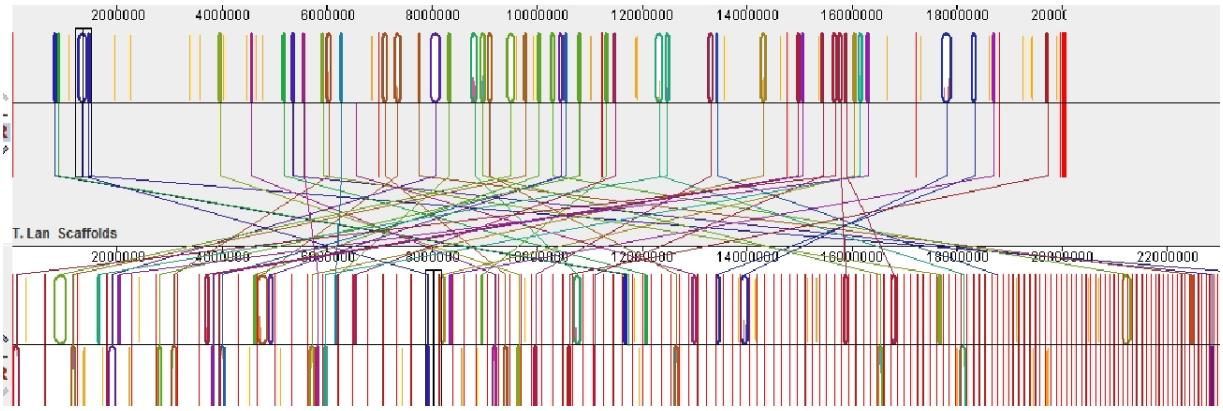

Figure 18: Synteny analysis of *T. lanuginosus* vs *A. fumigatus* genomes using WebACT software.

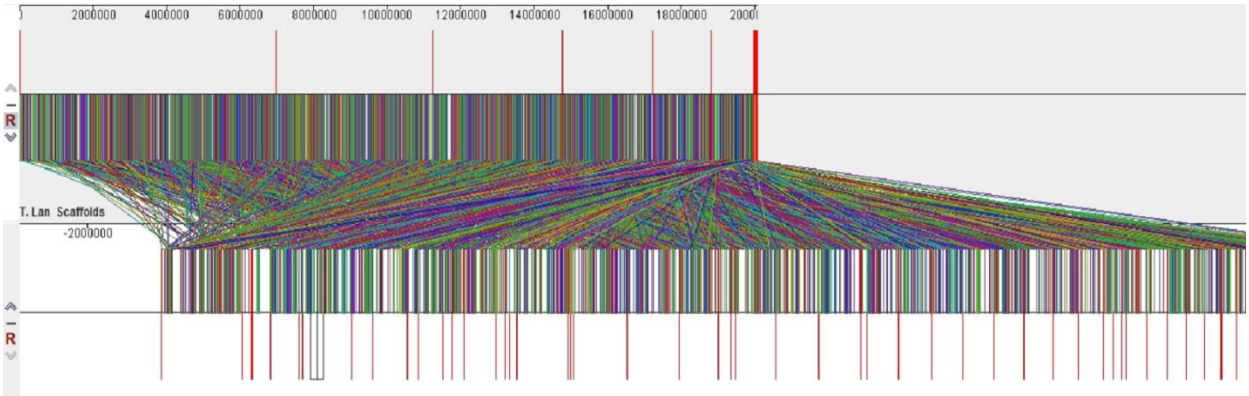

Figure 19: Synteny analysis of *T. lanuginosus* vs *T. reesei* genomes using WebACT software.

2.3.4 General metabolism

2.3.4.1Cluster of Orthologous Groups (KOG/COG) analysis

From a total of 5 105 predicted genes, 88% (4480) were mapped in the COG database (**Figure 20**). The predicted genes were divided into 4 groups *viz*., information storage and processing, cellular processes and signaling, metabolism and poorly-characterized genes. The cellular processes and signaling group had the highest number of genes with 33% followed by metabolism COGs representing 26 % while the information storage and processing group had 23%. The number of proteins poorly-characterized or with unknown function was 18%.

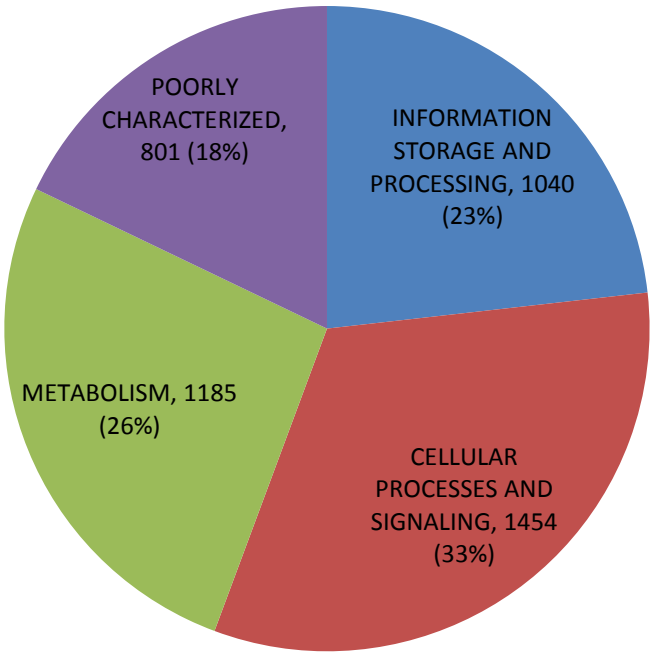

Figure 20: Annotation of *T. lanuginosus* genes into major groups represented in COG analysis. A total of 4480 proteins were annotated by KOG/COG. The proteins were assigned into different KOG/COG major functional categories.

2.3.4.2 Information, storage and processing

Each COG group was further analyzed in order to understand the distribution with the group. COG group for information, storage and processing were divided into five main components (**Figure 21**). The first, the translation and biogenesis-related gene sub-group, constituted 30% of the group with 310 genes followed by transcription with 27% (277 genes) and RNA processing and modification with 20% (214 genes). The last two sub-groups constituted a small part of this group with replication, recombination and repair sub-group making up 15% (160), while the chromatin structure and dynamics sub-group was the lowest at 8% (79).

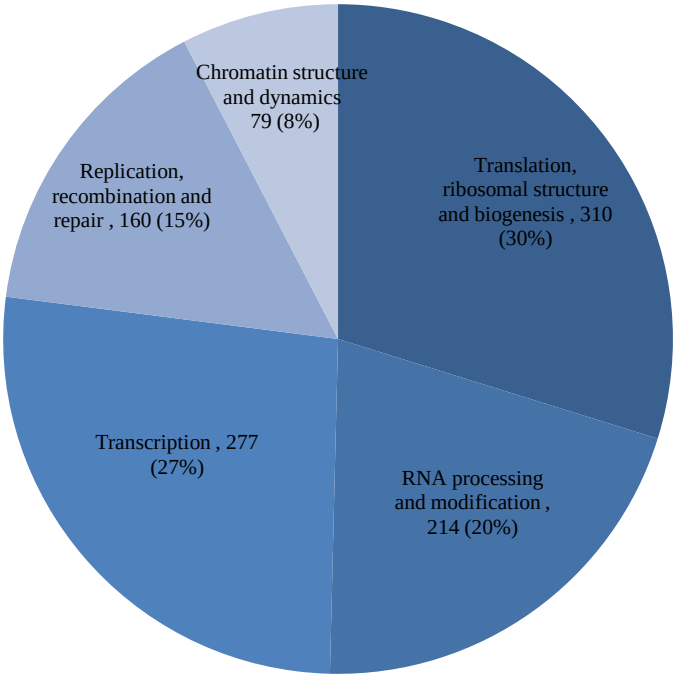

Figure 21: *T. lanuginosus* genes that belong in information storage and processing group. The sub-groups are presented both in percentage and the actual number of genes. The sub- groups found in information storage and processing belong in general to DNA, RNA and protein synthesis.

2.3.4.3 Cellular Processes and signaling

The cellular processes and signaling groups was divided into 8 sub-groups (**Figure 22)**. Post-translational modification, protein turnover and chaperons was the sub-group that contained the most genes with 27% (392), followed by the intracellular trafficking, secretion and vesicular transport sub-group with 22% (324) and the signal transduction mechanisms, with 22% (315 genes). Cell cycle control, cell division, chromosome and partitioning sub-group and cytoskeleton sub-group constituted 11% each, with 159 and 154 genes, respectively. The last 4 clusters constituted a small part of this group with the motility sub-group and cell wall/membrane/envelop biogenesis sub-group making up 4%, the nuclear structure sub-group 2% and defense mechanism sub-group with 1% (19) of the total of 1454 genes. Only 7 genes encoding extracellular structures were found.

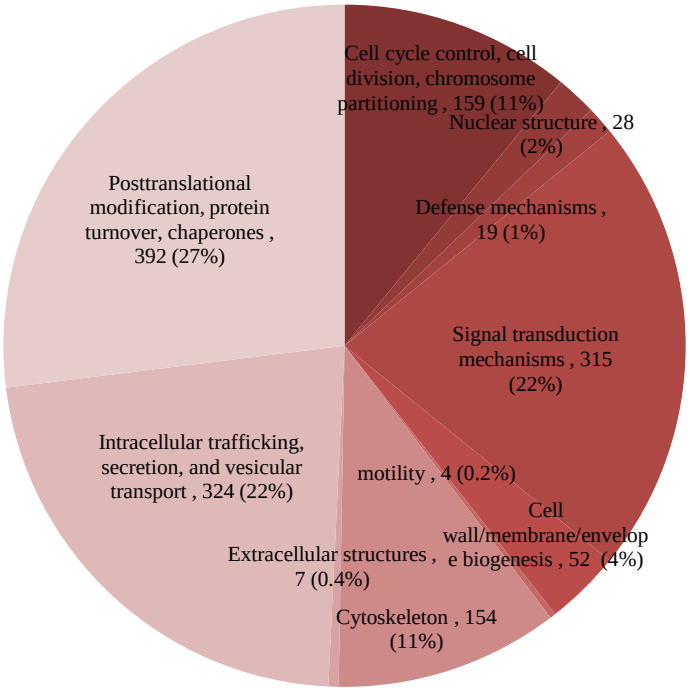

Figure 22: *T. lanuginosus* genes that belong to the cellular processes and signaling COGs. The sub-groups are presented both in percentage and the actual number of genes**.**

2.3.4.4 Metabolism

For the metabolism group, about 70% of the genes were divided between energy production and conversion (19%-228 genes) amino acid transport and metabolism (18%-217 genes) lipid transport and metabolism (17%-200) genes and carbohydrate transport and metabolism (15% - 178 genes). The secondary metabolites biosynthesis, transport and catabolism sub-group and inorganic ion transport and metabolism sub-group constituted 9% each, with 107 and 104 genes respectively. The transportation and metabolism of co-enzymes and nucleotide sub-groups contained 82 genes (7%) and 69 genes (6%), respectively (**Figure 23**).

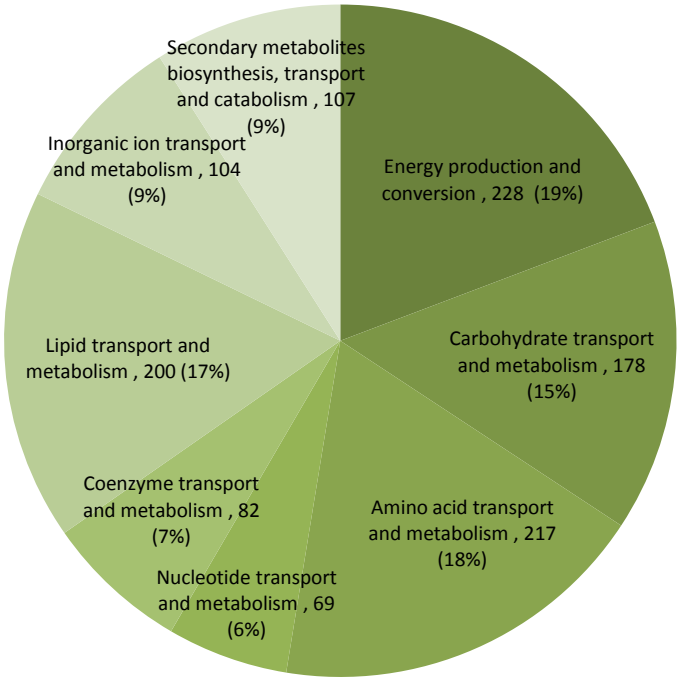

Figure 23: *T. lanuginosus* genes that belong in metabolism. The sub-groups are presented both in percentage and the actual number of genes**.**

2.3.4.5 Poorly -Characterized genes

The poorly-characterized genes group was divided into 2 sub-groups. The first sub-group constituting 70% (560 genes) of the group, represented genes that were assigned with general functions that were not linked to specific cellular processes. The second sub-group had genes with unknown function which constituted 241 genes (**Figure 24**).

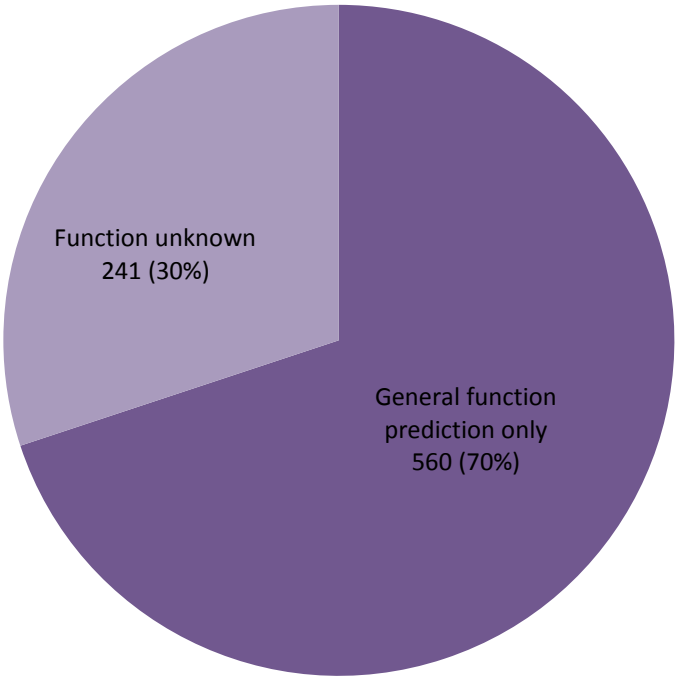

Figure 24: *T. lanuginosus* genes that belong in poorly characterized genes. The sub-groups are presented both in percentage and the actual number of genes. The sub- groups are genes with unknown function or general predictions.

2.3.5 Detection of CAZy family protein

2.3.5.1 Comparison with other fungi

The presence of carbohydrate-active modules (glycosylhydrolases, glycosyltransferases, polysaccharide lyases and carbohydrate esterases) and their associated carbohydrate binding modules (CBM) in genes from *T. lanuginosus* genome was assessed, as *T. lanuginosus* is commonly associated with carbohydrate-rich composting plant material. The total number of predicted proteins that were identified using the CAZY database was 224 (**Table 11**). *T. lanuginosus* contains 94 genes encoding glycoside hydrolases (GH) compared to more than 200 glycoside hydrolyases encoding genes found in other filamentous fungi. When glycosyltransferases (GT) were compared, *T. lanuginosus* contained 89 genes: similar to a number of other filamentous fungi. The number of GT versus GH genes in *T. lanuginosus* was similar, which is unique as other filamentous fungi have double the number of GH genes compared to GT genes. The enzymes involved in plant polysaccharide de-polymerization frequently carry a CBM joined to the catalytic domain. The number of CBM-containing proteins in the *T. lanuginosus* genome was also low (27) when compared to *A. fumigatus*, *F. graminearum* and *M. grisea*, but similar to *T. reesei*, *A. nidulans* and *A. orzyae*. Similarly, *T. lanuginosus* has the lowest number of carbohydrate esterases (12) and polysaccharide lyases (2) among the all filamentous fungi analyzed except for *T. reesei* which has similar numbers i.e, 16 carbohydrate esterases (CE) and 3 polysaccharide lyases. When the total number of CAZy proteins relative to the entire proteome is considered, however, the percentage in *T. lanuginosus* (4.38 %) was greater than most fungi, except for *A. fumigatus*.

Table 11: Comparison of CAZy enzyme families, by class, in the selected fungal genomes

Species	GH	GT	CBM	CE	PL	Total	%
A. nidulans	247	91	36	29	19	422	3.94
A. fumigatus	263	103	55	29	13	463	4.66
A. oryzae	285	114	30	26	21	476	3.94
T. lanuginosus	**94**	**89**	**27**	**12**	**2**	**224**	**4.38**
F. graminearum	243	110	61	42	20	476	4.08
M. grisea	231	94	58	47	4	434	3.37
N. crassa	171	26	39	21	3	260	2.44
T. reesei	200	103	36	16	3	358	3.92
S. cerevisiae	47	67	12	3	0	129	2.19

Enzymes: GH, glycoside hydrolase; GT, glycosyltransferase; CBM, carbohydrate-binding Module; CE, carbohydrate esterase; PL, polysaccharide lyase.

2.3.5.2 Cellulose-degrading genes

Although *T. lanuginosus* has been previously described as non-cellulolytic and it was suggested that it probably relies on commensal relationship in composts with cellulolytic fungi (Purkarthofer et al., 1993), cellulase-degrading genes were still identified. In this study, four, endo-ß-1,4-glucanases (cellulases), belonging to family 61 were identified in a total number of 96 glucoside hydrolyases. Three were similar to *T. reesei* cellulases (1, 3 & 4) and the other to *Aspergillus kawachi* (**Figure 25**). Alignment and phylogeny comparison confirmed that a *Trichoderma* cellulase shared ancestry as to cellulases 1, 3 and 4 from *Thermomyces* (**Figure 26**) while cellulase 2 is similar to *A. kawachi* (**Figure 27**) (Druzhinina et al., 2006; Gu, 2009; Martinez et al., 2009). Furthermore, 4 glucan 1,3-ß-glucosidases were identified belonging to GH5 which alsoplay a role in cellulose degradation (**Table 11**).

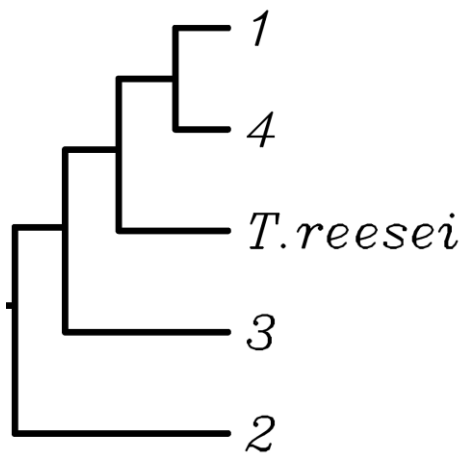

Figure 25: Phylogeny of cellulase enzymes found in *T. lanuginosus* created with ClustalW. The cellulases (1-4) were compared with the cellulase gene from *T. reesei* downloaded from the NCBI database.

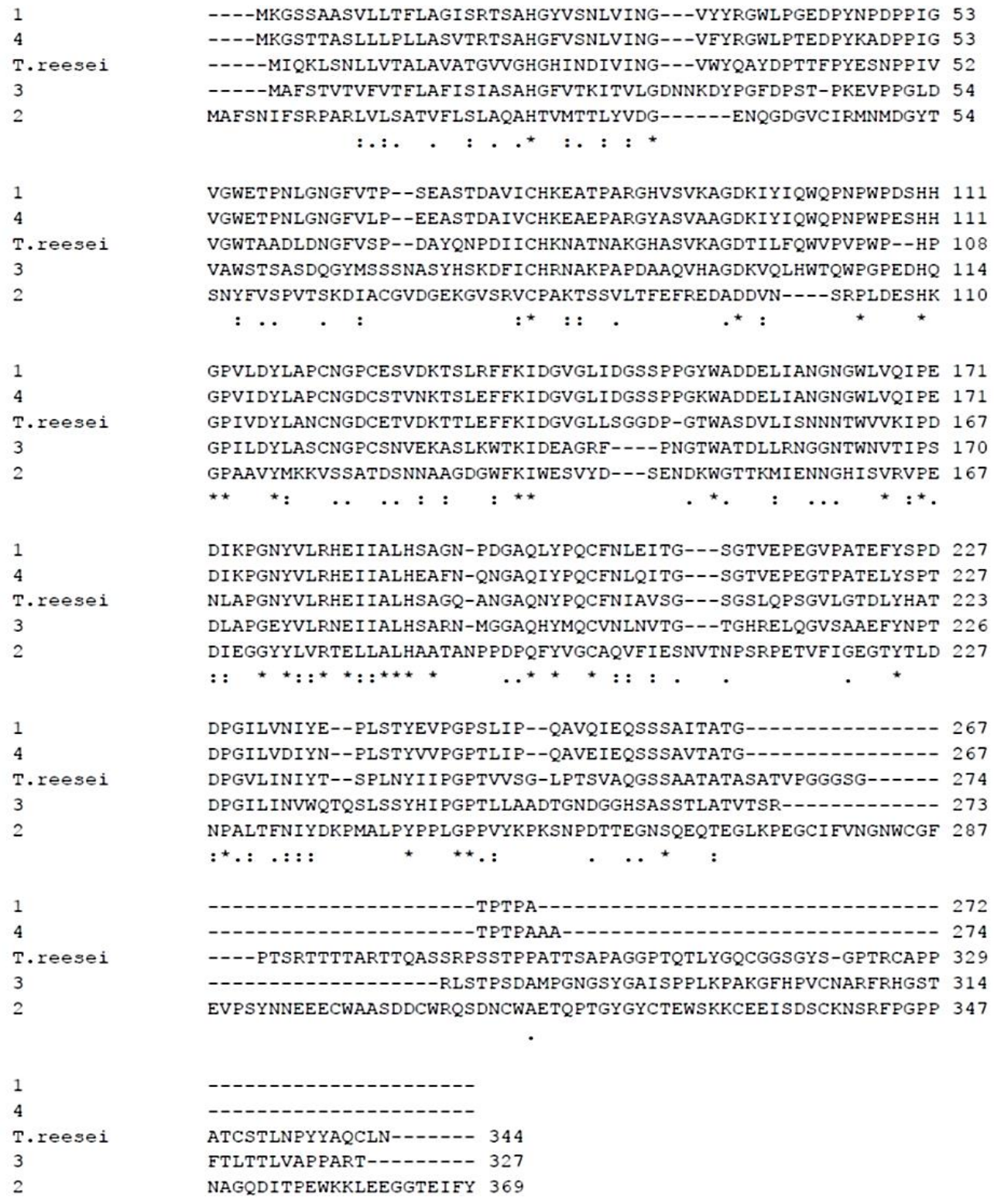

Figure 26: Alignment of putative *T. lanuginosus* cellulases with a cellulase *T. reesei*.

Figure 27: Alignment of a putative cellulase number 2 from *T. lanuginosus* compared with a cellulase gene from *A. kawachi*.

2.3.5.3 Xylan-degrading genes

Proteins that are involved in hemicelluloses degradation, which include glucomannan (mannose polymer) in softwoods and xylan (xylose polymer) in hardwoods were also analyzed. De-polymerization of xylan and xylo-glucan produces D-xylose, while, glucomannan is a straight-chain polymer that contains component sugars which are β-(1→4)-linked D-mannose and D-glucose. Thus, a variety of enzymes is required for the degradation of hemicellulose, unlike cellulose hydrolysis where fewer enzymes are needed. The ability of this thermophilic fungus to

hydrolyze xylans has been previously demonstrated (Singh et al., 2000a). Unexpectedly, only one gene was identified as an endo-β-1,4 xylanase which belonged to the GH11 family of proteins. While xylanase produces random hydrolysis of xylan, xylosidases with exo-activity are required for the conversion of this polymer to xylose monomers. Two genes were predicted to be xylosidase proteins. All possible candidates for proteins involved in substrates associated with other carbohydrates are listed in **Table 12** and **Table S2** (appendix).

Table 12: Selected enzymes that can degrade various carbohydrate substrates

Gene ID	EC Number	Function
NOKS4_001572-RA	EC:3.2.1.58	1,3-ß-glucosidase
NOKS4_001528-RA	EC:3.2.1.106	mannosyl-oligosaccharide glucosidase
NOKS4_004239-RA	EC:3.2.1.58	exo-ß-1,3-glucanase
NOKS4_003831-RA	EC:3.2.1.14	chitinase
NOKS4_002292-RA	EC:3.2.1.113	α-1,2-mannosidase
NOKS4_004387-RA	EC:3.2.1.84	α-1,3-glucosidase
NOKS4_000650-RA	EC:3.2.1.20	α-glucosidase
NOKS4_002370-RA	EC:3.2.1.8	endo-1,4-ß-xylanase
NOKS4_004873-RA	EC:3.2.1.21	ß-glucosidase
NOKS4_003394-RA	EC:3.2.1.25	ß-mannosidase
NOKS4_001716-RA	EC:3.2.1.1	alpha-amylase

2.3.6 Thermostability and thermo-tolerant features

2.3.6.1 DNA stability

The GC content of *T. lanuginosus* was measured at 52% which is high for filamentous fungi (**Table 8**). However, when the GC content for coding regions of the genome was calculated, it was significantly higher than that of the whole genome measuring 56%, with 42% of the genes above the average of 56%. *T. lanuginosus* is predicted to be capable of histone (Chapter 3) acetylation/deacetylation and has a high number of methylases. Histone acetylation/deacetylation

and methylation play an important role in packing and condensation (tight packaging) of DNA (Nowak and Corces, 2004) which is another strategy for stabilizing DNA. This fungus possesses the ability to performing poly-ADP-ribosylation, which involves the addition of one or more ADP-ribose moieties to a protein (Belenky et al., 2007). It also plays an important role in cell signaling and the control of many cell processes, including DNA repair and apoptosis (Berger et al., 2004, Corda and Di Girolamo, 2003). Also, 16 genes of possible DNAJ analogs are present in the genome.

2.3.6.2 Membrane stability and electron chain transport

T. lanuginosus genome contains a high number of methyltransferases. The 131 S-adenosyl-l-methionine (SAM)-dependent methyltransferases, which is the highest among the fungal species used for comparison in this study. The SAM-dependent methyltransferases represent a diverse and biologically important class of enzymes. These enzymes utilize the ubiquitous methyl donor SAM as a cofactor to methylate proteins, small molecules, lipids, and nucleic acids. Most interesting is that SAM-dependent methyltransferases are predominantly ubiquinone/menaquinone (53), phosphoethanolamine (9) and trans-aconitate methyltransferases (7) (**Table 13**).

Menaquinone/ubiquinone (MK/ coenzyme Q10) plays an essential role in the electron transport chain and participates in aerobic cellular respiration, generating energy in the form of ATP. Since its primary function in cells is to generate energy, the highest concentration is found on the inner membrane of the mitochondrion. Phosphatidylethanolamine methyltransferase (PEMT) catalyzes the first step in the conversion of phosphatidylethanolamine to phosphatidylcholine during the methylation pathway of phosphatidylcholine biosynthesis. Phosphatidylcholine is a major constituent of cell membranes.

Table 13: Comparative analysis of methyltransferase genes found in the *T. lanuginosus* genome and other selected fungal species (MTs, methyltransferases; MK, menaquinone; ubiE, ubiquinone; PETMs, phosphatidylethanolamine methyltransferases**)**

Species	Total MTs	MK/ubiE	PETMs	Growth temperature (°C)
A. niger	61	2	3	30
A. fumigatus	58	7	6	30
A. oryzae	44	1	4	30
T. lanuginosus	**131**	**53**	**9**	**50**
F. graminearum	24	3	3	30
M. grisea	61	2	2	30
N. crassa	62	1	1	30
S. cerevisiae	67	2	2	30

2.3.6.3 Cell stability at high temperature

Another mechanism that is used by microorganisms to resist adverse environmental conditions is the synthesis and accumulation of organic compounds that may serve a protective function. Trehalose utilisation has been widely reported as a part of the physiological adaptation to various environmental stresses in yeasts and filamentous fungi (Fillinger et al., 2001; Elbein et al., 2003). Metabolic reconstruction of the trehalose synthesis pathway indicates that trehalose could be possibly synthesised by *T. lanuginosus*. High assimilation and increased biomass was observed when the fungus was grown on trehalose (refer to chapter 3, **Figure 28**).

Aerobically growing cells are continuously challenged by potent oxidants produced during normal cellular metabolism. However, they can cause cell damage when present at toxic levels. Aerobic organisms possess extensive antioxidant systems to regulate oxidant levels. *T. lanuginosus* possesses IscX methyltransferase, which is an active component that is part of the Iron-sulphur cluster (ISC) system. It can bind to iron and may function as an iron donor in the

assembly of FeS clusters. Iron-sulphur (FeS) clusters are important cofactors for numerous proteins involved in electron transfer, in redox and non-redox catalysis, in gene regulation, and as sensors of oxygen and iron. *T. lanuginosus* also has mechanisms to protect the cell against oxidative damage. Five genes coding for a superoxide dismutases and five putative peroxiredoxin proteins were identified, which are enzymes that catalyze the dismutation of superoxide into oxygen and hydrogen peroxide (**Table 14).**

Table 14: Dimutases found in *T. lanuginosus*

Dismutase	**Prediction**	**Organism**
Superoxide dismutase 1	Similar to superoxide dismutase [Cu-Zn]	*Debaryomyces hansenii*
Superoxide dismutase 2	Similar to CCS1: superoxide dismutase one copper chaperone	*Debaryomyces hansenii*
Superoxide dismutase 3	Similar to SOD2: superoxide dismutase [Mn], mitochondrial	*Saccharomyces cerevisiae*
Superoxide dismutase 4	Similar to SOD: superoxide dismutase [Mn], mitochondrial	*Penicillium chrysogenum*
Superoxide dismutase 5	Similar to *sod*C: superoxide dismutase	*Neosartorya fumigata*

Reconstruction of the metabolic pathways of *T. lanuginosus* revealed that this fungus has a ubiquitin degradation system called N-end rule which is facilitated by arginine-tRNA-protein transferase. The ubiquitin system also plays an important role in the degradation of denatured proteins following environmental stress such as nutrient limitation, heat shock, and heavy metal exposure (Staszczak, 2008).

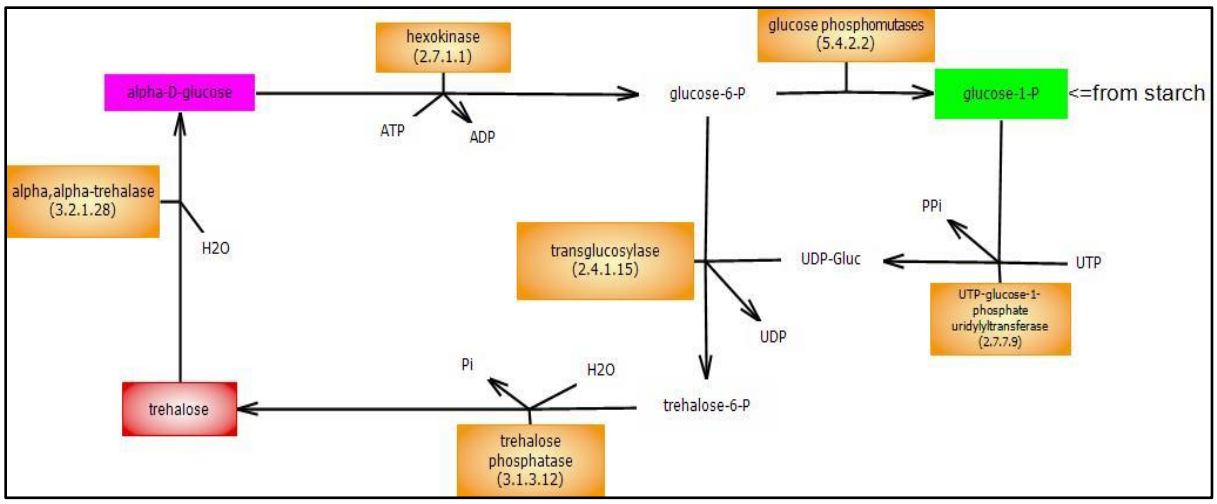

Figure 28: Reconstructed pathway for trehalose synthesis in *T. lanuginosus*. The pathways were constructed using Pathway Studio (Araidne Genomics).

2.3.6.4 Heat shock proteins

Another strategy used by cells to survive at high temperature is to express a group of proteins called heat shock proteins (Hsps). Heat shock proteins general function as intracellular chaperones for other proteins. They play an essential role in protein interactions such as folding and assisting in the establishment of proper protein conformation (shape) and prevention of unwanted protein aggregation. The genome of *T. lanuginosus* contains 19 genes coding for heat shock proteins, which was higher than the number found fin other fungi: *N. crassa* (11), *M. grisea* (11), *F. graminearum* (10), *A. nidulans* (9), *A. fumigatus* (6) and *A. oryzae* (2). Analysis of the diversity of these heat shock genes showed that they are mainly contained in three clusters with all heat shock proteins showing similarity to other fungal heat shock proteins (**Figure 29**). Other interesting temperature-dependent proteins that were identified are in a group consisting of several bacterial specific low-temperature-requirement A (LtrA) protein sequences which have been found to be essential for growth at low temperatures.

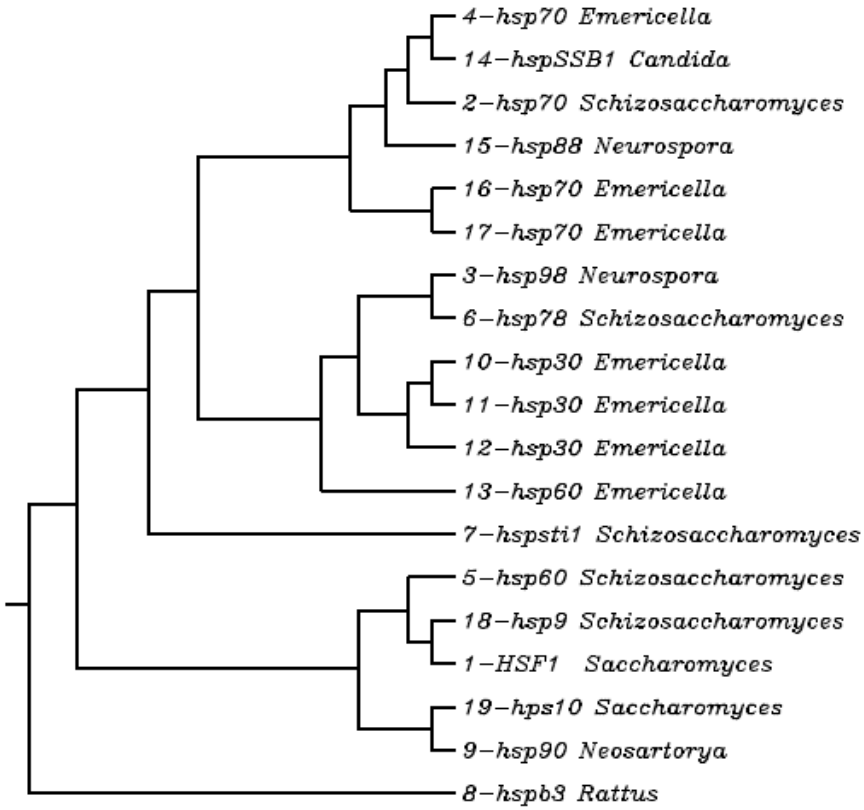

Figure 29: Phylogeny of heat shock protein analogs found in *T. lanuginosus* created with ClustalW.

2.4 DISCUSSION

Fungi are among the most widely used microorganisms for the benefit of humans in areas of agriculture, food, environmentally-useful and now biofuels. There is therefore raised a great amount of interest in bio-prospecting for industrially-useful fungi. Fungi exhibit a diverse range of metabolic activities, particularly as saprophytes, pathogens or symbionts. Saprophytic fungi acquire nutrients from dead organic matter e.g. plant waste, thus playing a major role in the degradation of cellulose and hemicellulose.

As mentioned previously, fungi are also of great importance to industry due to their ability to produce many industrially important drugs and enzymes. Filamentous fungi, especially, have become excellent production workhorses for industrial enzymes due to their natural ability to secrete large amounts of these enzymes. They are usually the preferred expression host for heterologous proteins in eukaryotic systems. Genomics has enhanced the ability to either search for new enzymes or understand how to improve existing systems (Andersen *et al.*, 2011). However filamentous fungal genomes have been elucidated at a slower rate compared to bacteria (Grossetête *et al.*, 2010) and genomes from thermophilic organisms are even more scarce.

T. lanuginosus is one such thermophilic, filamentous fungus whose enzymes have been used in different industries. This thermophilic filamentous fungus produces a wide range of thermostable enzymes including a large group of carbohydrate hydrolysases. These enzymes include: amylase, glucoamylase, xylanase, lipase, phytase, protease and chitinase (Singh *et al.*, 2003). These thermostable enzymes can be applied in different industries including the food industry for the production of sugar syrup, animal feed industry, pulp and paper industry and bioremediation/bio-conversion of waste industry (Kuttanpillai *et al.*, 2009b; Manimaran *et al.*, 2009; Puchart *et al.*, 1999; Singh *et al.*, 2000a). Therefore, this study focused on elucidation the genome sequence of *T. lanuginosus* SSBP, the highest xylanase hyper-producer among the *Thermomyces* xylanase-producing strains (Singh *et al.*, 2000a).

To do this, high quality genomic DNA and total RNA was isolated to ensure production of good-quality sequence reads and in-turn, good genome assembly. High quality starting material is essential for reliable next-generation sequencing results and accurate data analysis. In this study, a standard phenol/chloroform method was used for extraction (Melo *et al.*, 2006). The

genomic DNA appeared to be intact, in good condition and of good quality (**Figure 13**). Although transcriptome assembly and analysis was not the main goal of this study, mRNA sequences were important as they were used to validate gene predictions (Cantarel *et al.*, 2011). This required the isolation of the total RNA for the conversion of mRNA to cDNA, then sequencing. To make sure that the RNA samples contain only RNA molecules, the RNA samples were treated with DNase enzymes which only cleaves double-stranded DNA (**Figure 14**). Also, validation for the RNA integrity (no degradation) was done using the Agilent Bioanalyzer (**Figure 15**). This analytic technique identifies the 18S and 28s RNAs as they are the most abundant species of RNA. After the quality of both DNA and RNA samples were confirmed, sequencing commenced.

In general for model organisms where a wealth of genomic information is available, the massively parallel, short-read sequencing technologies like Illumina/Solexa and ABI SOLiD are most frequently used, as reads can easily be mapped to the reference genome or transcriptome. However, for non-model organisms this is more difficult to do as no reference assembly is available as was the case with *T. lanuginosus*. The most frequently-used method for non-model organisms is the Roche 454 pyrosequencing platform because the long reads generated are much easier to assemble using *de novo* assembly (DiGuistini *et al.*, 2009; Kumar and Blaxter, 2010). In this study, the platforms that were used for read generation was Roche 454 and Illumina/Solexa platforms. With the 454 Roche sequencing, both shotgun and pair-end genomic libraries were used while for Illumina pair-end and pair mate were used. The number of reads produced for the assembly was high to ensure that the coverage was good. This is a pre-requisite for *de novo* assembly. The depth coverage for 454 sequencing reached 48-fold while for Illumina, it was much higher at 138-fold (**Table 7**). The addition of Illumina technology increased the coverage depth, thus improving the overall assembly profile. The coverage for this genome was higher than most assemblies produced using the *de novo* assembling method. The repeat reads were less than 0.5% of the total reads generated. This was comparable to other reported assemblies like that of *Fusarium* species (Ma *et al.*, 2010; Qiang *et al.*, 2011). The transcriptome produced using Illumina technology, had fewer reads and therefore had a lower coverage compared to genomic DNA. However, in this study this was not major hurdle as transcriptome sequencing data was only to support genome assembly and annotation.

The assembly produced a 23.3 Mb *T. lanuginosus* genome sequence that was created by *de novo* assembly (**Table 8**). Using Newbler, 98% of the sequencing data generated was assembled into the genome. Although other assembly platforms are available, Newbler was chosen for its versatility and ability to assemble reads produced from different sequencing platforms. Also, comparison of the used assembly platforms for *de novo* assembly showed that Newbler produced the best assembly output with longer contigs, better alignment and was generally faster and easier to use (Kumar and Blaxter, 2010). This was also true for the *T. lanuginosus* genome. The assembly has long-range continuity as evidenced by the high N50 for the scaffold size being over 4 Mb. The high continuity compared favourably against other reported genomes with most having N50 sacffold sizes of around 2 Mb or less. Most of the assembly (90%) is contained in the 6 largest scaffolds, and the rest in 24 smaller scaffolds (**Table 9**). This suggests that the *T. lanuginosus* genome might be contained within 6 chromosomes but more needs to be done to confirm this. An alternate strategy, Pulse-field gel electrophoresis (PFGE) can be used to separate and identify the exact number of chromosomes (Bellis *et al.*, 1987). In addition, isolation of each chromosome after PFGE and DNA sequencing can localise identified genes onto each of these chromosomes. The accuracy of the assembly was further supported by mapping nearly 95% of the transcriptome reads to the genome, since transcriptome is snap-shot of the expression profile of a particular organism at a specific time. Also, this further validated the genome assembly and gene prediction, which very few genome assembly projects have achieved.

The genome of *T. lanuginosus* (23. 3 Mb) when compared to other filamentous fungi (generally between 30-40 Mb) is relatively small (**Table 10**). *A. fumigatus* and *A. niger* have small genomes but were still more than 20% larger that *T. lanuginosus* (Galagan *et al.*, 2005a; Galagan *et al.*, 2003; Goffeau *et al.*, 1996; Ma *et al.*, 2010; Machida M *et al.*, 2005; Martinez *et al.*, 2008). *T. lanuginosus* had a total number of 5105 genes which is lower compared to other filamentous fungi. This was surprising since the average number of genes reported in filamentous fungi is higher. 12 *T. melanosporum* genome, a 125 megabase (Mb) is the largest sequenced fungal genome to date but had only 7,496 protein coding genes identified, showing that a large genome does not necessary produce a large proteome (Martin *et al.*, 2010).

Synteny is defined as regions where DNA sequence or gene order is conserved between genomes. Closely-related species show extensive synteny with each other (Hane *et al.*, 2011). It is also used to identify the orthologous regions in different genomes. The inversion of the genome, especially around the origin of replication, is one of the dominant mechanisms for genome rearrangement. The factors affecting rearrangement processes, while not understood, are believed to be linked to abundant mobile elements and the repair/recombination system of the genome itself (Novichkov *et al.*, 2009). *T. lanuginosus* genome had very low synteny with the filamentous fungi analysed but exhibited some synteny with *T. reesei* but in a reverse direction (**Figure 19**). Comparison of many genome sequences has indicated divergence of fungi at the genome level, despite apparent morphological and physiological similarities (Hibbett *et al.*, 2007; Vilgalys and Johnson, 1987). Further, an analysis of fungal genomes has revealed a rapid loss of conserved synteny over a relatively short period (Galagan *et al.*, 2005b). A comparison of three *Aspergillus* species, *A. nidulans*, *A. fumigatus*, and *A. oryzae*, had around 60% amino acid similarities when synteny analysis was applied (Galagan *et al.*, 2005a).

The average gene length in *T. lanuginosus* is 1615 base bp containing 3.7 exons and 3 introns per gene (**Table 8**). The average gene length of most filamentous fungi ranges from 1,300 to 1700 bp. *T. lanuginosus* lies on the upper end of this scale, as well as the two thermophilic fungi use in the comparison, *T. terrestris* and *M. thermophile* with average gene lengths of 1649 and 1733 bp respectively (Berka *et al.*, 2011). Comparing this to some well-known mesophilic fungi, revealed that the length was generally low e.g. *A. fumigatus* (1226 bp), *Fusaruim oxysporum* (1292 bp), *Fusarium verticillioides* (1397 bp), *F. graminearum* (1355 bp) while others like *N. crassa* and *P. chrysogenum* were similar to thermophiles. This suggests that gene length may not be crucial to surviving high temperatures (Galagan *et al.*, 2003; Ma *et al.*, 2010; Marco *et al.*, 2008; Nierman *et al.*, 2005). The number of exons contained with the genes was relatively high for *T. lanuginosus* at 3.7 while *T. terrestris* and *M. thermophile* both had an average of two exons in length and other filamentous fungi ranged from around 2-3.6 exons (Andersen *et al.*, 2011; Berka *et al.*, 2011; DiGuistini *et al.*, 2009; Galagan *et al.*, 2005a) . The GC content of *T. lanuginosus* was generally similar to other filamentous fungal species which was measured to be around 52%. This was surprising as thermophilic organisms a have higher GC content than their mesophilic counter-parts. But deeper analysis revealed that the GC content in the coding region was much higher. This suggests the importance of maintaining DNA stability and integrity in

essential regions. Als,o two thermophilic fungi had similar GC content to that of *T. lanuginosus*. The higher GC percentages which is normally seen in bacteria, could be due to fewer strategies being available to bacterial organisms for maintenance of DNA integrity while eukaryotic systems are more sophisticated.

COG analysis revealed that general metabolism in *T. lanuginosus* is well distributed among the predicted genes. Eighty eight percent (4480) of the genes were clustered in the COG group with 625 genes having no classification are functionally uncharacterized COGs (**Figure 20**). During annotation a large fraction of ORF is assigned as conserved hypothetical protein or unknown function thus not classified in any of the COG clusters(Doerks *et al.*, 2012). However, this can be overcome with the manual curation of the genome annotation. For *T. lanuginosus*,this remains an on-going process as each gene the computer predictions for each gene has to be verified manually. But the high number of genes that were assigned into COG groups suggests that most of the gene predictions were accurate. Analysis of cellular processes and signaling COGs, which was the highest group, revealed that the post-translation, modification; protein turnover and chaperones sub-groups contained the most genes (**Figure 22**). In eukaryotic organisms, post-translation modification is important to the stability and functionality of proteins. But more important for thermophilic organisms is the ability to utilise or convert degraded or mis-folded peptides to amino acids, which are used as building blocks for new protein (Maheshwari *et al.*, 2000). This is vital at high temperatures because the amount of denatured protein is higher than inactive proteins which can still function when the temperatures become favourable again. The use of chaperones in high temperature is well reported and also it has been shown in this study that *T. lanuginosus* produces heat shock proteins. The other important group is genes that are used in defence mechanisms. In this fungus, the number of these genes is relatively low with only 19 genes. This may be due to this organism facing little threat or competition in nature as very few organisms can compete vigorously at high temperatures (Maheshwari *et al.*, 2000).

Metabolism of carbohydrate polymers or monomers is vital for the functioning of the cell as this provides energy and building blocks for other processes in the cell. In *T. lanuginosus,* genes responsible for metabolism and transport of lipids, carbohydrate, amino acid and energy production and conversion are equally represented (**Figure 23**). These genes are responsible for energy generation and are quite efficient in this organism as biomass production by*T.*

lanuginosus can be produced easily using different substrates. The COG clusters for information storage and processing showed that nearly 60% of these genes are used in protein synthesis (translation and transcription) with another 20% for RNA processing and modification e.g. mRNA splicing (**Figure 21**). This is in line with other COGs which show that functions that are related to protein synthesis of are important for this organism. Although the COG analysis revealed that this organism metabolises a variety of compounds, the focus in this study was on carbohydrate-based compounds.

It has been well documented that *T. lanuginosus* can use the following carbohydrates as a carbon sources: glucose, sucrose, galactose, arabinose, mannose, xylose, fructose, trehalose and starch (Kuttanpillai *et al.*, 2009a; Purkarthofer and Steiner, 1995). Some of these can be used for biofuel production. Ethanol is a potential substitute of gasoline as fuel for vehicles. It is currently derived from starch-rich substrates such as maize/corn and plants with high sugar contents, such as sugarcane (Nigam and Singh, 2011; Somerville, 2007). Most lignocellulosic material including woods, grasses, and agricultural residues which are primarily composed of cellulose, hemicelluloses and lignin, can be used in biofuel production instead of food crops (Peter Bosch, 2010). Currently, however, lignocellulose cannot be directly fermented to ethanol at commercially-viable scale. Before fermentation, the process requires breaking down the lignocellulose material into simple sugars which contributes to the high cost of ethanol production. Thus, the efficient conversion of plant biomass is essential in providing affordable biofuel (Nigam and Singh, 2011). CAZy proteins can fulfill this role because these proteins have the ability to cleave or add monomers to polysaccharides and many of them have a major importance in biotechnology including biofuels (Zhao *et al.*, 2013). Given the importance of this protein family to the biotechnology community, we performed a detailed analysis of the CAZy protein found in *T. lanuginosus*. Although this fungus has been widely reports to produce xylanase and amylase (Kuttanpillai *et al.*, 2009b; Puchart *et al.*, 1999; Purkarthofer *et al.*, 1993; Singh *et al.*, 2000a), little is known about proteins involved in metabolism of other carbohydrate compounds.

The total number of predicted proteins identified using the CAZy database as belonging to this group of proteins was 224 (**Table 11**). This was surprising low as it was expected that *T. lanuginosus* would contain a larger number of genes involved in digesting cell wall

polysaccharides. Other filamentous fungi used for the comparison contained more than 400 carbohydrate-related proteins. *T. lanuginosus* is not the only filamentous fungus with a low number of CAZy proteins. *N. crassa* had 260 and *T. reesei* contains 386 genes, although both were higher than *T. lanuginosus*. Due to the overall low number of these proteins in *T. lanuginosus*, the number of predicted genes encoding glycoside hydrolases (GH) was also low (94), which was lower by more than 50% compared to an average of 200 GH-encoding genes found in other filamentous fungi. But this was still almost twice the amount found in the eukaryotic yeast, *S. cerevisiae*. This low number of hydrolases can be overcome by a number of strategies. The first strategy is that a single protein can be highly expressed using strong inducer and promoter rather than having several genes that perform similar functions but with weak expression (Archer *et al.*, 1994). Secondly, the catalytic activity of a particular protein can be relatively high so that it can satisfy the metabolic needs of an organism. Thirdly, signaling and transport mechanisms can be made more efficient to transport a particular protein to the substrate (Gouka *et al.*, 1997). All three strategies could be the case in *T. lanuginosus* in known to produce large amount of enzymes *viz.*, xylanase and amylases (Kunamneni *et al.*, 2005; Singh *et al.*, 2000a). The number of glycosyltransferases (GT) from *T. lanuginosus* was 89 genes which are at levels similar to other filamentous fungi analysed which implied that signaling mechanisms are as efficient in comparison to other fungi. The enzymes involved in plant polysaccharide de-polymerization frequently carry a CBM joined to the catalytic domain. The number of CBM-containing proteins in the *T. lanuginosus* genome was also low (27) when compared to *A. fumigatus*, *F. graminearum* and *M. grisea*, but similar to *T. reseei*, *A. nidulans* and *A. orzyae*. Again this can also be attributed to the low total number of CAZy proteins in *T. lanuginosus*. In general high levels of CAZy proteins are found in phytopathogens *viz*., *F. graminearum* and *M. grisea*. Similarly, *T. lanuginosus* has the lowest number of carbohydrate esterases (12) and polysaccharide lyases (2) among the all filamentous fungi analyzed except for *T. reseei* which has 16 (EC) and 3 (PL). However, when the total number of CAZy proteins relative to the entire proteome are analysed, the percentage in *T. lanuginosus* (4.38 %) was high with only *A. fumigatus* (4.66%) higher that *T. lanuginosus*. This suggests that *T. lanuginosus* carries enough CAZy protein machinery to drive its metabolism efficiently.

Proteins that are involved in cellulose degradation have become crucial, as any advancement in this area will have a significance influence on the production of biofuels and related costs (Qing

and Wyman, 2011). *T. lanuginosus* has been previously described as non-cellulolytic and it was suggested that it probably relies on commensal relationship in composts with cellulolytic fungi by utilizing some of the sugars generated by the enzymes of these fungi and possibly by also using their mycelila breakdown products (Purkarthofer *et al.*, 1993). An important finding in this genome sequencing project, is the discovery of possible cellulase genes. Four endo-1,4-glucanases, belonging to family 61 were identified among the glycoside hydrolyases (Dodd *et al.*, 2010). More surprisingly, was that phylogenetic comparison showed that three of them share high similarity and identity with a *Trichoderma* cellulase (Cel61A) while the other is similar to *A. kawachi* cellulase (**Figure 25 & 26**). The number of cellulase genes present in *T. lanuginosus* is comparable to *T. reesei*, a well know cellulase super-producer, where four cellulase genes were identified after genome sequencing (Druzhinina *et al.*, 2005; Martinez *et al.*, 2008; Sukharnikov *et al.*, 2011). Previous characterization of Cel61A showed endoglucanase activity on substrates containing β-1,4 glycosidic bonds, like carboxymethylcellulose, hydroxyethylcellulose and β-glucan. This enzyme was able to produce soluble sugars from microcrystalline, cellulose phosphoric acid swollen cellulose, and CMC. However, Cel61A produced small amounts of oligosaccharides from these substrates and activity was several-fold lower compared to Cel7B (Sukharnikov *et al.*, 2011). The low activity of this cellulase could have made cellulase activity difficult to detect thus leading for *T. lanuginosus* to be identified as cellulase-free.

Another group of protein from four genes belonging to GH5 that were identified and have been associated with cellulase degradation is glucan ß-1,3-glucosidase. This enzyme is vital for the assimilation of cellobiose, the biodegradation product of cellulose for cellulolytic microorganisms (Ruel and Joseleau, 1991; Sukharnikov *et al.*, 2011). Moreover ß-glucosidases has been shown to participate in the hydrolysis of cellobiose, which is a major product of cellulose (Igarashi *et al.*, 2003). The indication of presence of cellulose hydrolysis enzymes was further strength by the results obtained when assessing the ability of this organism to utilize different compounds as energy source using the PM technology (**Chapter 3**). When *T. lanuginosus* was grown on cellobiose, which is an oligosaccharide of cellulose, it exhibited good growth and substrate utilization generally similar to starch derived oligosaccharides tested which included maltose.

Beside cellulose, another plant polymer that has important industrial application is hemicellulose. Xylan is the major hemicellulose components of plant cell walls and one of the most abundant polysaccharides in nature. De-polymerization of xylan and xyloglucan produces D-xylose which can be utilised to produce a number of industrial products (Dodd *et al.*, 2011). While, glucomannan is a straight-chain polymer that contains component sugars which are β-(1→4)-linked D-mannose and D-glucose. These heteropolymers can contain galactose, glucose, arabinose, and small amounts of rhamnose, glucuronic acid, methyl glucuronic acid, and galacturonic acid (Svastits-Dücső *et al.*, 2009). The ability of *T. lanuginosus* to hydrolyse xylans has been previously demonstrated with *T. lanuginosus* (Singh *et al.*, 2000a). One gene was identified as β-xylanase belonging to GH11 family, this is surprising because this organism produce the highest amount of extracellular xylanase ever reported (Kuttanpillai *et al.*, 2009b; Singh *et al.*, 2000b). While xylanase randomly cleaves xylan, xylosidases have exo-activity and are required for the conversion of this polymer to xylose monomers and two genes were predicted to be xylosidase proteins. The xylosidase activity has been previously reported for this strain with Singh *et al*., (2003) showed that this strain has xylosidase activity although in relatively low quantities.

When comparing *T. lanuginosus* with other fungi (**Tables 11**), *T. lanuginosus* has the smallest number of genes involved in the degradation of three major plant polysaccharides, cellulose, xylan and pectin. This is unexpected, as the ability of *T. lanuginosus* to grow in plant waste material is highly efficient and is one of the highest producers of hemicellulases. On the other hand the inability to use cellulose material has been reported but eight genes related to this function were identified, although functionality still has to be validated. The fact is, *T. lanuginosus* has a relatively low proteome with few CAZy proteins but sufficient enough to enable the organism to survive in nature with other fungi that degrade cellulose and hemicellulose. This also indicates our poor understanding of plant cell wall degradation and the co-operation and survival of microorganisms in their natural habitats.

Depending on their optimal growth temperature, microorganisms are classified into the following four groupspsychrophiles (-5 to 20°C), mesophiles (15–45°C), thermophiles (45–80°C), and hyperthermophiles (80 °C) (Vieille and Zeikus, 2001). Thermophiles and hyperthermophiles have evolved mechanisms to stabilize proteins at elevated temperatures. Thus, identifying

structural features involved in stability of proteins at high temperature is important, as this information can be used for designing more stable enzymes for industrial processes. In general eukaryotic organisms are less stable at high temperatures than prokaryotic organisms. Among the eukaryotic organisms, only a few species of fungi have the ability to thrive at temperatures between 45 and 55°C (Berka *et al.*, 2011; Maheshwari *et al.*, 2000; Morgenstern *et al.*, 2012; Salar and Aneja, 2007). *T. lanuginosus* is among these few exceptional fungi with the ability to grow well at temperatures close to 60°C, which is the upper limit for these organisms (**Table 3**).

The first difficulty thermophilic organisms encounter is protection of their DNA from instability at high temperatures. There are different mechanisms that have been identified that thermophilic organisms utilise to maintain DNA integrity. One school of thought emphasises that higher GC content in a genome, particularly in the protein coding regions, would be an essential feature to protect DNA from thermal extremes (Singer and Hickey, 2003). This was the case in *T. lanuginosus* with an overall GC content of 52.0 % which is high for eukaryotic organisms. But when the %GC for coding regions of the genome was analyzed, it was significantly higher than that of the whole genome (56%), with 42% of the genes above the average value of 56 %. Although increase in GC content may be important, it alone cannot be responsible for providing thermal stability to DNA. Hence, in the absence of high GC content there must be other features for thermal protection of the genetic material.

The most striking feature about the *T. lanuginosus* genome is the high number of methyltransferases, which is the highest when compared to other fungal genomes used in this study. SAM-depedent methyltransferases catalyze the transfer of a methyl group from a methyl donor (SAM) to nitrogen, oxygen, sulphur or carbon atoms of a large number of biologically active large and small molecules. These modifications play a major role in the regulation of various biological functions such as gene expression, signaling, nuclear division and metabolism. It is also involves methylation of biological molecules, including DNA, RNA, and histones and other proteins.

The *T. lanuginosus* genome encodes 53 ubiquinone/menaquinone methyltransferase genes which have been identified as being responsible for production quinones (**Table 13**). These methyltransferases are involved the conversion of dimethylmenaquinone to menaquinone (MK) or ubiquinone (coenzyme Q10). Menaquinone/ubiquinone (MK/ coenzyme Q10) plays an

essential role in the electron transport chain and participates in aerobic cellular respiration, generating energy in the form of ATP. Since its primary function in cells is in generating energy, the highest concentration is found on the inner membrane of the mitochondrion. Although fungi general have been classified to only possess ubiquinones, most of the methyltransferase genes found in *T. lanuginosus* were closely related to the menaquinone genes generally predominant in extremophilic prokaryotes (Brooijmans *et al.*, 2009; Farrand and Taber, 1973). The surprising high levels of these genes may imply that this organism uses this system to produce surplus energy during optimum conditions, especially in the ever changing conditions of the composting environment. Successive changes also occur in the compost physico-chemical characteristics, such as temperature, moisture content, O_2 and CO_2 levels, along with progressive breakdown of complex compounds to simpler molecules (Tiquia *et al.*, 2002). Studies done on a *Bacillus subtilis* mutant unable to synthesis menaquinone, exhibited reduced O_2 uptake, decreased the levels of cytochrome a, b and c, and decrease in sporulation (Farrand and Taber, 1973; Wu *et al.*, 2009). Therefore, the synthesis of large amounts of quinones in *T. lanuginosus* may be an adaptation process for living in composting environments where temperature and O_2 levels can change drastically, making it difficult to produce energy.

Many organisms vary the fatty acid composition of their membrane phospholipids as a function of growth temperature so that their membrane fluidity is maintained for the optimal functioning of membrane-localized transporters and proteins. An increase in temperature has been correlated to an increase in the proportion of saturated fatty acids incorporated into phospholipids, whereas at lower temperatures, a higher proportion of unsaturated fatty acids are incorporated (Sinesky, 1974). Phosphatidylethanolamine methyltransferase (PEMT) catalyzes the first step in the conversion of phosphatidylethanolamine to phosphatidylcholine during the methylation pathway of phosphatidylcholine biosynthesis. Phosphatidylcholine is a major constituent of cell membranes. Phosphatidylcholine is more commonly found in the exoplasmic or outer leaflet of a cell membrane. Phosphatidylcholine also plays a role in membrane-mediated cell signaling and phosphatidylcholine transfer protein activation of other enzymes. Recent studies have begun to indicate that choline also plays an important role in hyphal and mycelial morphology. The phospholipid is composed of a choline head group and glycerophosphoric acid with a variety of fatty acids, one being a saturated fatty acid and one being an unsaturated fatty acid. *T. lanuginosus* phosphatidyl groups consist of laureate, oleate, palmitate, myristate, stearate and

linoleate, all of which are saturated fatty acids. Wright et al. (1983) reported that in *T. lanuginosus*, the concentration of linoleic acid (18:2) was two-fold higher at 30°C than at 50°C. The degree of unsaturation of phospholipid fatty acids was 0.88-fold in culture grown at 50°C but one fold in temperature-shifted cultures from 50 to 30°C and 1.06-fold at constant temperature of 30°C. These results suggest that the relatively high number of PEMT genes in conjunction with the ability of the *T. lanuginosus* to synthesis saturated fatty acids is used as a mechanism to stabilize membranes at high temperature by this fungus (Maheshwari *et al.*, 2000; Wright *et al.*, 1983).

Another mechanism that is used by microorganisms to resist adverse environmental conditions, is the synthesis and accumulation of organic compounds that may have protective functions. Trehalose has been widely reported as a part of the physiological adaptation to various environmental stresses in yeasts and filamentous fungi (Elbein *et al.*, 2003; Fillinger *et al.*, 2001a). Trehalose was initially considered as an energy source for cells, but later studies demonstrated that trehalose was also important in other metabolic processes, mainly, when organisms are exposed to stressful growth conditions. In *S. cerevisiae*, trehalose accumulation was observed when the cells were exposed to thermal shock and it is positively associated with the viability of cells under stress conditions (Fillinger *et al.*, 2001a). Some studies have sought to determine the mechanism of cellular protection during stress conditions by trehalose e.g., in stabilizing intracellular enzymes and preventing the aggregation of denatured proteins (Elbein et al., 2003). Trehalose has shown a stabilizing effect on cellular membranes during stress conditions (Patist and Zoerb, 2005). Metabolic reconstruction of the trehalose pathway of *T. lanuginosus* shows that trehalose can be synthesised by the cells and high assimilation and increased biomass was observed when the fungus was grown in trehalose in phenotypic microarray experiments which will be discussed in the following chapter (**Figure 28**). The utilization and growth on trehalose was comparable to glucose and xylose which this organism predominantly uses, compared to other monomeric sugars. Ferreira et al. (2007) demonstrated that when trehalose was added to cultures of the filamentous fungus *Pisolithus* sp. exposed to thermal stress (42°C), the culture was able to recover at optimal temperature (28°C). It was also shown that one of the enzyme, β-glucosidase, was stabilized or the activity even enhanced when trehalose was added (Ferreira *et al.*, 2007). This suggests that this thermophilic fungus may also

be utilizing trehalose as a mechanism for thermal adaptation by accumulating and synthesis trehalose for membrane and enzyme stabilization.

Aerobically growing cells are continuously challenged by potent oxidants produced during normal cellular metabolism. These oxidants, including hydrogen peroxide and organic peroxides, are important components mediating various cell functions. However, they can also cause cell damage when present at toxic levels. Aerobic organisms possess extensive antioxidant systems to regulate oxidant levels. *T. lanuginosus* has IscX methyltransferase, which is an active monomer that is part of the ISC system. It can bind Fe, and may function as a Fe-donor in the assembly of FeS clusters. Iron-sulphur (FeS) clusters are important cofactors for numerous proteins involved in electron transfer, in redox and non-redox catalysis, in gene regulation, and as sensors of oxygen and iron (Beinert *et al.*, 1997). *T. lanuginosus* also has mechanisms to protect the cell against oxidative damage. Five genes coding for a superoxide dismutases and five putative peroxiredoxin proteins were identified, which are enzymes that catalyze the dismutation of superoxide into oxygen and hydrogen peroxide (Alscher *et al.*, 2002; Wood *et al.*, 2003). Thus, they are an important antioxidant defence in nearly all cells exposed to oxygen (**Table 14**).

T. lanuginosus has a ubiquitin degradation system regulated by arginine-tRNA-protein transferase. The ubiquitin system also plays an important role in the degradation of denatured proteins following environmental stress such as nutrient limitation, heat shock, and heavy metal exposure (Staszczak, 2008). Meheshwari *et al.* (2000) observed that breakdown rate of the soluble proteins in thermophilic fungi was two-fold compared to that of mesophilic fungi and suggested that the increased turnover rate of soluble protein is important in the survival of thermophilic fungi at high temperatures. Since *T. lanuginosus* is a thermophilic and grows generally on dead woody material, the system maybe an essential adaptation during rising temperature in composting materials.

T. lanuginosus is also capable of histone acetylation/deacetylation and has a high number of methylases and can perform poly-ADP-ribosylation. Histone acetylation/deacetylation and methylation play in important role in packing and condensation of DNA which is another strategy for stabilizing DNA (Nowak and Corces, 2004). Poly-ADP-ribosylation is the addition of one or more ADP-ribose moieties to a protein (Belenky *et al.*, 2007). It plays an important role

cell signaling and the control of many cell processes, including DNA repair and apoptosis (Berger *et al.*, 2004 ; Corda and Di Girolamo, 2003). This DNA condensation and repairing mechanisms may make this fungus well adapted to living and thriving in high temperatures.

Another strategy used by microorganisms to survive at high temperature is by expressing a group of proteins, called heat shock proteins (Hsps). Heat shock proteins general function as an intracellular chaperone for other proteins. They play an essential role in protein interactions such as folding and assisting in the establishment of proper protein conformation (shape) and prevention of unwanted protein aggregation. The synthesis of Hsps is suggested to be an adaptive response to increased thermotolerance and survival in stressful conditions. The synthesis of Hsps is related to an adaptive response to increased thermotolerance and survival in stressful conditions. This was observed in conidia of *T. lanuginosus*, germinated at 50°C and heat shocked at 55°C for 60 min prior to exposure to 58°C. The conidia showed enhanced survival compared to non-heat-shocked conidia and that thermotolerance was decreased or eliminated if protein synthesis during the heat shock period was inhibited (Trent *et al.*, 1994). The genome of *T. lanuginosus* contains 19 genes coding for heat shock proteins. This was higher than other fungi used in the analysis. Phylogenetic analysis of these heat shock genes showed there are mainly contained in three clusters with all heat shock showing similarity to other fungal heat shock proteins (**Figure 29**). Most of the Hsps were similar to 70 and 30 type Hsps of other filamentous fungi. The Hsp30 gene of *A. nidulans* codes for an α-crystallin-related protein, similar to Hsp26 gene of *S. cerevisiae* which is induced by a number of stressed cultivation conditions which includes high temperatures. Also among the heat shock proteins identified in *T. lanuginosus* is the Hsp70 protein, which is involved in proper protein folding after synthesis. Hsp70 is an evolutionarily highly conserved molecular chaperone that promotes the survival of stressed cells by binding tightly to partially-synthesized peptide sequences (incomplete proteins). Hsp70 prevents them from aggregating and being rendered non-functional. Once the protein is completely synthesized Hsp70 will detach. The protein is then free to fold on its own, or to be transferred to other chaperones for further processing. The accumulation of this Hsp was observed in *A. nidulans* upon prolonged heating. (Freitas *et al.*, 2011). The high presences of these genes in *T. lanuginosus,* in comparison to mesophilic counter parts, provides clues for organisms maintaining stability in stressful conditions. Also, 16 genes were identified as DNAJ proteins. DNAJ responsed to hyperosmotic and heat shock by preventing the aggregation of

stress-denatured proteins which occurs at high temperatures (Diefenbach and Kindl, 2000). Other interesting temperature-dependent proteins that were identified are a group consists of several bacteria specific low temperature requirement A (LtrA) protein sequences which have been found to be essential for growth at low temperatures (Zheng and Kathariou, 1995). It also contains a number of uncharacterized fungal proteins. Although the function of these proteins is unknown in *T. lanuginosus*, they may be related to survival in low temperature until it increases in the composting waste.

All these mechanisms that exist in *T. lanuginosus* are also present in numerous mesophilic organisms. However in most mesophilic organisms one or two mechanisms may be found for use in unexpected change of environment or introduction of stress. However, in *T. lanuginosus,* all these mechanisms are available to be utilised if it is going to thrive in an otherwise stressful thermophilic environments. Coupled with a relatively smaller genome means less energy for maintenance, and a wide range of metabolism by CAZy proteins makes *T. lanuginosus* well-suited to thrive in composting environments.

CHAPTER 3: METABOLIC RECONSTRUCTION, PHENOTYPIC AND DNA MICROARRAY

This chapter was published in: Mchunu, Nokuthula Peace, Kugen Permaul, Maqsudul Alam, and Suren Singh. (2013). Carbon utilization profile of a thermophilic fungus, *Thermomyces lanuginosus* using phenotypic microarray." Advances in Bioscience and Biotechnology 4: 24.

Advances in Bioscience and Biotechnology, 2013, 4, 24-32 **ABB**
http://dx.doi.org/10.4236/abb.2013.49A004 Published Online September 2013 (http://www.scirp.org/journal/abb/)

Carbon utilization profile of a thermophilic fungus, *Thermomyces lanuginosus* using phenotypic microarray

Nokuthula Peace Mchunu[1,2*], Kugen Permaul[1], Maqsudul Alam[2,3], Suren Singh[1]

[1]Department of Biotechnology and Food Technology, Durban University of Technology, Durban, South Africa
[2]Centre for Chemical Biology, University Sains Malaysia, Bayan Lepas, Penang, Malaysia
[3]Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii, Honolulu, USA
Email: *nokuthula@dut.ac.za

Received 29 June 2013; revised 30 July 2013; accepted 26 August 2013

ABSTRACT

The thermophilic filamentous fungus, *Thermomyces lanuginosus* produces the largest amount of xylanase reported. In addition to this, it expresses large amount of other enzymes that have been used industrially or have academic interest. Thus, this fungus has a potential to be applied for biomass conversion to produce biofuel or other applications. In this study, the Biolog system was used to characterize the utilisation and growth of *T. lanuginosus* on 95 carbon sources. The carbohydrates based compounds, both single sugars and oligosaccharide, showed the best utilisation profile, with the pentose sugar xylose inducing the highest growth, followed by trehelose, raffinose, D-mannose turanose fructose and glucose. Among oligosaccharides, sucrose had the highest mycelium formation followed by stachyose, maltose, maltotriose, glycogen and dextrin. Interestingly the fungus also grew well on cellobiose suggesting that this fungus can produce cellulose hydrolysing proteins. D-alanine was the best amino acid to promote fungal growth while the effect of other amino acids tested was similar to the control. These results demonstrate the ability of this fungus to grow relatively well on most plant based compounds thus making this fungus a possible candidate for plant biomass conversion which can be applied to a number of biotechnological applications including biofuel production.

Keywords: Filamentous Fungi; Thermophilic; Carbon Source; Hexose; Pentose

1. INTRODUCTION

The importance of fungi and other microorganisms is widely acknowledged, primarily due to their application in biotechnology industries as well as the effects they have on human health. Fungi are able to produce a variety of biotechnology products which include industrial enzymes, enzymes used in bioassays or for diagnostics, antibiotics, and enzymes involved in bioremediation [1,2]. During industrial application and scientific research, specific metabolic pathways or molecules that are related to a particular process are studied in depth. This however can lead to the overlooking of other molecules or useful products. The invention of genomics has produced a wealth of data, however to understand those data one must understand the relationship of genes within an organism and the interactions of gene products in metabolism.

The area of studying either gene or protein interactions on a larger scale is a relatively new field as it has spilled over from genomics. Although high-throughput screens for bacteria and unicellular fungi (yeast) using knock-out experiments are used frequently, this technique is labour intensive and time consuming. Even after obtaining mutants, methods of characterization can be limited or expensive as in the case of DNA microarrays. Alternative approaches for the characterization of functional genes are being developed and advanced [3]. One approach is to focus on the effect of a particular gene at a cellular level and to assess how it affects the organism as a whole. Therefore, phenotypic characteristics that the organism displays, become markers (for the effect) of a particular gene with relatively high certainty. Although phenotyping has been around for some time, it still provides a very useful way to describe biological differences between cells. As such, a specific phenotype is the final goal of any strain enhancement process for new products or processes. Therefore a good phenotypic assay method would be beneficiary in functional genomics [4].

Like many organisms, the natural habitat of fungi in-

*Corresponding author.

A1 Water	A2 Tween 80	A3 N-Acetyl-D-Galactosamine	A4 N-Acetyl-D-Glucosamine	A5 N-Acetyl-D-Mannosamine	A6 Adonitol	A7 Amygdalin	A8 D-Arabinose	A9 L-Arabinose	A10 D-Arabitol	A11 Arbutin	A12 D-Cellobiose
B1 α-Cyclodextrin	B2 β-Cyclodextrin	B3 Dextrin	B4 i-Erythritol	B5 D-Fructose	B6 L-Fucose	B7 D-Galactose	B8 D-Galacturonic Acid	B9 Gentiobiose	B10 D-Gluconic Acid	B11 D-Glucosamine	B12 α-D-Glucose
C1 Glucose-1-Phosphate	C2 Glucuronamide	C3 D-Glucuronic Acid	C4 Glycerol	C5 Glycogen	C6 m-Inositol	C7 2-Keto-D-Gluconic Acid	C8 α-D-Lactose	C9 Lactulose	C10 Maltitol	C11 Maltose	C12 Maltotriose
D1 D-Mannitol	D2 D-Mannose	D3 D-Melezitose	D4 D-Melibiose	D5 α-Methyl-D-Galactoside	D6 β-Methyl-D-Galactoside	D7 α-Methyl-D-Glucoside	D8 β-Methyl-D-Glucoside	D9 Palatinose	D10 D-Psicose	D11 D-Raffinose	D12 L-Rhamnose
E1 D-Ribose	E2 Salicin	E3 Sedoheptulosan	E4 D-Sorbitol	E5 L-Sorbose	E6 Stachyose	E7 Sucrose	E8 D-Tagatose	E9 D-Trehalose	E10 Turanose	E11 Xylitol	E12 D-Xylose
F1 γ-Amino-butyric Acid	F2 Bromosuccinic Acid	F3 Fumaric Acid	F4 β-Hydroxy-butyric Acid	F5 γ-Hydroxy-butyric Acid	F6 p-Hydroxyphenyl-acetic Acid	F7 α-Keto-glutaric Acid	F8 D-Lactic Acid Methyl Ester	F9 L-Lactic Acid	F10 D-Malic Acid	F11 L-Malic Acid	F12 Quinic Acid
G1 D-Saccharic Acid	G2 Sebacic Acid	G3 Succinamic Acid	G4 Succinic Acid	G5 Succinic Acid Mono-Methyl Ester	G6 N-Acetly-L-Glutamic Acid	G7 Alaninamide	G8 L-Alanine	G9 L-Alanyl-Glycine	G10 L-Asparagine	G11 L-Aspartic Acid	G12 L-Glutamic Acid
H1 Glycyl-L-Glutamic Acid	H2 L-Ornithine	H3 L-Phenylalanine	H4 L-Proline	H5 L-Pyroglutamic Acid	H6 L-Serine	H7 L-Threonine	H8 2-Amino Ethanol	H9 Putrescine	H10 Adenosine	H11 Uridine	H12 Adenosine-5'-Monophosphate

Figure 1. 95 Carbon sources found in FF MicroPlate from Biolog, Inc.

(**Figure 2**) and for biomass (**Figure 3**). The analysis for general assimilation showed that cluster I and II contain carbon sources that lead to very slow biomass formation. The most dominant compounds in these clusters are amino acids, except for alanine, and organic acids, esters, alcohols, phosphorylated sugars, rare sugars, rare polymers, a nucleotide and aromatics groups. Water (control) was grouped in cluster II not I as it had higher assimilation rate. The trend was similar when growth was analyzed with exception that cluster I was bigger than cluster II (**Figure 3**). Amino acids and some carbohydrates are also identified to give slow formation of biomass in these clusters. The other difference was that water had moved down to cluster I while tween 80 shifted up to cluster II.

Cluster III (assimilation) showed good assimilation for *T. lanuginosus*. This cluster contained mainly carbohydrates which are monosaccharide (sorbose, galactose, arabinose, ribose fucose and rhaminose), disaccharides (Lactose and Lactoluse), oligosaccharides and polysaccharides (cyclodextrine, tagose, gentibiose amd melibiose), some amino acids (asparagine and alanyl-glycine) and alcohol (sorbitol, glycerol, Maltitol and xylitol). Cluster IV contained carbon sources that enabled the fastest growth and included several monosaccharides, oligosaccharides (xylose, glucose, raffinose, glucose, fructose, cellobiose, maltose arabitol, NA-glucosamine, etc.). Growth analysis revealed that compounds found in Cluster III and IV were similar to those found in assimilation analysis cluster IV, however the cluster proportions were different. In growth analysis, most of the carbon sources clustered in group III, while only three carbon sources found in cluster IV (xylose, NAG and sucrose) which were classified as yielding higher biomass. Surprisingly, cellobiose also showed good biomass production and was clustered in group III.

3.2. Hexoses and Pentoses

Analysis of these 95 carbon sources was done further by analyzing carbon sources that fell into the following specific groups, hexose and pentose, oligosaccharides and

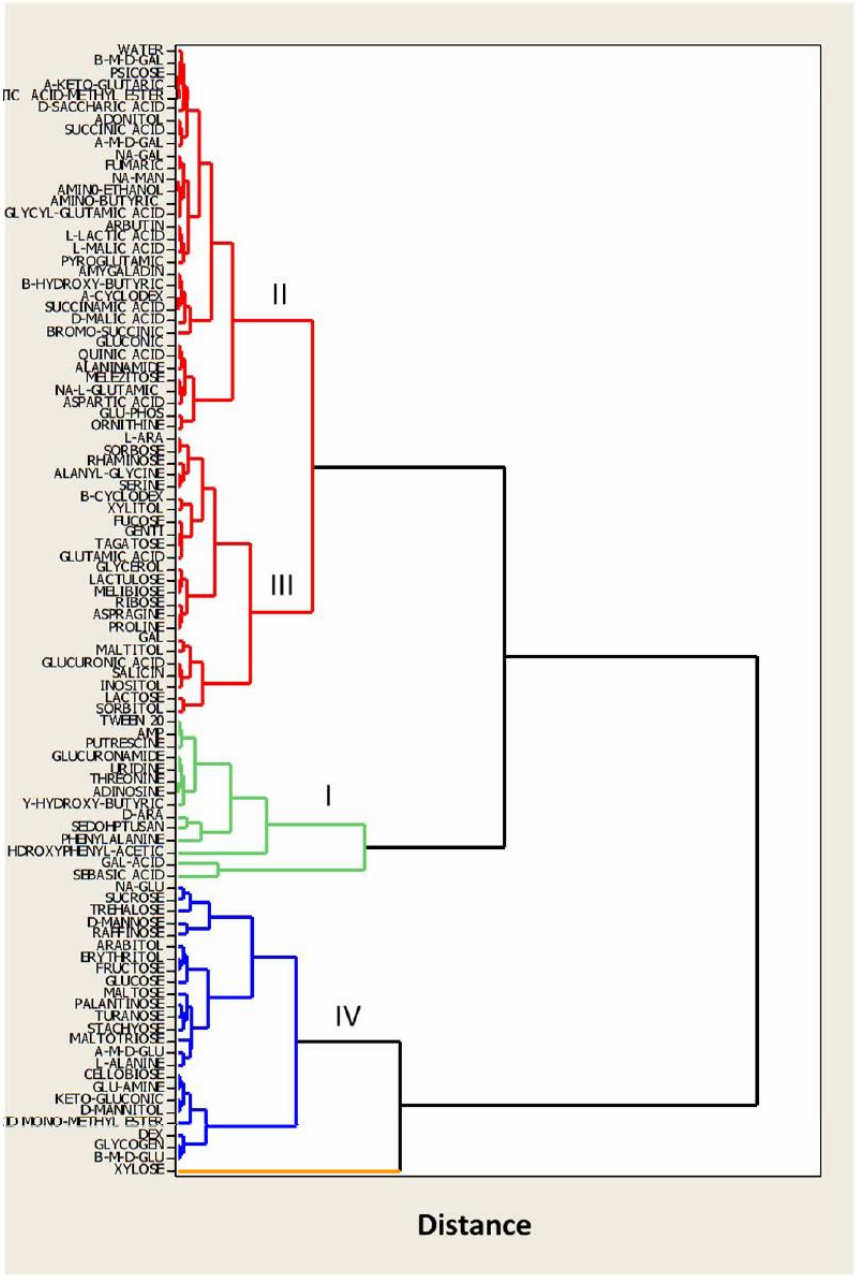

Figure 2. Joining cluster analysis applied to 95 carbon sources based on their assimilation and utilization of carbon sources by *T. lanuginosus* measured at 490 nm using the Biolog system (the standard deviation for absorbance values was an average of 0.041).

amino acid based compound and the rest were not assessed further. Analysis of hexose and pentose utilization revealed maximal assimilation of xylose followed by trehalose, NAG and mannose (**Figures 3** and **4**). Xylose exhibited 15% more assimilation than the second best compound trehalose with absorbance values of 3.1 and 2.6, repectively (**Figure 4**). Fructose, raffinose, glucose and turanose also showed good general assimilation. The

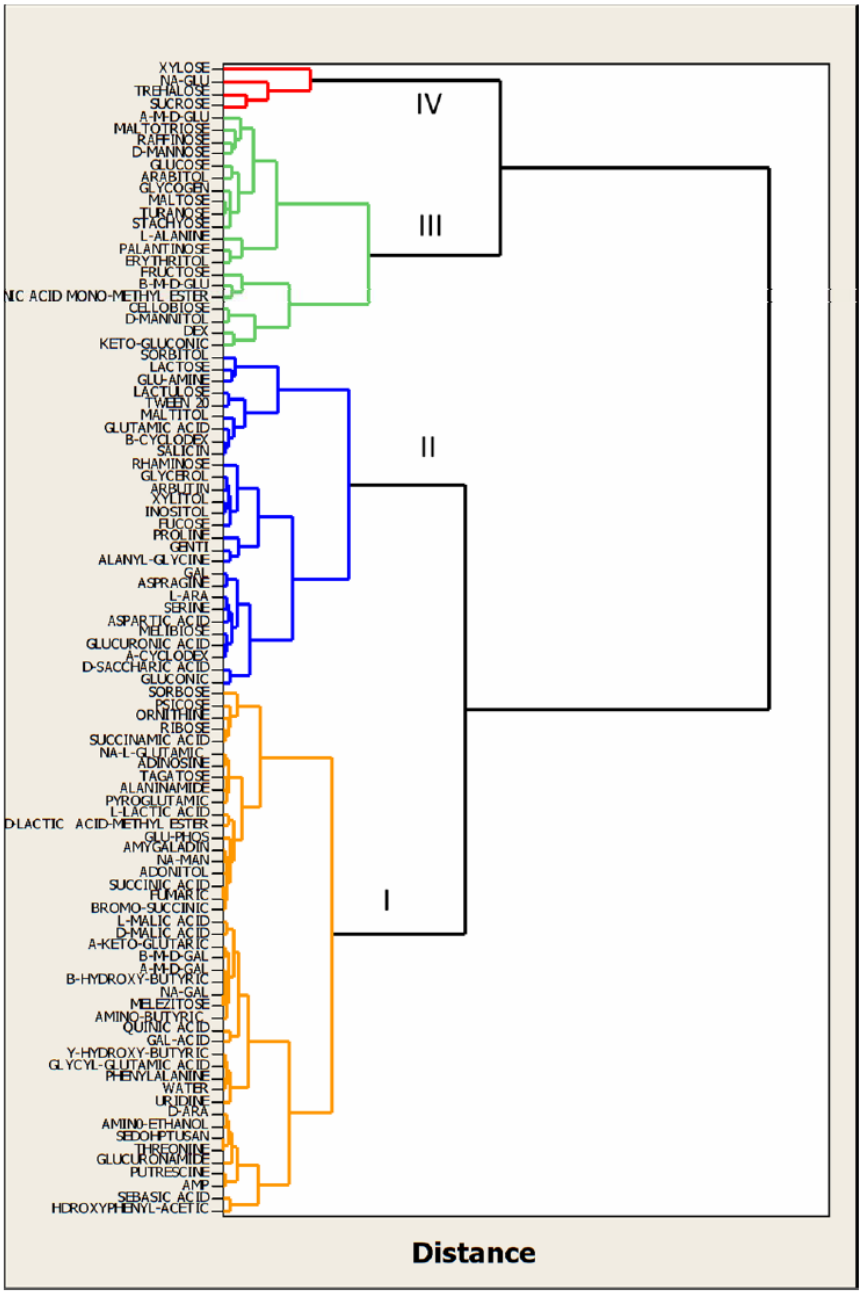

Figure 3. Joining cluster analysis applied to 95 carbon sources based on the growth of *T. lanuginosus* on these carbon sources. Groth was measured at 750 nm using the Biolog system.

biomass production showed that again xylose produced the highest biomass followed by NAG and trehalose (absorbance values, 1.4, 1.36 and 1.23, respectively, **Figure 5**). These were followed by glucose, mannose and raffinose with absorbance above 2, among the better hexose and pentose sugars. Water assimilation was measured at 1.36 and for tween 80 at 1.11. Nevertheless when the effect on growth was analyzed, tween 80 showed better biomass promotion than water with absorbance of 0.68 and 0.33, respectively.

3.3. Oligosaccharides

In oligosaccharide analysis, sucrose exhibited the best assimilation followed by maltotriose, stachyose, maltose, dextrin and glycogen (**Figure 6**). Cellobiose also showed

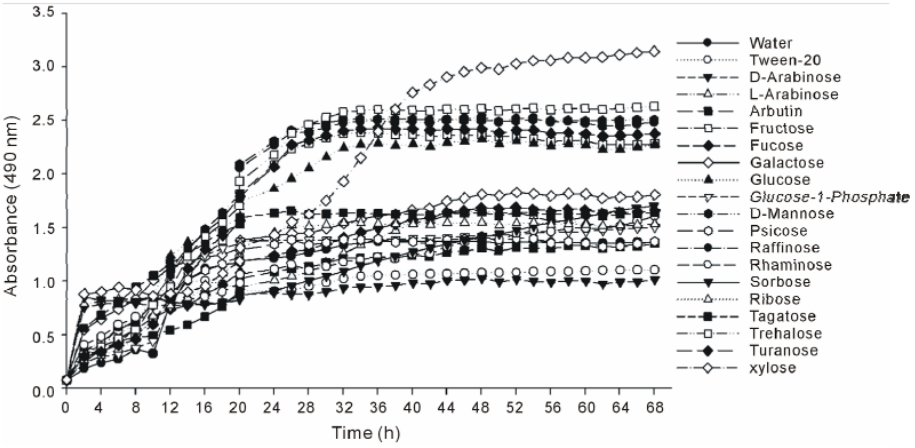

Figure 4. Assimilation of monomeric sugars (hexose and pentose) by *T. lanuginosus* SSBP. The assimilation was measured at an absorbance of 490 nm for 68 hours at 2 hour intervals.

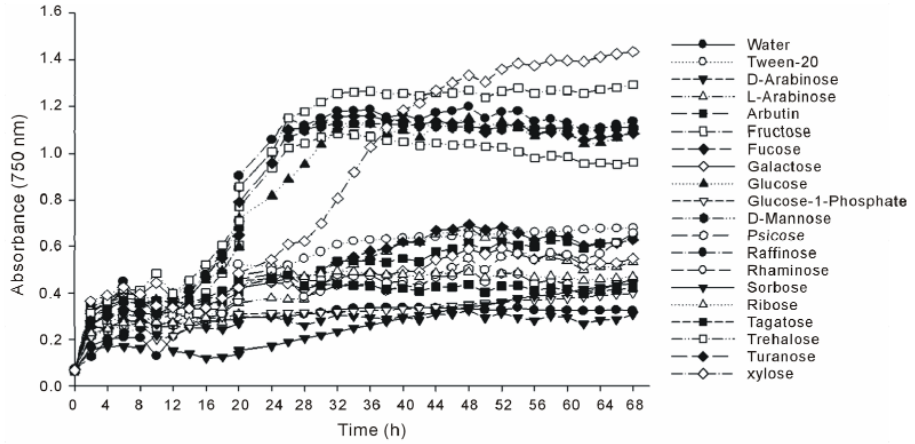

Figure 5. Growth of *T. lanuginosus* SSBP in monomeric sugars (hexose and pentose). The growth was measured at an absorbance of 750 nm for 68 hours at 2 hour intervals.

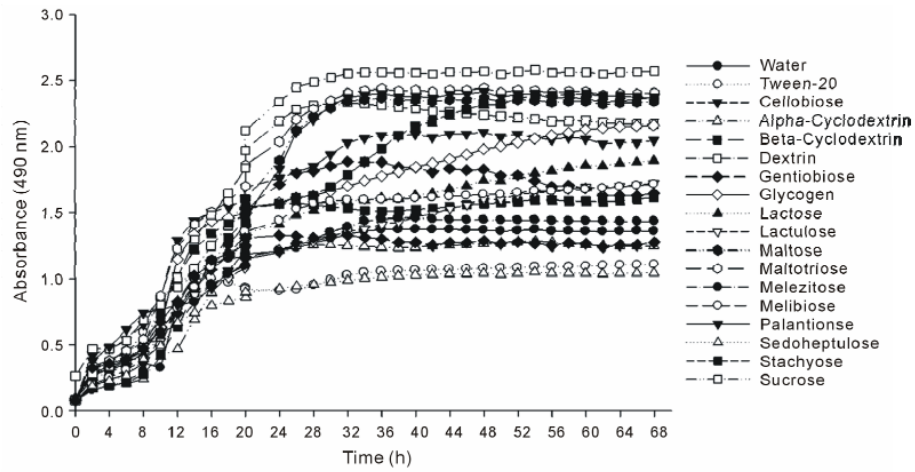

Figure 6. Assimilation of oligosaccharides by *T. lanuginosus* SSBP. The assimilation was measured at an absorbance of 490 nm for 68 hours at 2 hour intervals.

relatively good general assimilation. Water assimilation was lower than most of common carbohydrates while the assimilation of rare occurring carbohydrate compounds was even lower than water and tween 80 (sedoheptulose and gentibiose). In biomass production sucrose again produced the most biomass followed by maltose, glycogen, maltose, stachyose, palantiose, cellobiose and dextrin (**Figure 7**). Again common carbohydrate compounds supported more biomass production in *T. lanuginosus* than rare compounds.

3.4. Amino Acids

Amino acid analysis, L-alanine displayed the best assimilation followed by proline, asparagine, and glutamic acid (**Figure 8**). Gylcyl-glutamic acid gave the lowest assimilation even lower than water and tween 80. It was also noted that although most of the amino acid base compounds had high assimilation, they were unable to support significant biomass production. In biomass production L-alanine yielded greater biomass when compared to other amino acids (**Figure 9**). The rest of the amino acid compounds produced less biomass than tween 80 but more than water except for threonine which was lower.

4. DISCUSSION

In nature the ability of a microorganism to use a variety of compounds is vital for survival in composting environment as different substrates are degraded and utilised by different organisms. Filamentous fungi play a vital role in this ecological dynamics as they are responsible for the majority of the hydrolysis [8,9]. *T. lanugiosus* is

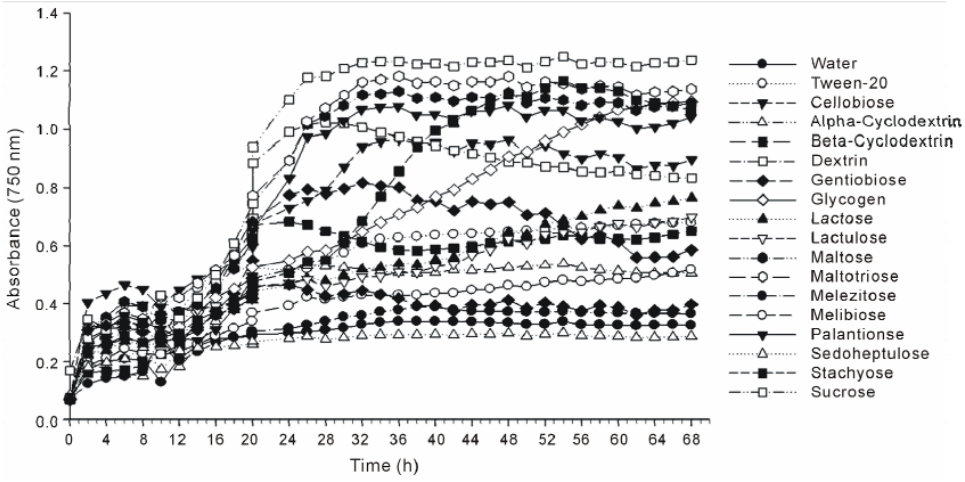

Figure 7. Growth of *T. lanuginosus* SSBP in oligosaccharide compounds. The growth was measured at an absorbance of 750 nm for 68 hours at 2 hour intervals.

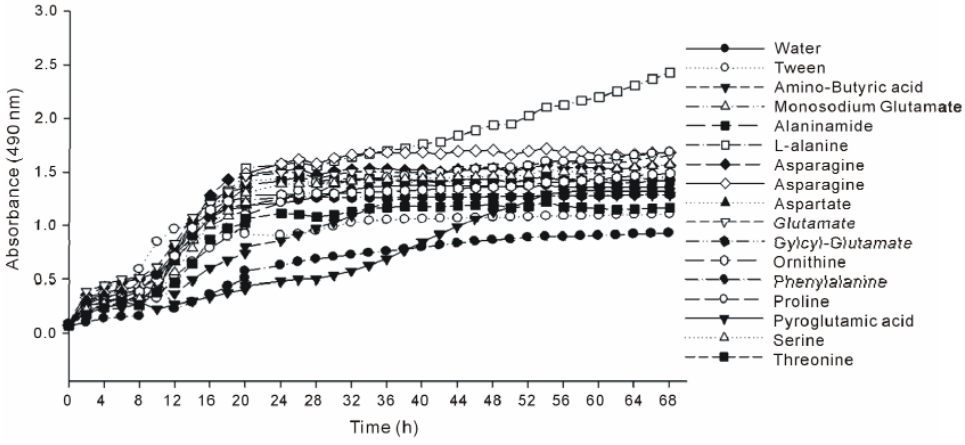

Figure 8. Assimilation of amino acid based compounds by *T. lanuginosus* SSBP. The assimilation was measured at an absorbance of 490 nm for 68 hours at 2 hour intervals.

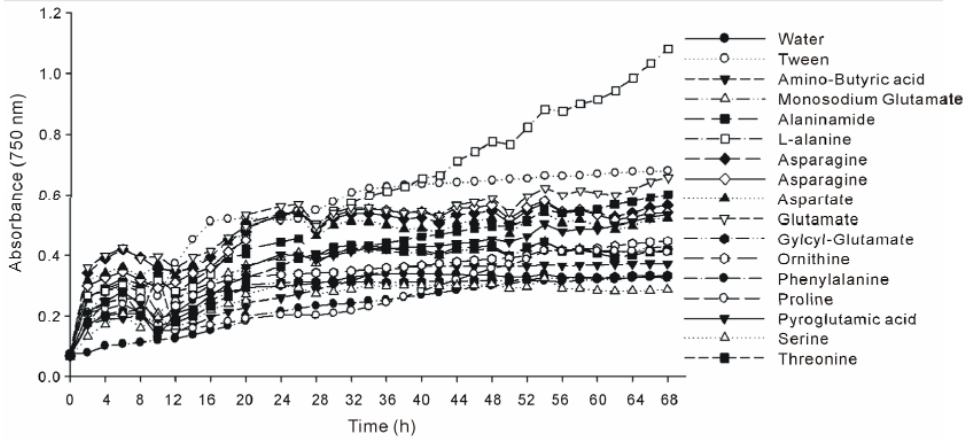

Figure 9. Growth of *T. lanuginosus* SSBP in amino acid based compounds. The growth was measured at an absorbance of 750 nm for 68 hours at 2 hour intervals.

among those fungal organisms that thrive in such environments with an added ability to survive high temperature which is only for a select few eukaryotic organisms [10]. The analysis of carbon source assimilation and utilization for biomass production in this organism revealed a similar profile to other filamentous fungi studies of this nature where glucose, xylose, trehalose and NAG produced high biomass in *Trichoderma reesei* and *Aspergillus niger* [5]. Although the clusters in these studies were similar to our findings, closer analysis of Cluster IV revealed that for *T. lanuginosus*, xylose is the preferred sugar compared to glucose. This concurs with reports that *T. lanuginosus* has the most powerful system for xylanase production and xylose utilization and thus it was expected that xylose would produce the most biomass and have the highest assimilation [11-13]

However, the most interesting finding was the high cellobiose utilization as this organism is well reported as a cellulose free organism. *T. lanuginosus* has been previously described as non-cellulolytic and it was suggested that it probably relies on commensal relationships in composts with cellulolytic fungi [13-15]. In this study, growth on cellobiose suggests that this fungus produces enzymes that have cellulose related activity. This is in agreement with unpublished data on genome sequencing of this fungus revealing that 8 predicted genes are with the possibility of having cellulose activity. Of the 8 genes, 3 were similar to *Trichoderma reesei* cellulases and the others to *Aspergillus kawachi* [16,17].

Trehalose also produced good biomass and assimilation in *T. lanuginosus*. The suggested reason for this is that trehalose is used by the organisms as an energy source; however there is a more important reason in thermophilic organisms. Trehalose has been widely reported as a part of the physiological adaptation to various environmental stresses e.g. high temperature, in yeasts and filamentous fungi [18]. NAG also had high assimilation and biomass production because it is the building block of fungal cell walls which contain chitin and also can be converted to energy molecule, therefore high assimilation and the ability to support growth were expected [19]. It was surprising that only one amino acid, alanine, produced significant biomass. This may be because alanine is one of the few amino acids that can transform into glucose and can be used in TCA cycle to provide energy for the cell, thus it may be preferred by this fungus to supplement the supply of mineral nitrogen and energy [20].

In conclusion, this study indicates that *T. lanuginosus* is a versatile organism that can utilize a diverse range of carbon sources, including carbohydrates, amino acids, carboxylic acids, polymers, aromatics, esters, phosphorylated and sugar alcohols. The application of Phenotypic Array as a tool of carbon utilization studies is a quick approach to studying and assessing filamentous fungi for specific activities.

5. ACKNOWLEDGEMENTS

This study was supported by grants from the National Research Foundation, Republic of South Africa and collaboration with the Centre for Chemical Biology, University Sains Malaysia. The authors are thankful to Dr. Alison Winger, a Research Associate in Dept. Biotechnology and Food Technology Durban University of Technology, for valuable suggestions and critical evaluation of the manuscript.

REFERENCES

[1] Ghorai, S., Banik, P., Verma, D., Chowdhury, S., Mukherjee, S. and Khowala, S. (2009) Fungal biotechnology in food and feed processing. *Food Research International*, **42**, 577-587. doi:10.1016/j.foodres.2009.02.019

[2] Singh, S., Madlala, A.M. and Prior, B.A. (2003) *Ther-*

momyces lanuginosus: Properties of strains and their hemicellulases. *FEMS Microbiology Reviews*, **27**, 3-16.

[3] Nitsche, B.M., Crabtree, J., Cerqueira, G.C., Meyer, V., Ram, A.F.J. and Wortman, J.R. (2011) New resources for functional analysis of omics data for the genus *aspergillus*. *BMC Genomics*, **12**, 486.

[4] Murat, C., Riccioni, C., Belfiori, B., Cichocki, N., Labba, J., Morin, E., Tisserant, E., Paolocci, F., Rubini, A. and Martin, F. (2012) Distribution and localization of microsatellites in the perigord black truffle genome and identification of new molecular markers. *Fungal Genetics and Biology*, **48**, 592-601. doi:10.1016/j.fgb.2010.10.007

[5] Druzhinina, I., S. , Schmoll, M., Seiboth, B. and Kubicek, C., P. (2005) Global carbon utilization profiles of wild-type, mutant, transformant strains of *Hypocrea jecorina*. *Applied and Environmantal Microbiology*, **72**, 2126-2133.

[6] Singh, M.P. (2009) Application of biolog ff microplate for substrate utilization and metabolite profiling of closely related fungi. *Journal of Microbiological Methods*, 77, 102-108. doi:10.1016/j.mimet.2009.01.014

[7] Maheshwari, R., Bharadwaj, G. and Mahakingesgwara, K.B. (2000) *Thermophilic fungi*: Their physiology and enzymes. *Microbiology and Molecular Biology Reviews*, **64**, 461-488. doi:10.1128/MMBR.64.3.461-488.2000

[8] Tang, J.-C. and Katayama, A. (2005) Relating quinone profile to the aerobic biodegradation in thermophilic composting processes of cattle manure with various bulking agents. *World Journal of Microbiology and Biotechnology*, **21**, 1249-1254.

[9] Yu, H., Zeng, G., Huang, H., Xi, X., Wang, R., Huang, D., Huang, G. and Li, J. (2007) Microbial community succession and lignocellulose degradation during agricultural waste composting. *Biodegradation*, **18**, 793-802.

[10] Puchart, V., Katapodis, P. and Biely, P. (1999) Production of xylanases, mannanases and pectinases by the themophilic *Fungus Thermomyces lanuginosus*. *Enzyme Microbial Technology*, **24**, 355-361.

[11] Dodd, D., Mackie, R., I and Cann, I., K.O. (2011) Xylan degradation, a metabolic property shared by rumen and human colonic bacteroidetes. *Molecular Microbiology*, **79**, 292-304. doi:10.1111/j.1365-2958.2010.07473.x

[12] Kuttanpillai, S.K., Ayyachamy, M., Kugen, P. and Suren, S. (2009) Production of β-xylanase by a *Thermomyces lanuginosus* mc 134 mutant on corn cobs and its application in biobleaching of bagasse pulp. *Journal of Bioscience and Bioengineering*, **107**, 494-498. doi:10.1016/j.jbiosc.2008.12.020

[13] Singh, S., Reddy, P., Haarhoff, J., Biely, P., Janse, B., Pillay, B., Pillay, D. and Prior, B.A. (2000) Relatedness of *Thermomyces lanuginosus* strains producing a thermostable xylanase. *Journal of Biotechnology*, **81**, 119-128.

[14] Manimaran, A., Kuttan-Pillai, S.K., Permaul, K. and Singh, S. (2009) Hyper production of cellulase-free xylanase by *Thermomyces lanuginosus* ssbp on bagasse pulp and its application in biobleaching. *Applied Microbiology and Biotechnology*, **81**, 887-893.

[15] Purkarthofer, H., Sinner, M. and Steiner, W. (1993) Cellulase-free xylanase from *Thermomyces lanuginosus*: Optimization of production in submerged and solid-state culture. *Enzyme and Microbial Technology*, **15**, 677-682.

[16] Martinez, D., Cullen, D., Danchin, E.G.J., Grigoriev, I.V., Harris, P., Jackson, M., Kubicek, C.P., Han, C.S., Ho, I., Larrondo, L.F., de Leon, A.L., Berka, R.M., Magnuson, J.K., Merino, S., Misra, M., Nelson, B., Putnam, N., Robbertse, B., Salamov, A.A., Schmoll, M., Terry, A., Thayer, N., Henrissat, B., Westerholm-Parvinen, A., Schoch, C.L., Yao, J., Barabote, R., Barbote, R., Nelson, M.A., Detter, C., Bruce, D., Kuske, C.R., Xie, G., Saloheimo, M., Richardson, P., Rokhsar, D.S., Lucas, S.M., Rubin, E.M., Dunn-Coleman, N., Ward, M., Brettin, T.S., Arvas, M., Baker, S.E., Chapman, J., Chertkov, O. and Coutinho, P.M. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnology*, **26**, 553-560. doi:10.1038/nbt1403

[17] Zhao, Z., Liu, H., Wang, C. and Xu, J.-R. (2013) Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*, **14**, 274. doi:10.1186/1471-2164-14-274

[18] Ferreira, A.S., Tótola, M.R. and Borges, A.C. (2007) Physiological implications of trehalose in the ectomycorrhizal fungus pisolithus sp. Under thermal stress. *Journal of Thermal Biology*, **32**, 34-41. doi:10.1016/j.jtherbio.2006.08.009

[19] Hartl, L., Zach, S. and Seidl-Seiboth, V. (2012) Fungal chitinases: Diversity, mechanistic properties and biotechnological potential. *Applied Microbiology and Biotechnology*, **93**, 533-543. doi:10.1007/s00253-011-3723-3

[20] Bechem, E.E.T. (2012) Utilisation of organic and inorganic nitrogen sources by *Scleroderma sinnamariense* mont. Isolated from *Gnetum africanum* welw. *African Journal of Biotechnology*, **11**, 9205-9213.

3.1 INTRODUCTION

The importance of fungi and industrial applicability has been already been discussed in detail in the previous chapters of this thesis. Of essence is that, specific metabolic pathways or molecules that are related to t particularindrustrial processes are usually studied in depth. This however, inevitably leads to other possible existing molecules or useful products being ignored. In addition, genomic research has produced a wealth of data on genes encoding enzymes, but to understand the data one must understand the relationship of genes within an organism and the interactions of these gene products in metabolic pathways (Kern *et al.*, 2007).

A metabolic pathway is a series of chemical reactions occurring within a cell. In each pathway, a molecule/chemical is modified through a series of other chemical reactions. Often enzymes catalyze these reactions with an addition of other molecules like minerals, vitamins, and other cofactors, in order to function efficiently and properly. Because of the many metabolites that may be involved, metabolic pathways can be quite complicated. In addition, numerous distinct pathways co-exist within a cell, forming metabolic networks which are important in the maintenance of homeostasis within an organism. Metabolic pathways can be divided into catabolic (degradative) and anabolic (synthesis) pathways, which often work interdependently to create new biomolecules as the final end-product (Lavoie *et al.*, 2009). Some pathways are anaplerotic where intermediates are formed for pathways such as the tricarboxylic acid cycle.

The study of metabolism has changed drastically during the last century. The concept of metabolic pathways was influenced by the limits of experimental methods available previously, resulting in a stepwise elucidation of metabolic pathways. In the last three decades, the discovery of the structure and information coding of DNA laid the foundations for recombinant technology, making microorganisms more amenable to metabolic engineering. Significant insights can also be obtained from the annotated genome sequence of an organism by using metabolic reconstruction. To fully exploit genome sequencing data, it requires that annotated genes are organised into full metabolic networks in order to understand if an organism has the ability to utilise specific metabolic pathways. The process of metabolic reconstruction begins with the annotation of the relevant genes. The metabolic reactions that the associated gene products

catalyze are delineated by incorporating data on the metabolites from databases (BRENDA, ExPASy and KEGG) and the literature using suitable software. The Pathway Tools software package (Ariadne Genomics) is used to generate a complete pathway including an inferred metabolic network for an organism using just an annotated genome and reference databases like those already mentioned (Ariadne Genomics). Metabolic network inference is a multi-step process that includes the matching of enzyme functions in the annotated genome to known reactions in metabolic reference databases. Pathway Studio software (Ariadne Genomics) used in this study, offers an exhaustive resource of molecular data along with inference tools that work in a graphical, biological context. Unlike manually-curated databases, Pathway Studio provides a direct access to underlying evidence so that it can be independently validated and determined (Bonnet *et al.*, 2009; Nikitin *et al.*, 2003). This is because the software includes Medscan, which scans the literature for genes or molecules related to the gene of interest. The construction of these pathways from genomic data is vital as it provides further evidence of the integrity of the assembly and the accuracy of gene annotation. If genome assembly and annotation is flawed, it would be difficult to build complete metabolic pathways including the most basic.

Studying single gene or protein interactions has been investigated for many decades, however new methodologies are required as the amount of data generated from genomics is enormous. Thus high-throughput equipment and methodologies have been developed to study the phenotypic effect of single gene changes on a cell. High-throughput screening of bacteria and unicellular fungi (yeast) using knock-out experiments is used frequently. However, in filamentous fungi this technique is labour intensive and time consuming (Cohen *et al.*, 2001). Even after successfully obtaining mutants, further research can be limited or expensive. Techniques such as DNA microarray focus on specific genes or processes by studying the effect of altered genes at genetic or protein levels using electrophoresis and proteomics. New approaches to characterise gene function are being developed e.g., proteomics, metabolomics, secretome studies and phenotypic microarrays. Phenotypic microarrays focus on effects of a particular gene at cellular level, where it is easier to study. Phenotypic characteristics are markers for the effect of a particular gene and its particular function, with relatively high certainty. Phenotypes provide a very useful way to describe biological differences between cells. They reflect both genetic and epigenetic differences among cells. A phenotype is the manifestation of

a specific characteristic as the product of genes and habitat. As such, a specific phenotype is the final goal of any strain enhancement process for new products or processes. Therefore, a good phenotypic assay method is beneficial in functional genomics (Bochner *et al.*, 2001).

Like with many organisms, the natural habitat of fungi determines what phenotypes it will exhibit. The natural environments of fungi contain many factors, which includes nutrients that they metabolise, where nutrients are the major controllers of phenotypic characteristics, they are vital in comparative assays. Idealy, these assays should comprise a combination of hundreds of carbon, nitrogen, phosphate sulphur and other nutrients. This will push the boundaries with assay numbers of hundreds of thousands when including other physical factors such as temperature, pH and O_2. Such scales are a not feasible for most laboratories due to labour and cost restrictions (Clemons, 2004).

Thus, the introduction of Phenotypic MicroArray thechnology (PM) from Biolog Incorporated offers a viable option for most researchers and industries. The Biolog system is designed for high-throughput screening of different basic nutrient sources and additives required for growth and also can screen antagonistic compounds towards numerous microorganisms including filamentous fungi. This method has proven successful using bacteria strains and their respective mutants in phenotypic profiling (Bochner *et al.*, 2001; Borglin *et al.*, 2012; Ito *et al.*, 2005; Lea and Irina, 2010). Phenotypic microarray technology was developed to provide scientists with a simple and efficient way to test nearly 2000 phenotypes of a microbial cell and gain a comprehensive overview of pathway functions in a single experiment. The phenotypic assays are designed from a physiological perspective to survey, in vivo, the function of diverse pathways including both metabolic and regulatory pathways. Included in the tests are basic cellular nutritional pathways for C, N, P, and S metabolism, pH growth range and regulation of pH control, sensitivity to NaCl and various other ions, and sensitivity to chemical agents that disrupt various biological pathways. There are numerous scientific publications (>80) that have used PM technology in basic cellular research, pathogenesis and epidemiology, gene function and genome annotation, drug discovery, bioprocess development, and cell identification and taxonomy which highlights the applicability of this technique (Bochner *et al.*, 2001; Druzhinina *et al.*, 2005; Ito *et al.*, 2005; Lea and Irina, 2010; Mohammada *et al.*, 2012; von Eiff *et al.*, 2006) . The

filamentous fungi (FF) database of known fungi stores data of fungal growth via turbidimetric analysis (Biolog, Inc.). Analysis of both colour development and turbidity allows for accurate identifications to the species level. Thus, the use of phenotypic microarray, a technology that allows over 1000 cellular traits to be measured simultaneously as well as accurately, allows for more phenotypic traits to be identified than conventional testing (Borglin *et al.*, 2012; Lea and Irina, 2010).

One of the phenotypic characteristics that is desired by numerous industries is the ability of an organism to utilize any plant biomass. Plant biomass conversion is one of the key areas for improving the feasibility of biofuel production. However one of the obstacles in doing this has been finding a suitable organism that converts different carbohydrate compounds and has the requisite biological and physiological characteristics compatible with the process. One type of screening system from the PM technology is an FF MicroPlate specifically designed for testing of carbon utilization in filamentous fungi and yeast, including species from the genera *Aspergillus*, *Penicillium*, *Fusarium*, *Alternaria*, *Mucor*, *Gliocladium*, *Cladosporium*, *Paecilomyces*, *Stachybotrys*, *Trichoderma*, *Zygosaccharomyces*, *Acremonium*, *Beauveria*, *Botryosphaeria*, *Botrytis*, *Candida*, and *Geotrichum*. The assay is based on reduction of tetrazolium dye during metabolism (Biolog, Inc.). This technology according to Biolog is compatible with any carbon source with greatly simplifies the testing process, as no color developing chemicals need to be added after inoculation. The current system contains 95 different carbon sources that can be tested simultaneous for multiple activities on different carbon material. This system has been used to study the utilization of different carbon sources by different strains and mutants of *T. reesei* and *A. niger* (Druzhinina *et al.*, 2005; Mohammada *et al.*, 2012).

T lanuginosus produces a wide range of thermostable enzymes including a large group of carbohydrate hydrolases. These enzymes include amylase, glucoamylase, xylanase, lipase, phytase, protease and chitinase. This wide variety of thermostable enzymes can be applied in different industries. Amylases and glucoamylases can be used in the food industry for production of sugar syrup, in the baking industry and in animal feeds. Xylanases have immense potential in the pulp and paper industry as they can be used for bleaching of pulp, instead of chlorine which

produces toxic waste. This enzyme can also be used in the pre-digestion of animal feed and in the food industry. Phytase is used in the pre-digestion of animal feed in order to release phosphate, making it bio-available (for animal usage). Chitinases have potential for application in bioremediation of seafood waste by degrading the chitin component which makes up crustacean shells. Based on this, the *T. lanuginosus* genome may contain previously unidentified proteins that are active on the different carbohydrate materials analysed using PM technology, however, the elucidation of genes involved is essential.

DNA microarray analysis is one of the fastest-growing new technologies in the field of genetic research and is used to analyse gene expression is organisms. Gene expression is one of the most accurate methods of understanding how organisms adapt to changes in the environment. Thousands of genes are being discovered for the first time by genome sequencing (Dorrell *et al.*, 2002; Hedegaard *et al.*, 2009; Heller, 2002). DNA microarrays provide a natural vehicle for this exploration as many genes can be studied in a single reaction using high-throughput screening. Exploration of the genome using DNA microarrays and other technologies could increase our understanding of gene function and interactions between different genes.

A typical microarray experiment involves the hybridization of an mRNA molecule to the DNA probe attached to a slide or array. The level of expression is determined by the amount of mRNA bound to each site on the array measured by using a fluorescent dye. All the data is collected and a profile is generated for gene expression in the cell or organism. Microarrays can be categories into three (i) expression analysis, (ii) mutation analysis and, (iii) comparative genomic hybridization (Dorrell *et al.*, 2002; Kim *et al.*, 2008a). DNA microarrays has found many applications including, (i) gene discovery, (ii) disease diagnosis, (iii) drug discovery and, (iv) toxicological research (Heller, 2002) .

This chapter will focus on the use phenotypic and DNA microarray using the FF MicroPlate to identify the ability of *T. lanuginosus* to utilize different carbon sources. This also was used to confirm known activities that were reported in literature. Furthermore, metabolic reconstruction of selected pathways will provide further evidence of accurate gene annotation and pathways that are functional in *T. lanuginosus*. DNA microarrays was used for expression analysis of *T.*

lanuginosus grown in glucose and xylose in order to understand and identify genes involved in the metabolism of the sugars (pentose and hexose) as this organism is able to utilise the efficiently.

3.2 Materials and Methods

3.2.1 Pathway Reconstruction

The fungal database from Ariadne Genomics contains a collection of 125 metabolic pathways imported from KEGG mainly those from *A. Niger*, *A. flavus*, *T. reesei*, *S. cerevisiae* and *N. crassa*. These pathways are represented as a collection of functional classes (enzymes) and a set of corresponding chemical reactions. Every functional class in the database can contain an unlimited number of protein members encoding corresponding enzymatic activity. Usually a set of members includes paralogs of catalytic and regulatory subunits necessary to perform enzymatic activity. Manual population of functional classes by protein members represents the initial reconstruction of metabolic pathways in Pathway Studio. The process is equivalent of closing gaps in a metabolic network. TBLASTN was used against assembled DNA sequences of the *T. lanuginosus* genome to manually identify proteins overlooked by automatic annotation by orthologs identified with best reciprocal hit method from BLASTP results. The typical workflow for closing gaps in the *T. lanuginosus* metabolic network involved downloading protein sequences that could perform the missing enzymatic activity from other fungi or bacterial genomes and then using it as query for TBLASTN.

3.2.2 Phenotypic MicroArray

Biology Phenotypic MicroArray was used to compare gene model validation of some of the CAZY family proteins which are proteins involved in utilization of carbohydrate compounds. The experiments were performed by growing *T. lanuginosus* on 2% malt extract agar at 50°C for 5-7 days until spore formation was visible. Global carbon assimilation profiles were evaluated by using Biolog FF MicroPlate. The FF MicroPlate test panel contains 95 wells, each with a different lyophilised carbon-containing compound, and one well with water as control (**Figure 30**). The inoculum for the 96 well FF plates for the biolog system collection the spores from plate and resuspending then in 16 ml of FF inoculum media supplied by Biolog in glass tubes. The spore suspension displayed 75% transmittance (recommended by the manufacturer) at 590 nm using the Biolog tubidometer. Hundred microlitres of the spore suspension was dispensed

into each well. Microplates were incubated at 50°C and readings were taken using the Biolog Microstation, at 2 h intervals until 68 h had elapsed.

Readings were taken at absorbance at 490 nm and 750 nm, which were used to estimate mycelial growth based on turbidity and the change in optical density. The growth was measured at the absorbance of 750 nm, while assimilation (general uptake) of each carbohydrate was evaluated at 490 nm. Mitochondrial activity which plays an important rule in evaluation of metabolic reaction, resulted in the formation of a reddish-orange colour and was measured at absorbance 490 nm. Joining cluster analysis was used to group carbon sources utilized by *T. lanuginosus* using Minitab 16 software (Minitab Inc.) and was applied to identify the different groups of carbon sources from the experimental data set. The joining cluster analysis was designed by means of the Euclidean distance with complete linkage. Out of the 95 compounds used in this analysis, only compounds belonging to carbohydrates and amino acid group will be discussed in this chapter.

A1 Water	A2 Tween 80	A3 N-Acetyl-D-Galactosamine	A4 N-Acetyl-D-Glucosamine	A5 N-Acetyl-D-Mannosamine	A6 Adonitol	A7 Amygdalin	A8 D-Arabinose	A9 L-Arabinose	A10 D-Arabitol	A11 Arbutin	A12 D-Cellobiose
B1 α-Cyclodextrin	B2 β-Cyclodextrin	B3 Dextrin	B4 i-Erythritol	B5 D-Fructose	B6 L-Fucose	B7 D-Galactose	B8 D-Galacturonic Acid	B9 Gentiobiose	B10 D-Gluconic Acid	B11 D-Glucosamine	B12 α-D-Glucose
C1 Glucose-1-Phosphate	C2 Glucuronamide	C3 D-Glucuronic Acid	C4 Glycerol	C5 Glycogen	C6 m-Inositol	C7 2-Keto-D-Gluconic Acid	C8 α-D-Lactose	C9 Lactulose	C10 Maltitol	C11 Maltose	C12 Maltotriose
D1 D-Mannitol	D2 D-Mannose	D3 D-Melezitose	D4 D-Melibiose	D5 α-Methyl-D-Galactoside	D6 β-Methyl-D-Galactoside	D7 α-Methyl-D-Glucoside	D8 β-Methyl-D-Glucoside	D9 Palatinose	D10 D-Psicose	D11 D-Raffinose	D12 L-Rhamnose
E1 D-Ribose	E2 Salicin	E3 Sedoheptulosan	E4 D-Sorbitol	E5 L-Sorbose	E6 Stachyose	E7 Sucrose	E8 D-Tagatose	E9 D-Trehalose	E10 Turanose	E11 Xylitol	E12 D-Xylose
F1 γ-Amino-butyric Acid	F2 Bromosuccinic Acid	F3 Fumaric Acid	F4 β-Hydroxy-butyric Acid	F5 γ-Hydroxy-butyric Acid	F6 p-Hydroxyphenyl-acetic Acid	F7 α-Keto-glutaric Acid	F8 D-Lactic Acid Methyl Ester	F9 L-Lactic Acid	F10 D-Malic Acid	F11 L-Malic Acid	F12 Quinic Acid
G1 D-Saccharic Acid	G2 Sebacic Acid	G3 Succinamic Acid	G4 Succinic Acid	G5 Succinic Acid Mono-Methyl Ester	G6 N-Acetly-L-Glutamic Acid	G7 Alaninamide	G8 L-Alanine	G9 L-Alanyl-Glycine	G10 L-Asparagine	G11 L-Aspartic Acid	G12 L-Glutamic Acid
H1 Glycyl-L-Glutamic Acid	H2 L-Ornithine	H3 L-Phenylalanine	H4 L-Proline	H5 L-Pyroglutamic Acid	H6 L-Serine	H7 L-Threonine	H8 2-Amino Ethanol	H9 Putrescine	H10 Adenosine	H11 Uridine	H12 Adenosine-5'-Monophosphate

Figure 30: Carbon sources (95) found in the filamentous fungal (FF) MicroPlate from Biolog, Inc.

3.2.3 DNA MicroArray

Total RNA was isolated as described in section 2.2 in chapter 2 after *T. lanuginosus* was grown in media containing containing 0.1% KH_2PO_4, 2% glucoseor xylose, 3% yeast extract at pH 6.5, for 3 days at 50°C. This was followed by cDNA synthesis using the Invitrogen SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen) to synthesize double-stranded cDNA following the manufacturer's instructions. RNA removal was achieved by the addition of 4 mg/ml RNase A solution to the cDNA tubes and incubated the samples at 37°C for 10 minutes. The samples were centrifuged at 12 000 *g* for 2 minutes. The cDNA was extracted by adding of phenol-chloroform-isoamyl alcohol (25:24:1) to samples and centrifuging at 12 000 g for 5 minutes. After centrifugation the upper, aqueous layer was transferred to a new tube and cDNA precipitation by adding 7.5 M ammonium acetate and ice-cold absolute ethanol to the samples. The samples were then centrifuged at 12 000 *g* for 20 minutes and the pellet washed with ice-cold 80% ethanol. The pellet was dried in a DNA vacuum concentrator and samples suspended in sterile water. The cDNA samples were then labelled using cyanine 3 (Cy3) random nonamers (Roche). The Cy3-labeled cDNA samples were hybridised to the NimbleGen 385k Array slide using the NimbleGen Hybridization Kit (Roche) at 42°C for 16 hours.The array slide contained straight 60mer probes selected from the last 1500 bp of the 5105 sequences. Each sequence had an average of 8 probes and these were replicated 9 times on the array with additional random probes.The samples were scanned using MS 200 Microarray Scanner and the MS 200 Data Collection Software (Roche). The microarray scanned images were imported and data extracted using the NimbleScan software. The data was imported in ArrayStar (Roche) for analysis of expression profiles.

3.3 Results

3.3.1 Pathway reconstruction

Metabolic pathway reconstruction revealed at least 84 possible pathways that are functional in *T. lanuginosus* which contained 1231 metabolic reactions and were catalyzed by 798 proteins. This involved using a collection of 125 metabolic pathways templates with enzyme classes represented by their respective numbers. There are 113 gaps in the 84 reconstructed pathways, i.e. enzymatic steps that do not have a protein assigned to this function or gaps have been highlighted in blue. Only a few important metabolic pathways (**Figure 31-36**) will be described in detail but full presentation of the pathways are displayed in the **appendix (Figure S1-17)**.

Metabolic reconstruction revealed that this thermophilic fungus can use several common carbohydrates for energy production *viz*., glucose, sucrose, galactose, arabinose, mannose, xylose, fructose, trehalose, NAG, starch, glycogen and cellulose through the respiratory oxidative phosphorylation pathway. The organism uses glycolysis, pentose phosphate shunt, TCA cycle and hexose transformation for carbon transformation. This fungus also is able to perform gluconeogenesis which is a metabolic pathway that produces glucose from non-carbohydrate carbon substrates such as pyruvate, lactate, glycerol, glucogenic amino acids, and odd-chain fatty acids. Sugar transporter of these compounds uses 50 genes involved in sugar and organic major facilitator superfamily (MFS) transporters which make part of the 199 transporter genes involved in amino acids, peptides and other nutrient needs (**Figure 31-35 and Table S1 appendix**).

The cell wall is made from chitin and the cell membrane contains diglycerols, triglycerides, phosphotidyl inositol, phosphatidylcholine, ergosterol, ergosterol glucoside, ergosterol oleate, sphingosine, ceramide, phytosphingosine, phytoceramide and phosphatidylethanolamine. *T. lanuginosus* phosphatidyl groups consist of laureate, oleate, palmitate, myristate, stearate and linoleate. The *T. lanuginosus* genome has a ubiquitin degradation system controlled by N-end rule arginine-tRNA-protein transferase. *T. lanuginosus* is able to synthesis all the essential amino acids and is able to deaminate uric acid using the urea cycle to release nitrogen. Amino acids that contain sulphur are able to incorporate this element from their sulphate metabolism.

Also, purine and pyrimide synthesis takes place *de novo* and the salvage cycle also is used to recycle nucleotides. Heme is an essential molecule for the cell due to its involvement in many essential processes providing precursors for cytochrome prosthetic groups, photosynthetic pigments, and vitamin B12. Furthermore, *T. lanuginosus* contains the full pathway for heme biosynthesis (**Appendix, Figure S1-S17**).

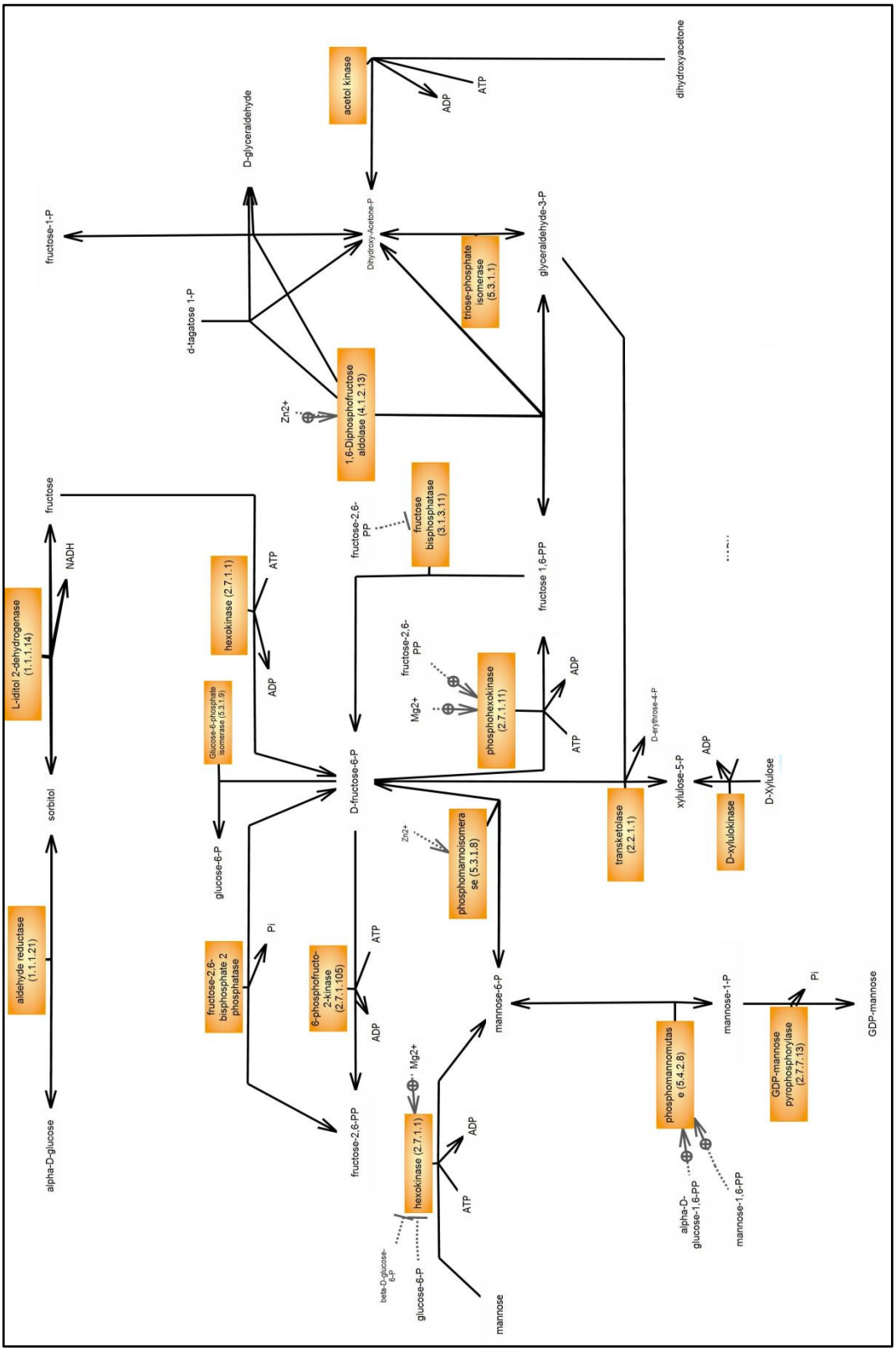

Figure 31: Predicted hexose transformation pathway in *T. lanuginosus*.

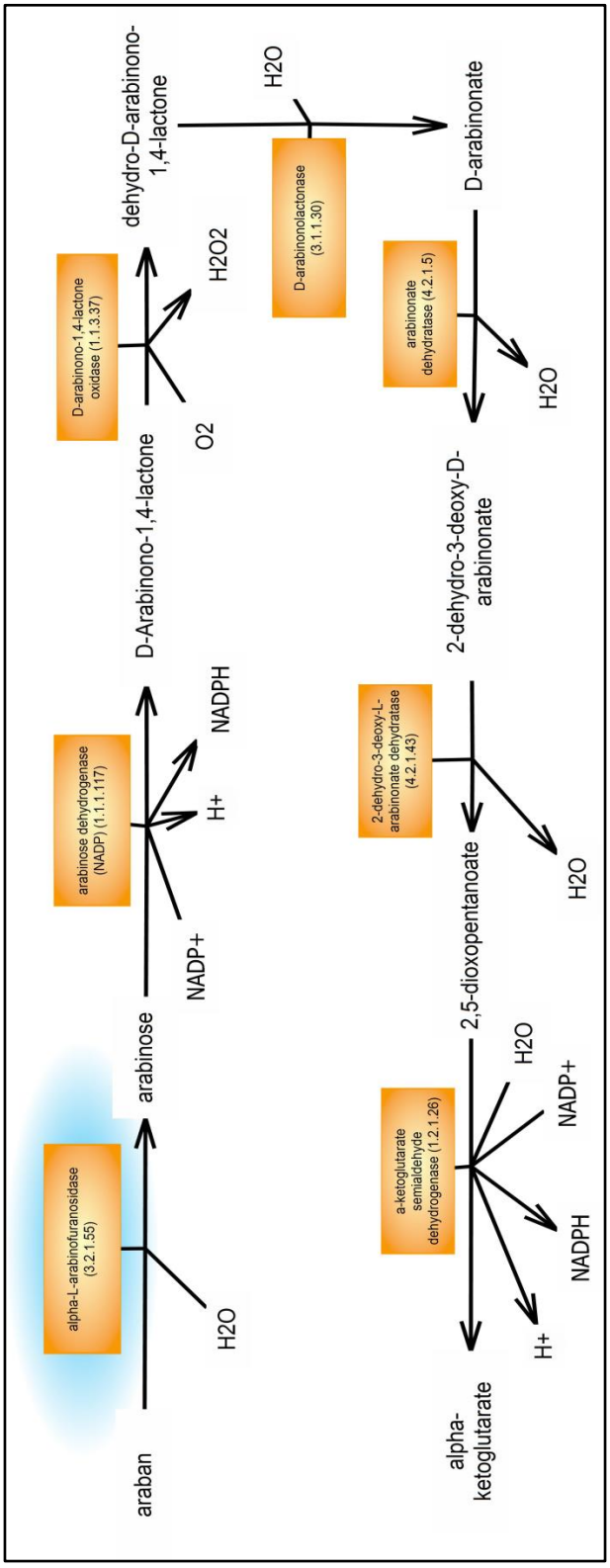

Figure 32: Predicted arabinose degradation pathway in *T. lanuginosus*.

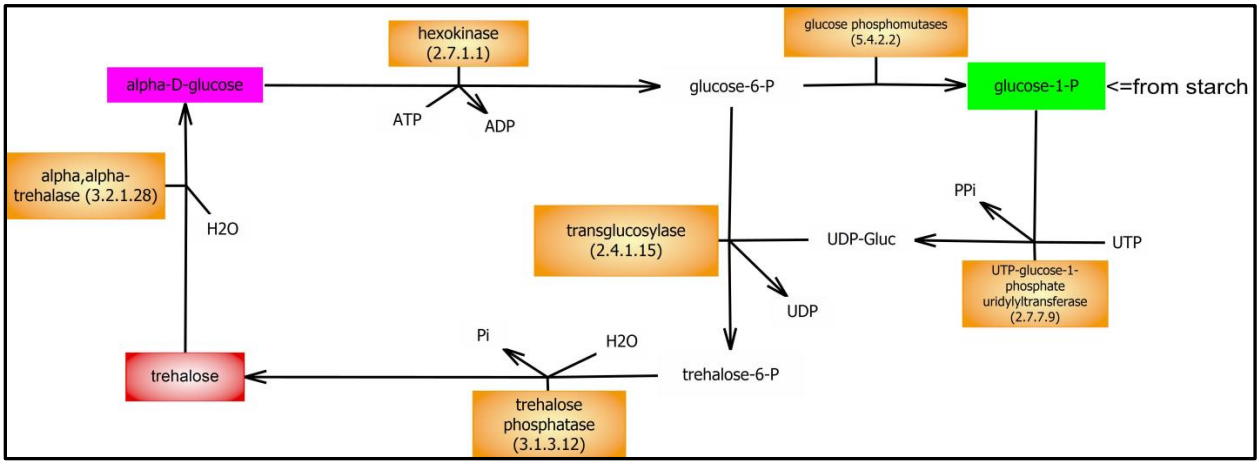

Figure 33: Predicted trehalose biosynthesis and conversion pathway in *T. lanuginosus*.

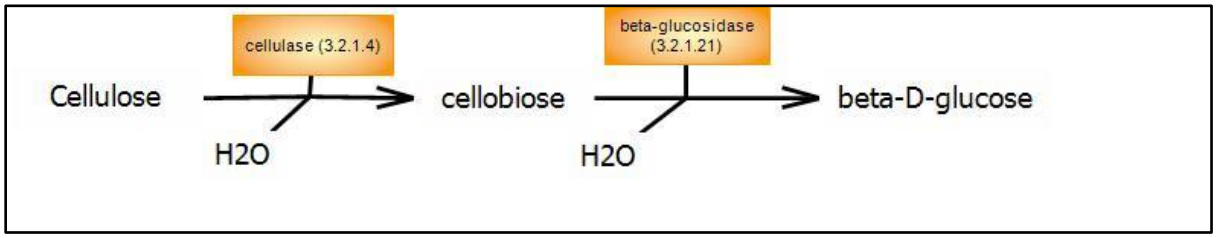

Figure 34: Predicted cellulose degradation pathway in *T. lanuginosus*.

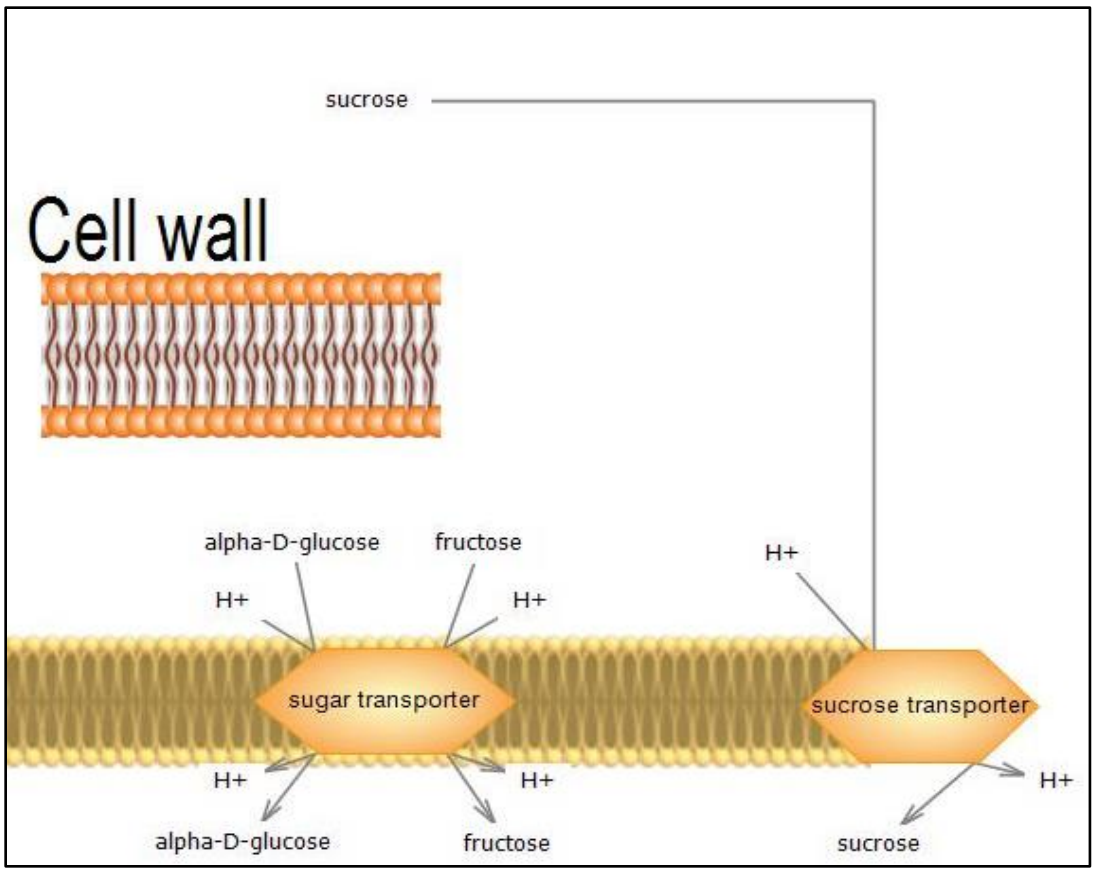

Figure 35: Predicted general sugar and sucrose transportation in *T. lanuginosus*.

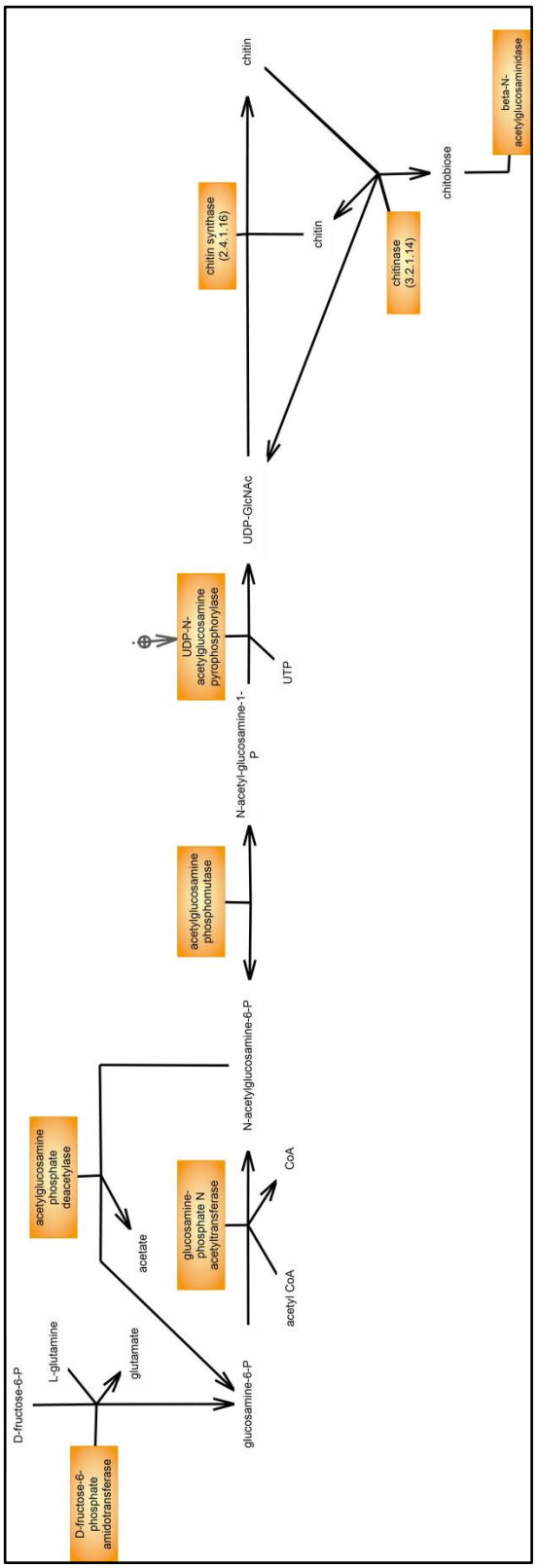

Figure 36: Predicted amino-sugar biosynthesis pathway in *T. lanuginosus*.

3.3.2 Phenotypic MicroArray

3.3.2.1 Cluster analysis of carbon source utilization profiles

The carbon source utilisation profile for *T. lanuginosus* was analysed using cluster analysis. The data generated was divided into 4 distinct clusters (**Figure 37 and 38)** based on the absorbance values obtained and final values (**Table 15 and 16)**. The analysis for general metabolism (490 nm) showed that cluster I and II contain carbon sources that lead to very slow growth of biomass formation and some even caused inhibition to the growth due to the absorbance being lower than water (highlighted in red in **Table 15 and 16**). For the control experiment containing water, it was grouped within Cluster II. The most dominant groups of this in these clusters are amino acids except for alanine and organic acids, esters, alcohols, phosphorylated sugars, rare sugars, rare polymers, nucleotide bases and aromatics groups. The trend was similar when mycelial growth was analyzed (750 nm) with exception that cluster I was larger than cluster II. Amino acids and some carbohydrates are also identified to give slow formation of biomass in these clusters.

Cluster III (490nm) depicted good metabolism of carbon sources for this strain, which led to reasonable growth. This group of carbon sources contained mainly carbohydrates which are monosaccharides *viz.,* sorbose, galactose, arabinose, ribose, fucose and rhamnose, disaccharides (lactose and lactulose), oligosaccharides (cyclodextrin, tagatose, gentiobiose and melibiose), some amino acids (asparagine and alanyl-glycine) and alcohol (sorbitol, glycerol, maltitol and xylitol). Cluster IV contained the carbon sources that enabled rapid growth and these included several monosaccharides and oligosaccharides *viz*., xylose, glucose, raffinose, glucose, fructose, cellobiose, maltose arabitol and N-Acetyl-glucosamine. High cellobiose utilisation was surprising as this organism is well known as a cellulase free organism.

The compounds found in clusters III and IV for growth (750 nm) were similar to clusters found in metabolism analysis (**Figure 38**), however group compositions were different. Following growth analysis most of the carbon sources clustered were in group III, while only three carbon sources (xylose, NAG and sucrose) were found in cluster IV which were classified as producing high biomass. Cellobiose also induced good biomass production and was placed in cluster III.

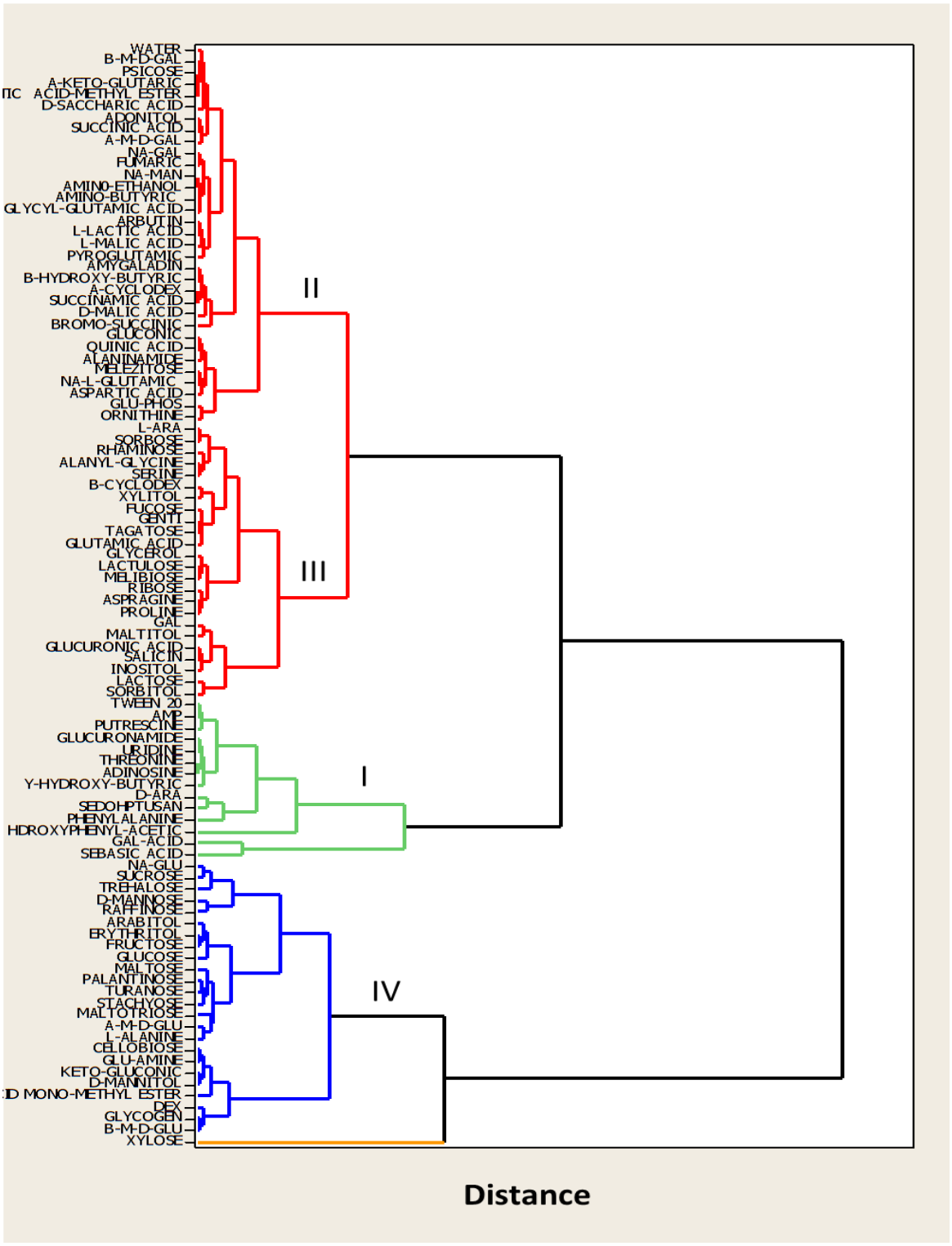

Figure 37: Joining cluster analysis applied to 95 carbon sources based on their metabolism of carbon sources by *T. lanuginosus* measured at 490 nm using the Biolog system.

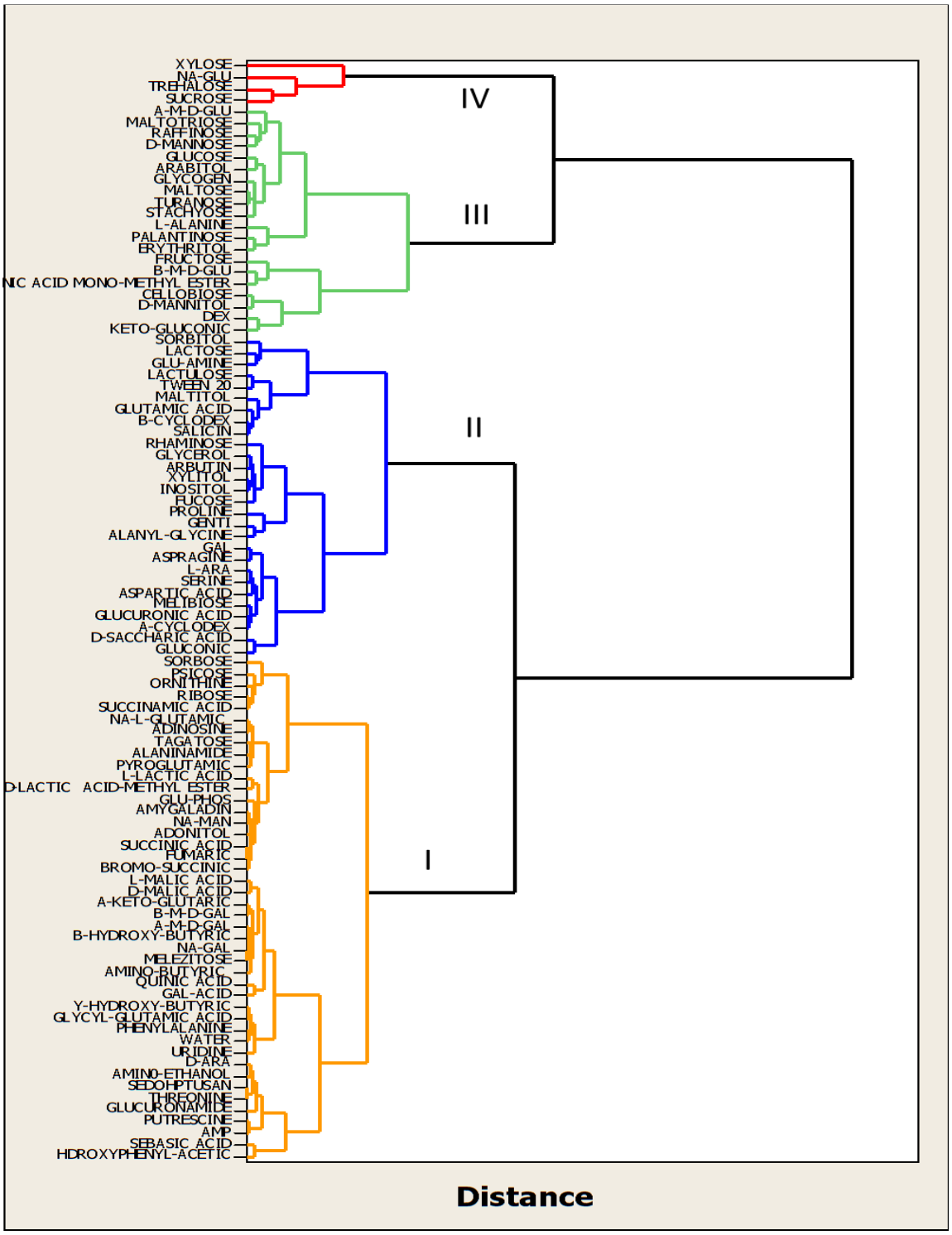

Figure 38: Joining cluster analysis applied to 95 carbon sources based on the growth of *T. lanuginosus* on these carbon sources. Growth was measured at 750 nm using the Biolog system.

Table 15: Metabolism of the 95 carbon sources by *T. lanuginosus* measured at 490 nm using phenotypic microarray

Compound	**Abs (490 nm)**
Xylose	3.15
Trehalose	2.63
Sucrose	2.57
N-Acetyl-glucosamine	2.54
D-mannose	2.50
Raffinose	2.48
α-Methyl-D-Glucoside	2.45
L-alanine	2.43
Maltotriose	2.41
Turanose	2.38
Palantinose	2.38
Stachyose	2.37
Maltose	2.34
Arabitol	2.31
Fructose	2.28
Glucose	2.28
Erythritol	2.27
β-Methyl-D-Glucoside	2.18
Dextrin	2.17
Glycogen	2.16
Succinic Mono-Methyl Ester	2.10
Glucosamine	2.06
Keto-D-gluconic acid	2.05
Cellobiose	2.05
D-mannitol	2.05
Sorbitol	1.91
Lactose	1.90
Maltitol	1.82
Galactose	1.81
Glucuronic acid	1.79
Salicin	1.78
Inositol	1.78
Glycerol	1.73
Lactulose	1.72
Melibiose	1.72
Ribose	1.71
Proline	1.69

Compound	Abs (490 nm)
Asparagine	1.68
Fucose	1.66
Glutamic acid	1.65
Gentibiose	1.65
Tagatose	1.64
B-cyclodextrin	1.61
Rhaminose	1.61
Xylitol	1.59
Serine	1.57
Alanyl-glycine	1.56
L-arabinose	1.54
Sorbose	1.53
Glu-phosphate	1.50
Ornithine	1.49
Aspartic acid	1.46
Gluconic	1.45
Quinic acid	1.45
Na-l-glutamic	1.45
Melezitose	1.43
Alaninamide	1.42
D-saccharic acid	1.40
Succinic acid	1.39
Adonitol	1.39
A-m-d-gal	1.38
Psicose	1.37
D-lactic acid-methyl ester	1.37
α-keto-glutaric acid	1.37
β-Methyl-D-Galactoside	1.36
Water (control)	1.36
L-lactic acid	1.36
Arbutin	1.35
L-malic acid	1.35
Pyroglutamic	1.32
Na-galactose	1.31
Fumaric	1.30
D-Malic acid	1.30
Na-mannose	1.29
Glycyl-glutamic acid	1.29
Amino-ethanol	1.29

Compound	Abs (490 nm)
Amino-butyric acid	1.29
Amygdalin	1.27
β-hydroxy-butyric acid	1.27
Succinamic acid	1.25
α-cyclodextrin	1.25
Bromo-succinic	1.22
γ-Hydroxybutyric acid	1.20
Uridine	1.18
Adenosine	1.17
Glucuronamide	1.17
Threonine	1.17
Putrescine	1.13
Adenosine-Mono-Phosphate	1.11
Tween 20	1.11
Sedoheptulosan	1.04
D-arabinose	1.02
Phenylalanine	0.93
Hydroxyphenyl-acetic	0.76
Galicic acid	0.49
Sebacic acid	0.27

Table 16: Growth analysis of *T. lanuginosus* in 95 different carbon sources measured at 750 nm using phenotypic microarray.

Compound	Abs (750 nm)
Xylose	1.43
N-Acetyl-Glucosamine	1.31
Trehalose	1.29
Sucrose	1.24
α-Methyl-D-Glucoside	1.15
Maltotriose	1.14
Raffinose	1.13
D-mannose	1.11
Glucose	1.1
Glycogen	1.1
Maltose	1.09
Turanose	1.09
Arabitol	1.08
L-alanine	1.08

Compound	Abs (750 nm)
Stachyose	1.07
Palantinose	1.04
Erythritol	1.02
Fructose	0.96
β-Methyl-D-Glucoside	0.94
Succinic acid Mono-Methyl Ester	0.91
Cellobiose	0.9
D-mannitol	0.88
Dextrin	0.83
Keto-D-Gluconic acid	0.81
Sorbitol	0.8
Lactose	0.77
Glu-amine	0.76
Lactulose	0.7
Tween 20	0.68
Maltitol	0.66
Rhamnose	0.66
Glutamic acid	0.66
B-cyclodextrin	0.65
Glycerol	0.65
Arbutin	0.64
Inositol	0.64
Salicin	0.64
Xylitol	0.64
Fucose	0.63
Proline	0.6
Genti	0.59
Alanyl-glycine	0.57
Galctose	0.55
L-arabinose	0.54
Asparagine	0.54
Serine	0.54
Melibiose	0.52
D-saccharic acid	0.52
Aspartic acid	0.52
α-cyclodextrin	0.51
Gluconic acid	0.51
Glucuronic acid	0.51
Sorbose	0.47
Psicose	0.46
Ornithine	0.45
Ribose	0.44

Compound	Abs (750 nm)
Succinamic ac	0.43
Na-l-glutamic	0.43
Adinosine	0.43
Tagatose	0.42
L-lactic acid	0.42
Alaninamide	0.42
Pyroglutamic	0.41
Adonitol	0.4
Amygadin	0.4
Glucose-phosphate	0.4
D-lactic acid-methyl ester	0.4
Na-mannose	0.39
Bromo-succinic acid	0.39
L-malic acid	0.39
Succinic acid	0.39
β-Methyl-D-Glucoside	0.38
Fumaric	0.38
α-keto-glutaric	0.38
D-malic acid	0.38
Quinic acid	0.38
Na-galactose	0.37
Galacic acid	0.37
Melezitose	0.37
α-Methyl-D-Galactose	0.37
Amino-butyric	0.37
β-hydroxy-butyric acid	0.37
γ-Hydroxybutyric acid	0.34
Water (control)	0.33
Glycyl-glutamic acid	0.33
Phenylalanine	0.33
Uridine	0.32
D-arabinose	0.31
Amino-ethanol	0.31
Glucuronamide	0.3
Sedoheptulosan	0.29
Threonine	0.29
Putrescine	0.29
Adenosine-Mono-Phosphate	0.28
Hydroxyphenyl-acetic	0.24
Sebacic acid	0.24

3.3.2.2 Hexose and pentose metabolism and growth induction

The metabolism and growth induction on the 95 carbon sources was analysed further by grouping into (a) hexose and pentose, (b) oligosaccharides and (c) amino acid based compounds whereas the remaining compounds were not assessed further. Analysis of hexose and pentose compounds revealed that xylose was best assimilated (490 nm) followed by trehalose, NAG and mannose. Xylose had around 15% more assimilation than the second best compound trehalose with absorbance values of 3.1 and 2.6. Similarly fructose, raffinose, glucose and turanose showed good general assimilation **(Figure 39).**

Biomass production measured at 750 nm showed that xylose also produced the highest biomass followed by N-acetyl-glucosamine and trehalose with absorbance values, 1.4, 1.36 and 1.23, respectively (**Figure 40)**. This was followed by glucose, mannose and raffinose, all with absorbance values above 2. Water metabolism was measured at 1.36 and for Tween 80 at 1.11. Nevertheless when their effect on growth was analyzed Tween 80 showed better biomass promotion than water with absorbance of values 0.68 and 0.33, respectively.

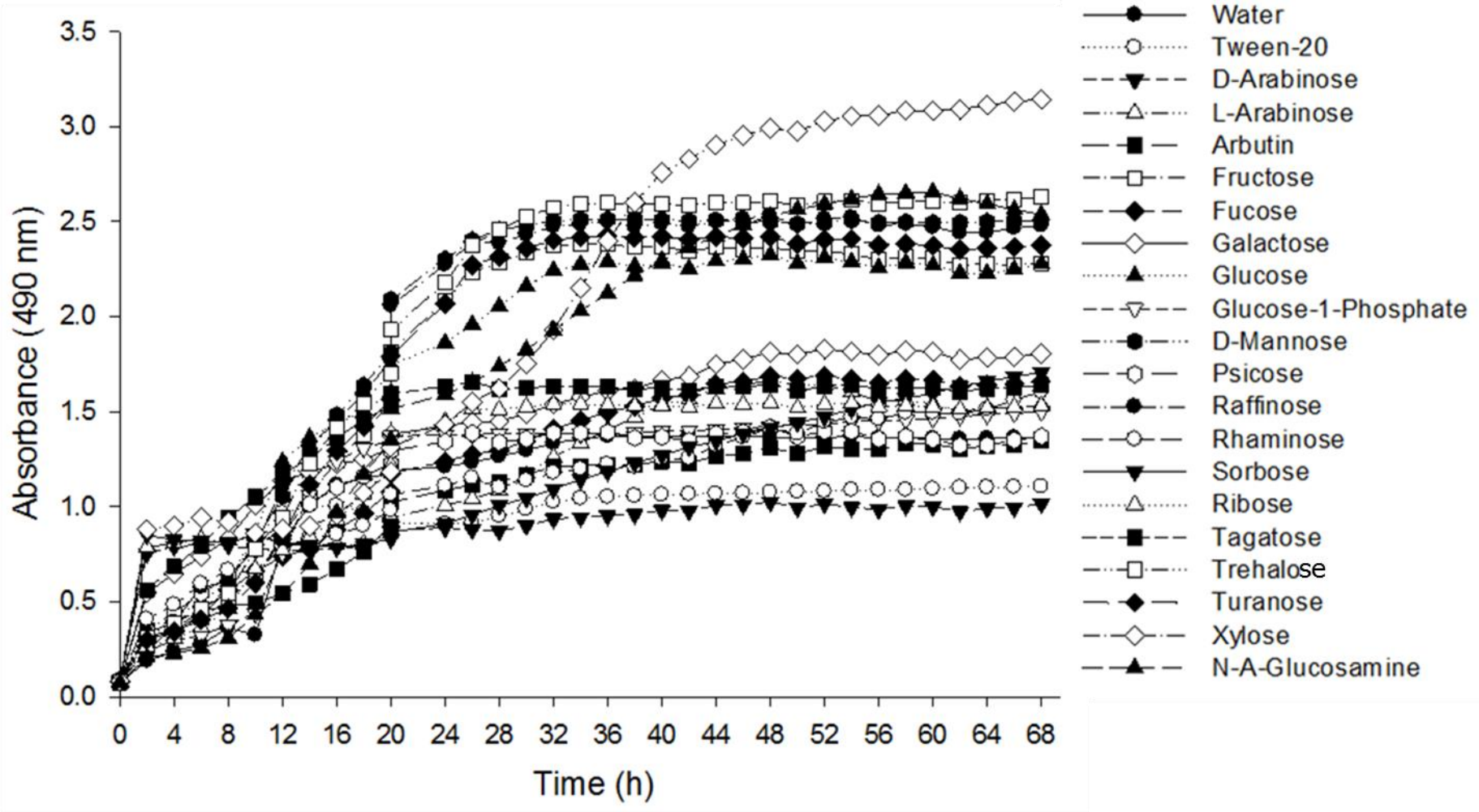

Figure 39: Metabolism of monomeric sugars (hexose and pentose) by *T. lanuginosus* SSBP. Metabolism was measured at an absorbance of 490 nm for 68 hours at 2 h intervals using Phenotypic MicroArray.

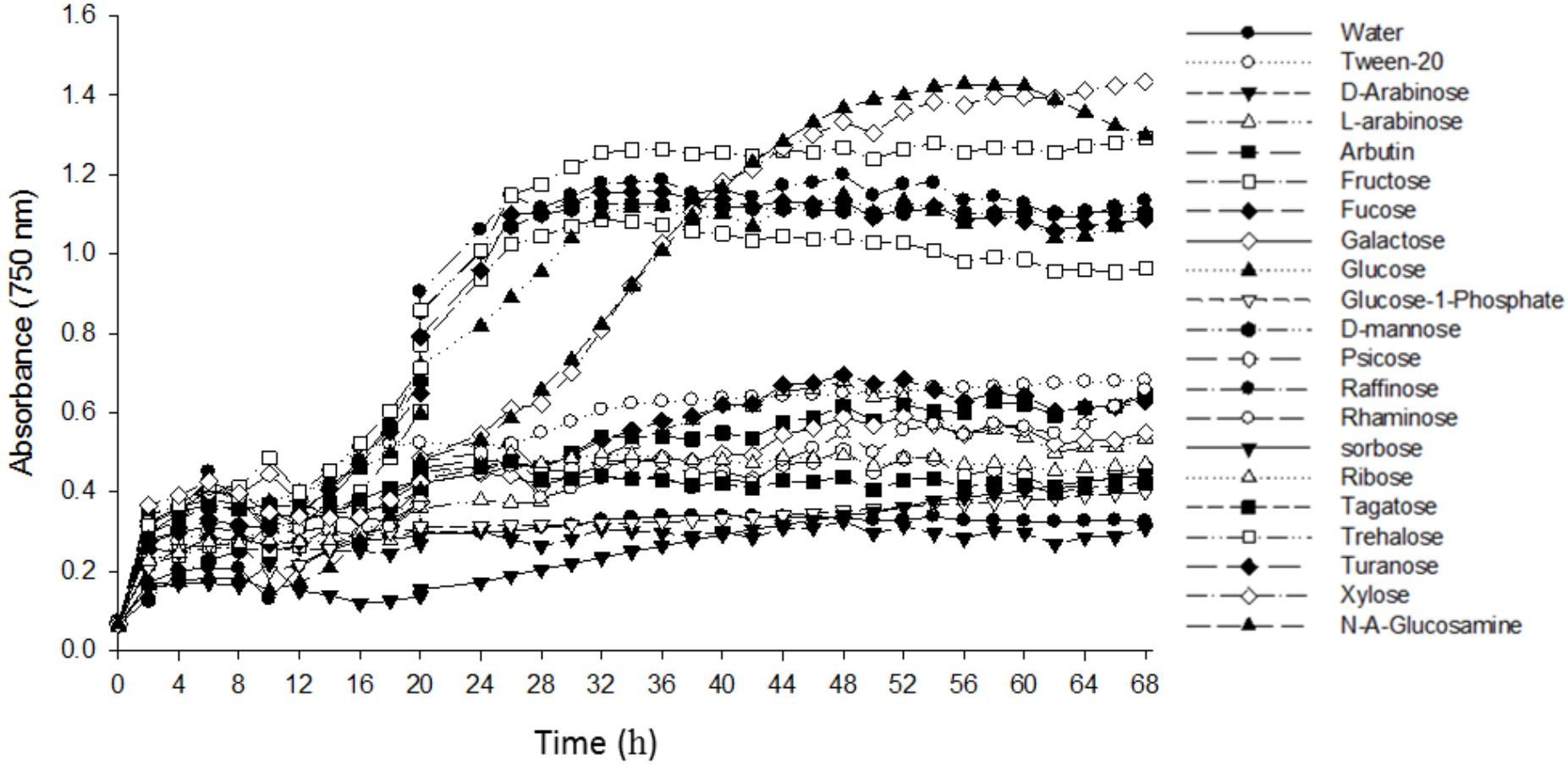

Figure 40: Growth of *T. lanuginosus* SSBP on monomeric sugars (hexose and pentose). Growth was measured at an absorbance of 750 nm for 68 hours at 2 h intervals using Phenotypic MicroArray**.**

3.3.2.3 Oligosaccharide assimilation and growth

Oligosaccharide assimilation analysis confirmed sucrose to be the best assimilated (490 nm) followed by maltotriose, stachyose, maltose, dextrin and glycogen (**Figure 41**). Interestingly cellobiose was well metabolise although *T. lanuginosus*, meaning if the glucanases 4 identified in the genome are fuctional, this organism can be used for biofuel production. Water assimilation was lower than most of common carbohydrates while the assimilation of rare occurring carbohydrate compounds (sedoheptulose and gentibiose) was even lower than water and Tween 80.

Sucrose produced the highest biomass among the oligosaccharides tested followed by maltose, glycogen, maltose oligosaccharides, stachyose, palantiose, cellobiose and dextrin (**Figure 42**). Again common carbohydrate compounds supported higher biomass production by *T. lanuginosus* than the rare compounds such as sedoheptulose and gentiobiose.

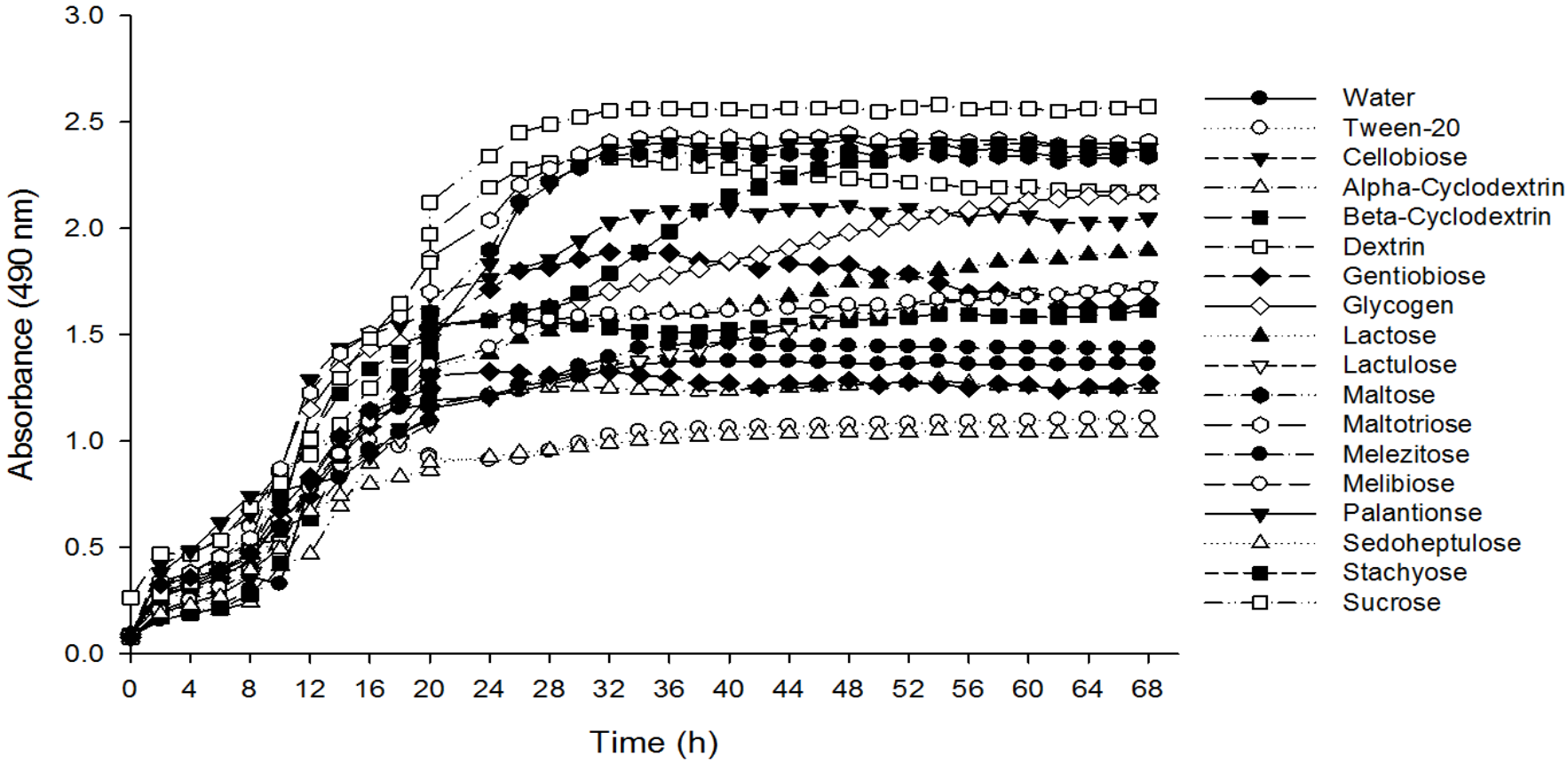

Figure 41: Metabolism of oligosaccharides by *T. lanuginosus* SSBP. Assimilation was measured at an absorbance of 490 nm for 68 hours at 2 h intervals using Phenotypic MicroArray.

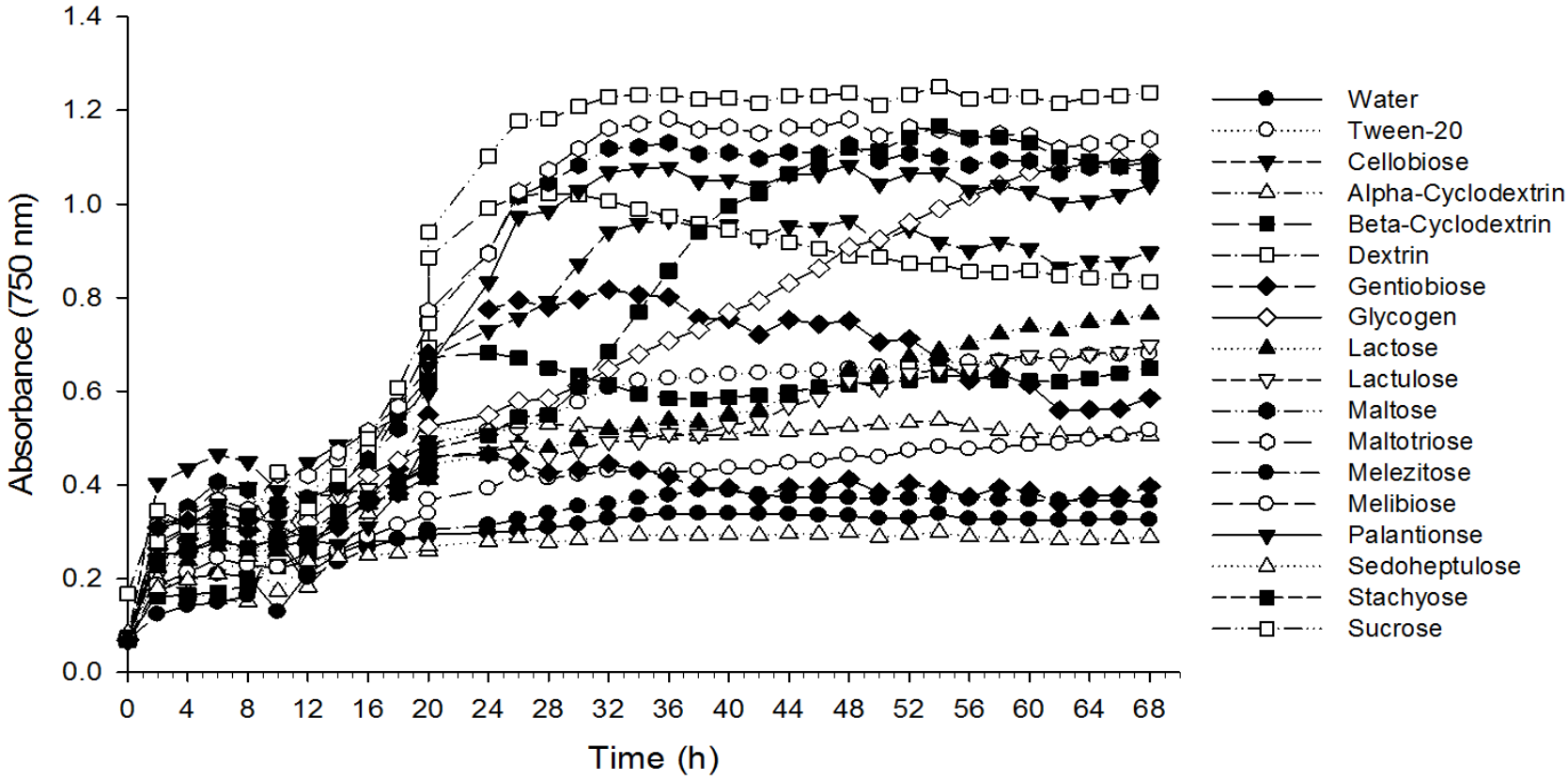

Figure 42: Growth of *T. lanuginosus* SSBP on oligosaccharide compounds. Growth was measured at an absorbance of 750 nm for 68 hours at 2 h intervals using Phenotypic MicroArray.

3.3.2.4 Amino acid assimilation and growth

The analysis of amino acid assimilation revealed that L-alanine was the best assimilated followed by proline, asparagine, and glutamic acid (**Figure 43**). Gylcyl-glutamic acid was the least assimilated even lower than water (control) and Tween 80. L-alanine also supported higher biomass production when compared to the other amino acids. This was in agreement with the cluster analysis resulted with grouped alanine among better metabolised compounds and biomass inducers. The other amino acid compounds produced less biomass than Tween 80 but more than water, except for threonine which was much lower (**Figure 44**).

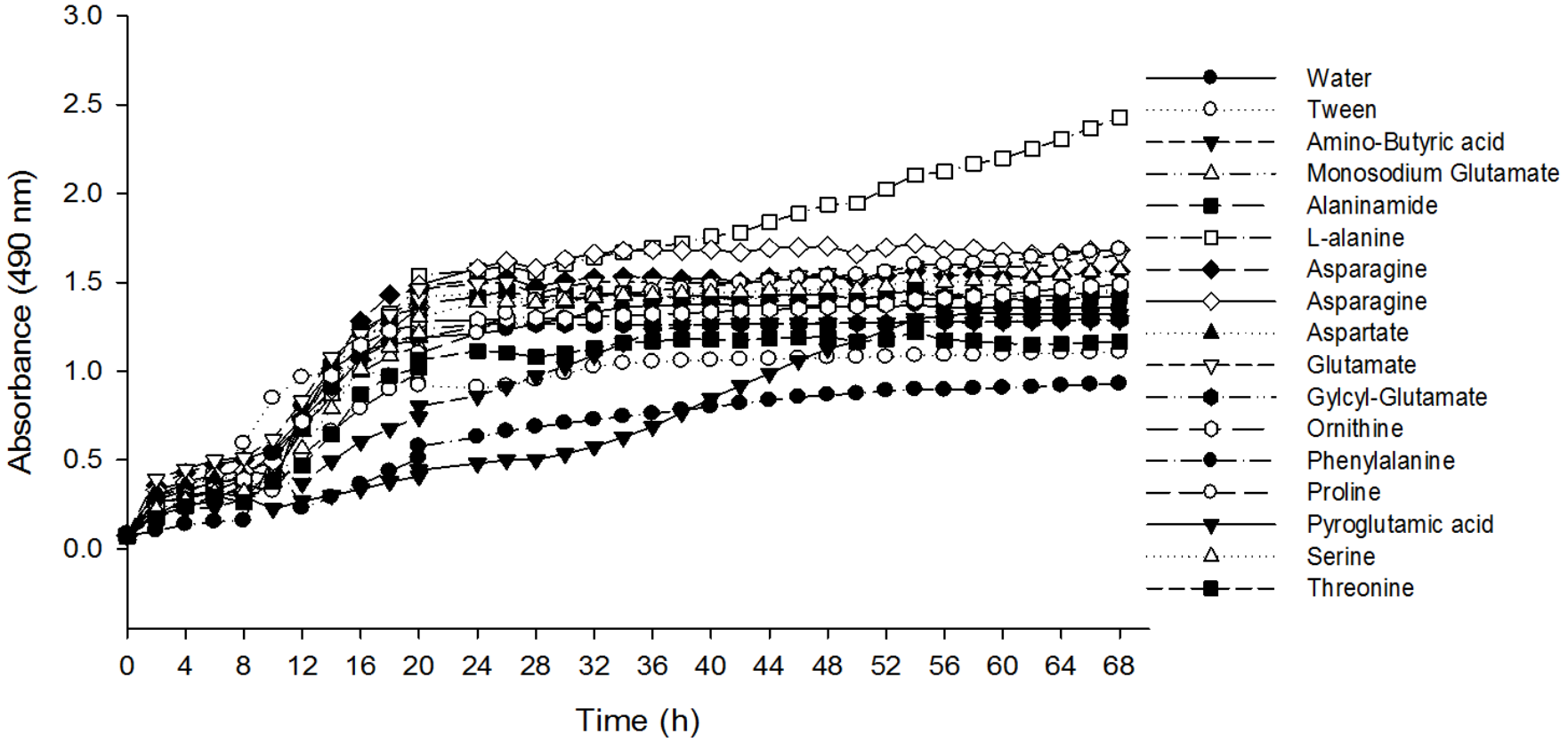

Figure 43: Metabolism of amino acid based compounds by *T. lanuginosus* SSBP. Metabolism was measured at an absorbance of 490 nm for 68 hours at 2 h intervals using Phenotypic MicroArray.

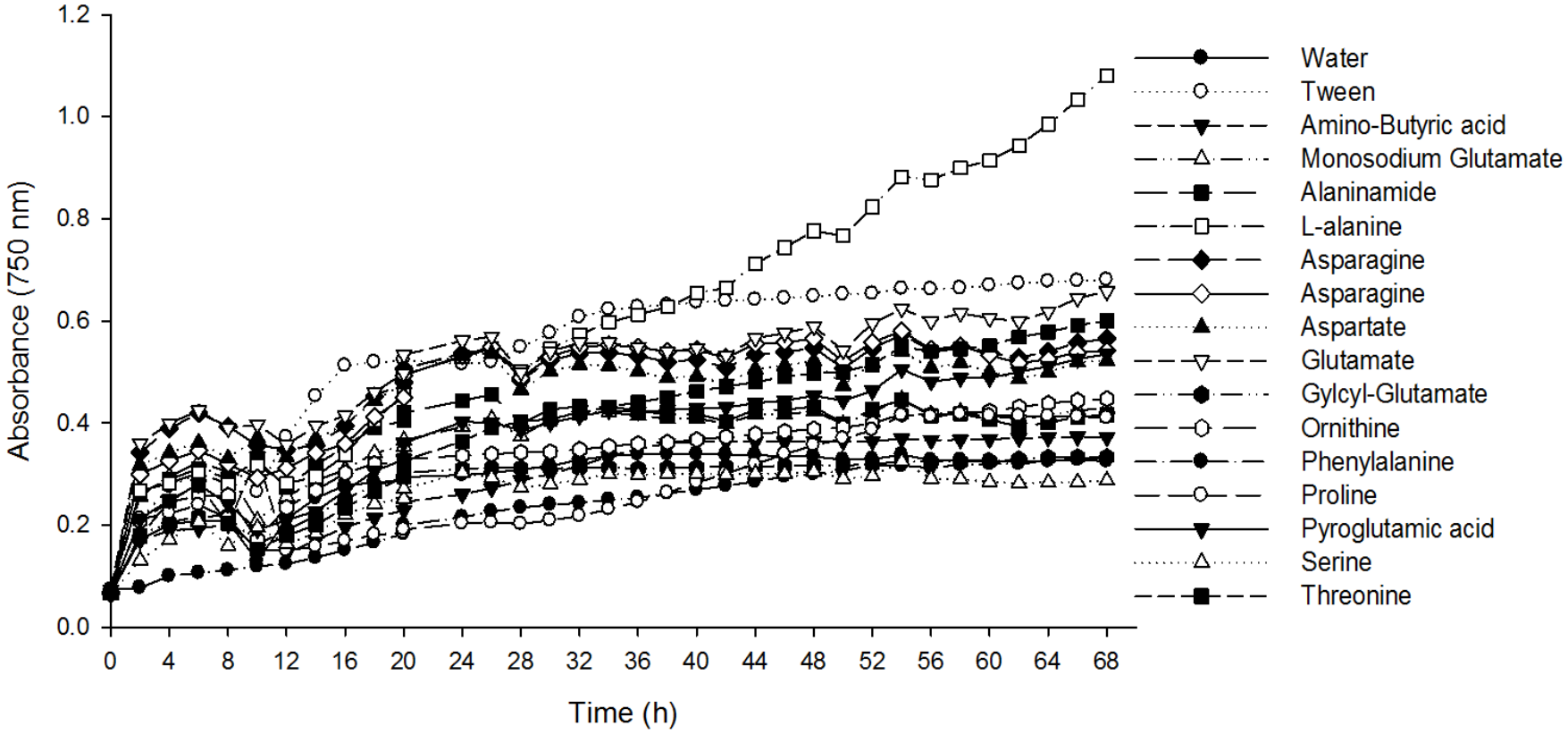

Figure 44: Growth of *T. lanuginosus* SSBP on amino acid based compounds. Growth was measured at an absorbance of 750 nm for 68 hours at 2 h intervals using Phenotypic MicroArray.

3.3.3 DNA microarray

3.3.3.1 Xylose

The gene expression profile of *T. lanuginosus* grown in xylose was analysed using DNA microarray designed using genome data. Microarray analysis revealed 208 proteins that were up-regulated by more than 2-fold in xylose compared to glucose. There were 54 proteins that were up-regulated 3-fold as shown in **Table 17**. Surprisingly, the highest up-regulated protein was a permease with 72-fold increase which was followed by α-xylosidase at an increase of 15-fold. As expected, other xylose-related proteins were also up-regulated including β-xylanase (7-fold), alcohol dehydrogenase (6-fold), xylose reductase (6-fold), β-xylosidase (4-fold) and xylulokinase (4-fold). Also, β-mannosidase expression was increased by 4-fold. Another surprise was up-regulation of endoglucanase 4 (4-fold) and 1,3-β-glucosidase, as these are normally involved in cellulose degradation (**Table** 17).

Table 17: DNA microarray of *T. lanuginosus* in xylose showing genes that were up-regulated 3-fold

Gene Prediction	**Fold change**
Permease	71.73
Alpha-xylosidase	15.42
glucosidase	13.93
Increased recombination centers protein 22-1	7.36
Versicolorin reductase	6.73
Endo-1,4-beta-xylanase	6.65
Zinc-type alcohol dehydrogenase-like protein C1773.06c	6.23
Zinc-type alcohol dehydrogenase-like protein	6.17
Uncharacterized oxidoreductase ygbJ	6.14
Probable glucan 1,3-beta-glucosidase A	5.95
Probable NAD(P)H-dependent D-xylose reductase xyl1	5.94
Adenylate kinase	5.62
Rhomboid family member 1	4.73
Probable beta-mannosidase B	4.25
Endoglucanase-4	4.23
Menaquinone biosynthesis methyltransferase ubiE	4.15
Regucalcin	4.09
Uncharacterized protein	4.07
Xylosidase	4.03
Tyrosine-protein phosphatase 3	3.98
Hexokinase	3.95
Regulatory protein abaA	3.91
Uncharacterized transporter	3.83
Probable D-xylulose kinase A	3.82
High-affinity glucose transporter	3.77
Conidiation-specific protein 10	3.74
Molybdopterin molybdenumtransferase	3.71

Tripeptidyl-peptidase sed3	3.65
Gene Prediction	**Fold change**
UPF0347 membrane protein	3.64
Maltose permease MAL31	3.64
Hit family protein 1	3.63
Menaquinone biosynthesis methyltransferase ubiE	3.59
Trehalose phosphorylase	3.58
Protein MNN4	3.51
Prestalk A differentiation protein A	3.51
mRNA turnover protein 4 homolog	3.48
Sorbitol dehydrogenase	3.4
Serum paraoxonase/arylesterase 1	3.39
Probable 3-hydroxymethyl-3-methylglutaryl-CoA lyase 2	3.39
Putative aminoacrylate hydrolase RutD	3.25
Nitrate reductase	3.21
Vesicle-associated membrane protein	3.21
Probable hydroxyacylglutathione hydrolase	3.17
DNA-directed RNA polymerase II subunit rpb4	3.13
Sterigmatocystin biosynthesis regulatory protein	3.12
Ribosome maturation factor rimM	3.11
Peroxisomal catalase 1 (Caenorhabditis elegans)	3.09
Alcohol dehydrogenase [NADP+] (Gallus gallus)	3.07
Sorbose reductase SOU1	3.07
31 kDa ribonucleoprotein	3.07
Probable 4-hydroxyphenylpyruvate dioxygenase 1	3.05
50S ribosomal protein L31	3.04
Prestalk A differentiation protein A	3.01

3.3.3.2 Glucose

Microarray analysis for glucose revealed 332 proteins that were up-regulated by more than 2-fold in glucose compared to xylose. There were 88 proteins that were up-regulated by 3-fold and 55 of them are shown in **Table 18**. In the glucose medium, a transporter protein (acetyl-CoA acetyltransferase) was the highest up-regulated gene (18-fold). Other up-regulated genes are well known to be involved in glucose metabolism included α-glucosidase (13-fold), glucose transporter (8-fold), High affinity glucose transporter (4-fold), another 1,3-β-glucosidase (4-fold), ABC-transporter (4-fold) and amylase (3-fold). Another interesting gene was up-regulated was the lipase gene (4-fold).

Table 18: DNA microarray of *T. lanuginosus* in glucose showing genes that were up-regulated 3-fold

Gene Prediction	**Fold change**
Acetyl-CoA acetyltransferase	18.52
Probable alpha/beta-glucosidase agdC	13.29
Cell surface glycoprotein 1	11.24
Ubiquitin-40S ribosomal protein S27a	10.04
Sterigmatocystin 8-O-methyltransferase	9.14
Probable glucose transporter rco-3	8.46
6-phospho-beta-galactosidase	6.27
Calcium-transporting ATPase 2	5.97
Leptomycin B resistance protein pmd1	5.68
Cytochrome b-c1 complex subunit 9	5.65
Tryptophan dimethylallyltransferase 1	5.58
Superoxide dismutase [Cu-Zn]	5.28
Cytochrome d ubiquinol oxidase subunit 1	5.27
Retinol dehydrogenase 12	5.21
Patatin	5.19
Class E vacuolar protein-sorting machinery protein hse1	5.18
Uncharacterized protein	4.92
Nonsense-mediated decay protein 4	4.9
Ribosomal N-lysine methyltransferase 4	4.87
IQ domain-containing protein F2	4.68
Lipase	4.56
Menaquinone biosynthesis methyltransferase ubiE	4.55
Lysine/arginine permease	4.55
Uncharacterized mitochondrial carrier	4.52
KiSS-1 receptor	4.48
Probable scytalone dehydratase	4.26
Ring assembly protein 3	4.26
High-affinity glucose transporter HXT2	4.22

Gene Prediction	Fold change
C-hordein	4.15
H/ACA ribonucleoprotein complex subunit 1	4.1
Probable electron transfer flavoprotein subunit alpha	3.98
Uncharacterized ATP-dependent helicase	3.98
Cobyric acid synthase	3.96
Glucan 1,3-beta-glucosidase	3.93
Protein transport protein Sec61 subunit gamma	3.92
Leucine carboxyl methyltransferase 1	3.91
Histone H4	3.89
60S ribosomal protein L10-A	3.8
Exodeoxyribonuclease 7 small subunit	3.74
Nucleoporin nup189	3.68
Uncharacterized membrane protein	3.67
Probable ATP-dependent RNA helicase	3.66
Uncharacterized transcriptional regulatory protein	3.61
Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase	3.58
ABC transporter ATP-binding protein/permease	3.55
External NADH-ubiquinone oxidoreductase 2	3.54
Stress-induced-phosphoprotein 1	3.54
Vegetative incompatibility protein	3.51
Cyclic AMP receptor 4	3.51
Alpha-glucosidase	3.49
GPI ethanolamine phosphate transferase 1	3.47
Uncharacterized transporter	3.46
Uncharacterized AAA domain-containing protein	3.45
Alpha-amylase A type-3	3.43

3.4 DISCUSSION

Functional and efficient metabolic pathways are vital for the survival of any organism. The biochemical assimilation and dissimilation of nutrients and including all enzymatic reactions within the cell and how such reactions are regulated, determine if these metabolic pathways function optimally. This chapter discusses metabolic pathway that were reconstructed from the genomic data, to validate the accuracy of genome annotation and the biochemical results obtained using a phenotypic microarray with 95 carbon sources, with emphasis on carbohydrate-based compounds as they would be more indicative of possible plant biomass degradation abilities.

Metabolic reconstruction revealed 84 functional metabolic pathways. Although more than 113 gaps were identified in the constructed pathways, this did not pose a major problem to complete predictions as most of the validated pathways contained all the necessary genes to be complete. The pathways that have gaps can be later filled or closed using the unknown group of genes which was obtained during annotation. This was because gene prediction revealed more than 800 genes with functions which were poorly characterised or unknown. Thus, the genes missing from these pathways maybe be part of this group with time, these genes may be further characterised as more functional characterisation of these genes or orthologs from other organisms become available.

It is well known that carbohydrates are usually the major carbon sources for fungi and metabolized to provide energy and precursors for the synthesis of cellular structures. Most organisms prefer carbohydrate monomers or simple sugars as carbon and energy sources since less energy is required to convert these compounds. In this study, possible metabolic pathways inferred from the *T. lanuginosus* genome provided further validation of the genome assembly accuracy. The metabolic pathways were constructed using metabolic templates from the KEGG database and other fungal genomes mention in section 3.2.2. *T. lanuginosus* grows in composting plant waste and requires O_2 for survival. Singh et al. (2000) showed that this organism grows very well on xylose and glucose. Glucose is used by most organisms as a source of energy and *T. lanuginosus* has both functional glycolytic and TCA pathways, as was expected. In these

pathways, glucose is converted to glucose-6-phosphate or fructose-6-phosphate then to pyruvate, through the glycolytic pathway. Under aerobic conditions, pyruvate is usually oxidized to CO_2 through in TCA cycle. Glycolysis and the TCA cycle perform a dual role in organisms. They produce energy in the form of ATP, NADH or NADPH and produce other molecules that are required in different pathways like nucleotide synthesis. Also as an alternative to glycolysis, *T. lanuginous* is predicted to possess the pentose phosphate shunt pathway (**Figure 36**). The oxidative phase of this pathway generates NADPH and the non-oxidative phase synthesises 5-carbon sugars. Although this pathway involves oxidation of glucose, it is primary involved in anabolism rather than catabolism. The data suggests that *T. lanuginosus* can also generate glucose using the gluconeogenesis pathway. This metabolic pathway generates glucose from non-carbohydrate carbon substrates such as glycerol, pyruvate, glucogenic amino acids, lactate and some fatty acids which is essential when the availability of glucose is low or decreases (Scholz *et al.*, 1973).

Beside glucose, fructose, galactose and mannose are also highly metabolized hexoses by *T. lanuginosus*. This fungus is able to take up these extracellular substrates using the sugar transport system (**Figure 35**). Fructose metabolism follows a similar path to that of glucose. Pathway analysis suggests that D-fructose is first converted to into D-fructose-1,6 biphosphate, which is cleaved by aldolase into dihydroxyacetone phosphate and glyceraldehyde. Dihydroxyacetone phosphate is isomerized into glyceraldehyde and enters glycolysis. The metabolism of mannose and galactose is different as these compounds derived from galactomannans which make up a component of hemicellulose (Naganagouda *et al.*, 2009). *T. lanuginosus* produces mannanases which cleave the backbone of galactomannans, thus the existence of a complete pathway for mannose and galactose is not surprising (Puchart *et al.*, 1999).

Trehalose utilisation has been discussed in chapter 2 and metabolic contruction validated the present of trehalose utilization and synthesis pathways (**Figure 33**). *T. lanuginosus* possibly uses trehalose as an energy source and accumulates trehalose to aid in thermostability. Trehalose is a non-reducing disaccharide composed of two α-(1,1)-linked glucose molecules. It is common in fungi and plants. Trehalose is produced in a multi-step process, whose prime substrate is glucose. Glucose is converted into glucose-6-phosphate which, together with UDP-glucose, leads to the

formation of trehalose-6-phosphate and subsequently trehalose. Trehalose can also be split into two molecules of glucose, thus closing the trehalose cycle. Trehalose is now being recognized as a crucial defence component that stabilizes proteins and biological membranes under various stress conditions, including increased temperature, hydrostatic pressure, desiccation, nutrient starvation, osmotic or oxidative stress, and exposure to toxic chemicals (Elbein *et al.*, 2003; Ferreira *et al.*, 2007; Fillinger *et al.*, 2001b; Garg *et al.*, 2002; Van Dijck *et al.*, 1995; Wiemken, 1990).

Another two common but important metabolic pathways in *T. lanuginosus*, are starch and glycogen degradation. Starch constitutes the most abundant rapidly renewable source of energy for living organisms. Starch is composed of 1,4-linked D-glucose residues (amylose) and branches containing α-1,4- and α-1,6-linked D-glucose residues called amylopectin. It is predominantly degraded by hydrolytic enzymes. *T. lanuginosus* has many enzymes responsible for starch degradation like presents α-amylase which cleaves α-1,4-glycosidic bonds, and α-1,6-glucosidase (cleaves α-1,6-glycosidic bonds) (Kunamneni *et al.*, 2005). Glycogen degradation occurs by the combined action of glycogen phosphorylase which releases glucose-1-phosphate, and a de-branching enzyme, which transfers a maltosyl unit to the end of an adjacent linear α-(1,4)-chain and releases glucose. A two-step reaction catalyzed by phosphoglucomutase and UDP-glucose pyrophosphorylase leads to the synthesis of UDP-glucose of this energy molecule. Another potential function of glycogen and trehalose is the progression of cell division. Mobilization of storage carbohydrates temporarily increases the sugar flux, thus enabling the cells to go through the next round of division (Guillou *et al.*, 2004). Besides these common carbohydrate compounds, genome analysis, *T. lanuginosus* or ancestors might possess the ability to hydrolyse cellulose (**Figure 34**) as four cellulases (which are responsible for the hydrolysis of cellulose) and four ß-glucosidases to produce glucose were identified. DNA microarray also revelead that one of the genes identified as a glucane was induced by xylose (**Table 17**).

Pentose sugars represent one of the most abundant sources of energy but not all organisms are able to utilize them. However, *T. lanuginosus* grows as well or better on the pentose sugar xylose than glucose, which makes this organism important. This may indicate that the mechanism for a xylose uptake and utilization in this organism is highly efficient, prpbably using the hexose

transformation pathway (**Figure 31**). Moreover, our current knowledge of transcriptional regulation of thexylose utilisation pathway in *T. lanuginosus* is not well understood. This prompted us to perform a detailed analysis of xylose utilization pathway in *T. lanuginosus*. There are generally two major pathways for xylose utilization in microorganisms. The first one is known as "Oxo-Reductive" or "Xylose Reductase-Xylitol Dehydrogenase" pathway which is used by fungi (Seiboth et al., 2003). In this pathway xylose is converted to xylitol by xylose reductase, then to xylulose by xylitol dehydrogenase. In the last step D-xylulose is phosphorylated by an ATP-utilizsing kinase, xylulose kinase, to result in D-xylulose-5-phosphate which is an intermediate of the pentose phosphate pathway. The second pathway is the isomerase pathway which is generally found in bacteria (Gu *et al.*, 2010; Mach-Aigner *et al.*, 2010). In this pathway, the enzyme xylose isomerase converts D-xylose into D-xylulose. D-xylulose is then phosphorylated to D-xylulose-5-phosphate as in the Xylose Reductase-Xylitol Dehydrogenase pathway, so in this pathway xylitol is not produced and it a two-step process instead of three. Using metabolic reconstruction, *T. lanuginosus* was found to be able to convert xylose using the Xylose Reductase-Xylitol Dehydrogenase pathway. Although in our analysis there is no evidence to explain the reason for high xylose utilization by this fungus, this can be hypothesised that perhaps the rate of xylose assimalation is relatively high in *T. lanuginosus* as DNA microarray showed an up-regulation by 72-fold of a permease gene. Another pentose sugar that this fungus can use is arabinose (**Figure 32**). Although *T. lanuginosus* is missing α-arabinofuranosidase enzyme, this fungus is still able to utilize arabinose as this enzyme is only required for the breakdown of the araban polymers (arabinoxylans and arabinogalactans) which may be provided by other organisms in the composting environment (Bernhard and Benjamin, 2011).

Electron transfer from NADH to molecular oxygen occurs via a linear respiratory chain, which is indispensable for ATP synthesis in most catabolic pathways. The mitochondrial chain includes four electron-transferring oligomers, which are located in the inner membrane of the organelle. The energy released from electron transfer reactions is used to pump protons across the membrane outwards, thus generating an electrochemical gradient that can be used to synthesize ATP by ATPase or complex V (Hannemann *et al.*, 2007; Joseph-Horne and Hollomon, 2000). The whole process is known as oxidative phosphorylation and is largely conserved in most

organisms. *T. lanuginosus* performs oxidative phosphorylation by classical means, in that genes corresponding to components of complexes I, II, III, and IV were identified in the metabolic reconstruction (Marcet-Houben *et al.*, 2009).

The complete pathway for fatty acids biosynthesis and metabolism includes 48 genes, distributed equally in fatty acid biosynthesis and degradation. Metabolic reconstruction revealed that cell membranes and wall synthesis pathwaysin *T. lanuginosus* has the following biomolecules with cell membrane function, diglycerols, triglycerides, phosphotidyl inositol, phosphatidylcholine, ergosterol, ergosterol glucoside, ergosterol oleate, sphingosine, ceramide, phytosphingosine, phytoceramide and phosphatidylethanolamine. Also, metabolic reconstruction revealed that the cell wall from *T. lanuginosus* is made from chitin so the existence of chitinase genes was not surprising as chitinase gene form other strain of T. lanuginosus have been isolated (Guo *et al.*, 2008; Hartl *et al.*, 2012). The chitin is made up of an amino sugar known as *N*-acetylglucosamine (**Figure 36**) which is synthesis by D-fructose to glucosamine followed by acetylation using acetyl-CoA.

Amino acids can be regarded as connectors of metabolic pathways and *T. lanuginosus* bio-synthesizes all amino acids, although only alanine biosynthesis is shown in (**Appendix-Figure S14**). Amino acids are the monomers of proteins and they are crucial to the synthesis of various biologically important molecules, such as the nitrogenous bases of nucleic acids, coenzyme electron carriers, and the oxygen-binding ring of hemoglobin. They are also structural components of hormones and neurotransmitters. When these molecules are deaminated, their carbon component can enter the TCA cycle for energy production. Also, the urea cycle is connected to amino synthesis as it provides nitrogen, which is important in arginine and ornithine synthesis (Cruz *et al.*, 2007). Arginine is one of the most versatile amino acids in eukaryotes, being used as a precursor for the synthesis of proteins and for a great range of compounds, including nitric oxide, urea, polyamines, proline, glutamate, creatine and agmatine (Bauer *et al.*, 2010).

In filamentous fungi and yeasts, sulphur metabolism is involved in amino acid biosynthesis and in other major synthetic pathways. Thus, these organisms possess complex regulatory circuitry

that governs the expression of a diverse set of permeases and enzymes used to take up and assimilate a steady supply of sulphur. Most sulphur in living organisms is present in a reduced form, as organic thiols. For their synthesis, inorganic sulphate is reduced and incorporated into organic compounds in the assimilatory sulphate reduction pathway. After uptake by the cells, inorganic sulphate is first phosphorylated via adenosine triphosphate in two enzymatic steps and then to sulfide, which is condensed with *O*-acetylserine to generate cysteine, which also serves as an intermediate in the synthesis of methionine and S-adenosylmethionine, of genome analysis of *T. lanuginosus* revealed that this fungus contains a highnumber of these genes. It is known that a sulphur regulatory circuit operates to insure that the cells maintain an adequate source of sulphur and, conversely, to repress the synthesis of various sulphur catabolic enzymes, whenever cells have an adequate supply of the substance. Regulation may occur both at the level of entry of sulphur from various sources into the assimilatory pathway and via distinct steps within the main pathway itself (Banszky *et al.*, 2003; Maruyama-Nakashita *et al.*, 2004). Sulphur metabolism has also been implicated in cellulase gene expression in the filamentous fungus *T. reesei* (Gremel *et al.*, 2008).

De novo purine synthesis appears to be identical in nearly all organisms, and it begins with their metabolic precursors: amino acids, ribose 5-phosphate, CO_2, and NH_3, and this was also the case in *T. lanuginosus*. Although the free bases are not intermediates in the *de novo* pathways, they are intermediates in some of the salvage pathways (**Figure 49**). Analyzing data concerning *de novo* purine synthesis enzymes in the *T. lanuginosus*, seven enzymes were identified in pathway out of the 15 that are known from all organisms. *De novo* pyrimidine biosynthesis proceeds differently from purine nucleotide synthesis. The six-member pyrimidine ring is made first and then attached to ribose 5-phosphate, and carbamoyl-phosphate is required in this process. Also, in *T. lanuginosus*, genes for the pyrimidine pathway are present. Thus it is likely that *T. lanuginosus* is able to produce both purines and pyrimidines *de novo*. Salvage reactions are important in the metabolism of purine nucleotides, because of the energy required for the synthesis of purines (Nara *et al.*, 2000). A free purine base that has been cleaved from a nucleotide can produce the corresponding nucleotide by reacting with the compound phosphoribosylpyrophosphate, formed by a transfer of a pyrophosphate group from ATP to ribose-5-phosphate. Two different enzymes with different specificities catalyze salvage

reactions: adenine phosphoribosyltransferase and hypoxanthine-guanine phosphoribosyltransferase. Analysis of the genome also indicated that *T. lanuginosus* can produce nucleotides through the salvage pathway.

T. lanuginosus transporter systems contain 199 proteins, largely due to importers for organic nutrients like sugars, amino acids and peptides. Prediction for transport proteins included paralogs (**Appendix-Table S1,**) for sugar (**Figure 35**) and organic MFS transporters (50), ABC superfamily transporters (21), mitochondria carriers (31) and amino acid permeases (10). Transporters in these families use ATP hydrolysis as the primary source of energy to drive the transport process. These transporter proteins may participate in the uptake of several organic nutrients and inorganic compounds, as well as the export of iron compounds, peptides and multiple drugs. *T. lanuginosus* grows very well on sugar molecules which, this corroborates with the presence of a large group of transportation genes.

Heme plays a very important role in many cellular processes. It is an important cofactor for proteins involved in oxygen transport and storage, mitochondrial electron transport complexes (II–IV), drug and steroid metabolism (cytochromes), signal transduction (nitric oxide synthases, soluble guanylate cyclases) and transcription and regulation of antioxidant-defense enzymes. Heme is also a regulatory molecule. Its cytosolic to nuclear ratio and the absolute amount of its concentration affects gene transcription and translation. Thus, the intracellular heme biosynthesis and regulation is vital. *T. lanuginosus* has a complete pathway for heme biosynthesis (**Figure 53**) and it is important since this organism is an aerobe but thrives in environments where oxygen availability varies regularly (Gazi *et al.*, 2007; Sharma *et al.*, 2008; Tang *et al.*, 2004; Yu *et al.*, 2007).

In nature, the ability of a microorganism to use a variety of compounds is vital for survival in a composting environment, as different substrates are degraded and utilised by different organisms. Filamentous fungi play a vital role in this ecological dynamics as they are responsible for the majority of the hydrolysis reactions (Tang and Katayama, 2005; Yu *et al.*, 2007). *T. lanuginosus* is among those fungal organisms that thrive in such environments, with an added ability to survive high temperature which only a select few eukaryotic organisms can achieve

(Puchart *et al.*, 1999). The analysis for carbon source for metabolism and utilisation for biomass production in this organism revealed a similar profile to other filamentous fungal studies of this nature where glucose, xylose, trehalose and NAG produced high biomass in *T. reesei* and *A. niger* (Druzhinina *et al.*, 2005). This also corroborated with previous finding where *T. lanuginosus* was grown on the above substrate and produce good biomass (Kuttanpillai *et al.*, 2009a; Purkarthofer and Steiner, 1995; Singh *et al.*, 2000b). Although, the clusters in these studies were similar to our findings, closer analysis of cluster IV revealed that for *T. lanuginosus*, xylose is the preferred sugar compared to glucose. This concurs with reports that *T. lanuginosus* has the most efficient system for xylanase production and xylose utilisation and thus it was expected xylose would produce the most biomass and have the highest assimilation (Dodd *et al.*, 2011; Kuttanpillai *et al.*, 2009b; Singh *et al.*, 2000a)

Trehalose also induced high biomass and was metabolised the best by *T. lanuginosus*. The reason for this is that trehalose can be used by the organisms as an energy source. However, there is a more important reason in thermophilic organisms (Argüelles, 2000; Gancedo and Flores, 2004). Trehalose has been widely reported as a part of the physiological adaptation to various environmental stresses e.g. high temperature, in yeasts and filamentous fungi (Ferreira *et al.*, 2007). Also, NAG was high metabolised and produce high biomass because it is the building block of fungal cell walls which contain chitin and also can be converted to energy molecules therefore high metabolism and the ability to support growth was expected (Hartl *et al.*, 2012). It was surprising that only one amino acid, alanine, produced significant biomass. This may be because alanine is one of the few amino acids that can transform into glucose and can be used in TCA cycle to provide energy for the cell and thus may be preferred by this fungus to supplement the supply of mineral nitrogen and energy (Bechem, 2012).

The phenotypic microarray had confirmed that *T. lanuginosus* was able to metabolise xylose and glucose equally well (Chen *et al.*, 2012a; Christopher *et al.*, 2005; Singh *et al.*, 2003). However, the identity of other genes involved in xylose utilisation and reasons for this organism using xylose efficiently, were still unanswered. The power of the microarray approach was demonstrated by the fact that it was able to show clear differences between up-regulated genes when the organism was grown in glucose or xylose. The most interesting find was the up-

regulation of a permease gene (sharing some similarity with other xylose permeases or transporter) which was the most up-regulated gene for glucose and xylose analyses (**Table 17 and 18**). This data provided clues how *T. lanuginosus* evolved to be an efficient utiliser of xylose. *T. lanuginosus* secretes huge amounts of β-xylanase which degrades xylan randomly. The funfus also produces at least two different xylosidases to convert xylooligosaccharides into xylose monomers. Most of the xylose and xylo-oligosaccharrides would be outside the organism thus available for all organisms associated with composting environments. However, the presence of an efficient permease would be an advantage to the organism as it will be able to transport the xylose into the cell rapidly (Cook *et al.*, 1993). Also, xylose induction of xylan-related genes was confirmed in this analysis as the xylanase and xylosidase genes were up-regulate by more than 5-fold (Kuttanpillai *et al.*, 2009a; Puchart *et al.*, 1999; Purkarthofer and Steiner, 1995; Singh *et al.*, 2000b). Also up-regulated were xylose utilisation genes such as xylose dehydrogenase, xylose reductase and xylose kinase. This shows that genes predicted in the genome are actively expressed.

Another interesting find, was the up-regulation of glucanase by xylose. These results support the genome data where these genes were identified to be present *in T. lanuginosus* although no previous report has ever report this enzyme in this organism. The up-regulation of the glucanase and glucosidase with xylose suggests that in *T. lanuginosus* the regulation of these two systems are interlinked as the substrate (xylan and cellulose) in nature although the xylan system seems more dominant system. Thus, further investigation of a possible cellulose degradation pathway is necessary to understand it as this is the first DNA microarray report for this organism. In glucose grown culture, most of the up-regulated genes were expected *viz*., a transporter protein acetyl-CoA acetyltransferase, 3 glucosidases, glucose transporters and amylase (Chambergo *et al.*, 2002). However, induction of the lipase gene was unexpected. This information can provide a base for further studies of this enzyme as it is being used for industrial application (Hasan *et al.*, 2006; Saxena *et al.*, 1999).

Genome sequencing and metabolic reconstruction can provide clues on how organisms are organised. However, inactive (damage/mutated) genes may confound results. However, phenotypic and cDNA microarray can be used to compliment the data and obtain the true picture.

In conclusion, this study indicates that *T. lanuginosus* is a versatile organism that can utilise a diverse range of carbon sources, including carbohydrates, amino acids, carboxylic acids, polymers, aromatics, esters, phosphorylated and sugar alcohols. The application of phenotypic and DNA microarray as an investigative tools for carbon utilization studies is a quick approach to studying and assessing filamentous fungi for specific activities.

CHAPTER 4: GENERAL DISCUSSION

4.1 RESEARCH PERSPECTIVE

Fungi play a critical role in the environment through the decomposition of organic material and through symbiotic relationships with prokaryotes, plants, animals and their economic significance. Humans have been using fungal organisms for millennia in processes like brewing, baking and cheese making. Industrial applications of fungal organisms and their biocatalysts have been identified and used for human benefit. The application of enzymes, especially, has been largely responsible for the development of biotechnology, with applications in industries, such as starch, food and beverage, detergents, pulp and paper, textile, leather, forest products, animal feed and biofuels. In addition, the environmentally-friendly approach to industrial production has been led by the application of microbial enzymes. Fungi continue to be an important source of enzymes, both in terms of their role as a source for biological diversity and as a host for production of industrial enzyme products (Galante *et al.*, 1998; Ghorai *et al.*, 2009; Maheshwari *et al.*, 2000; Marquez *et al.*, 2011; Pariza and Cook, 2010).

For the identification and application of new enzymes, scientists have concentrated on a group of organisms know as extremophiles. Extremophiles are organisms that have evolved to exist in a various extreme environments which include halophiles, thermophiles, acidophiles, alkalophiles, psychrophiles, and barophiles. Biocatalysts derived from these microorganisms are stable in diverse conditions. Mesophilic enzymes are often not well suited for the harsh conditions which prevail in industrial processes (Niehaus *et al.*, 1999). Although many methods are available for enhancing biocatalysts to make them more suitable for industrial applications, the search for novel enzymes still continues and presents the greatest challenge to switching from chemical to green processes. The development of genomics has given renewed hope in this area as more data is becoming available from genomes of high temperature microorganisms.

However, genomic sequencing of fungi especially filamentous fungi has been limited. *S. cerevisiae* was the first eukaryotic and fungal organism that was sequenced (Goffeau et al. 1996).

However, the sequencing of *S. pombe* (Wood et al. 2002) and *N. crassa* (Galagan et al. 2003) revealed that the yeast was no longer a model organism for all other fungi. The genome of *N. crassa*, the first filamentous fungus to be sequenced, was double in size and even *S. pombe* lacked homologs to known yeast genes. Several fungal organisms have been sequenced which includes: *Candida albicans*, *Yarrowia lipolytica*, *Aspergillus clavatus, A. flavus, A. fumigatus, A. nidulans, A. niger, A. terreus*, *Coccidioides immitis* (strain RS), *Cryptococcus neoformans, F. graminearum, F. oxysporum*, *Fusarium verticillioides* (*Gibberella moniliformis*), *Magnaporthe oryzae*, *Puccinia graminis*, *Rhizopus oryzae* and *T. reesei* (Stajich *et al.*, 2012).

However, very few thermophilic fungal genomes are available, even though these thermophilic organisms can be a rich source of industrial enzymes. *T. lanuginosus* is one such organism which has the ability to produce a variety of enzymes that have industrial application (Bennett *et al.*, 1998; Chen *et al.*, 2012a; Kuttanpillai *et al.*, 2009b; Puchart *et al.*, 1999; Purkarthofer *et al.*, 1993; Singh *et al.*, 2003; Singh *et al.*, 2000b). In lieu of the above, this study focused on sequencing the genome of *T. lanuginosus* SSBP, which produces the highest amount of a thermostable xylanase reported in literature. A further aim was to analyse proteins involved in carbohydrate-processing, which can have application for a wide range of industrial processes. In addition, the genome sequence of a thermophilic organism can provide insight to the mechanisms used for survival at elevated temperatures. Thus with these two main objectives in mind, *T. lanuginosus* was sequenced and annotated.

The quality of a genome assembly is affected by a variety of factors including the relative difficulty of obtaining DNA from various types of cells, availability of data analysis tools, and the implications of sequence errors and gaps downstream. Depth coverage, amount of reads assembled and continuity are the most used characteristics for identifying a good quality assembly. The genome assembly for the *T. lanuginosus* has good sequence coverage with the depth of 48-fold for the reads obtained in 454 sequencing and 137-fold with Illumina due to the high number of reads (98%) that were used for genome assembly. The depth of coverage is a measure of the number of times that a specific genomic site is sequenced. There are published genomes that have lower coverage including *A. oryzae* (9-fold), *A. nidulans* (13-fold), *N. crassa* (20-fold), *Mycosphaerella graminicola* (32-fold) and *T. reesei* (9-fold) which may have

sacrificed depth for breadth coverage (Galagan *et al.*, 2005a; Galagan *et al.*, 2003; Martinez *et al.*, 2008; Stukenbrock *et al.*, 2010).

The trade-off between breadth and depth coverage of genomes has been a challenge because the choice between spreading available data over a larger extent of targeted DNA or obtaining deeper coverage arises in several cases. However, this compromise for the *T. lanuginosus* genome was avioded as a depth of 48-fold was achieved whilst ensuring there were only minimum gaps. This is evident with the amount of scaffolds created in this genome and the high continuity of the scaffolds with an N50 value of approximately 4 Mb. The 6 larger scaffolds and the 24 smaller ones made up the genome which when considering that this was a *de novo* assembly was highly accurate. Genomes that have reference assemblies, means that both accuracy and coverage can be improved by using these genomes from other species but for de novo assemblies it is difficult.

The small size of the assembled genome for *T. lanuginosus* was unexpected as most filamentous fungal genomes range from between 30 Mb to 40 Mb on average. However, this did not seem to place this organism at a disadvantage as metabolic reconstruction showed that this organism contained the necessary tools to survive in its natural composting environment. Although the reasons for this huge difference is unclear, it could that a small genome is easier to maintain at high temperatures than a larger one, as more genome sequences become available, they may provide further information for this characteristic. Due to the size of the genome, the proteome of *T. lanuginosus*, with 5105 predicted genes, is very small with gene duplication and transposition events not evident. Again the *T. lanuginosus* proteome was much smaller than other filamentous fungi with *T. reseei* having the second smallest proteome of more than 9000 genes but still nearly twice the size of *T. lanuginosus*. Even yeasts like *S. cerevisiae*, *K. lactis* and *K. thermotolerans* had slightly larger proteomes.

The proteome was further analyzed for the ability of this organism to survive in its habitat. In order for organisms to grow and survive, energy-generating molecules are required and most organisms have high preference for carbohydrate-based molecules for energy production. Thus, the presence of carbohydrate-active enzymes (CAZy) was analyzed in *T. lanuginosus* and

revealed that it comprises the lowest amount of CAZy proteins (224) compared to other filamentous fungi (average 400), but higher than most yeasts (below 200). Glycoside hydrolases were significantly lower in *T. lanuginosus* with 94 genes, whilst other fungi had more than 200 GHs. The low number of GH proteins was attributed to the small size of the genome as discussed previously. Also, another factor responsible is the duplication of genes in other organisms, e.g., most *Aspergillu*s species have high number of ortholougs for GH proteins. However, when the total percentage of the CAZy proteins versus the total number of predicted protein was done, *T. lanuginosus* contained 4.38% CAZy proteins, second to only *A. fumigatus* which had the highest percentage of 4.66. This demonstrates that *T. lanuginosus* has an efficient system for carbohydrate utilisation where the balance between the amount of required genes and redundant genes being kept to a minimum as energy expenditure is crucial for thermophilic organisms.

In depth analysis of the carbohydrate utilization profile included the uptake of carbohydrate substrate common in composting plant waste, which contains compounds like cellulose, hemicellulose (xylans and galactomannans), starch and other compounds. The ability of this organism to grow on xylan polymers and xylose is well documented (Bennett *et al.*, 1998; Chen *et al.*, 2012a; Kuttanpillai *et al.*, 2009a; Puchart *et al.*, 1999; Purkarthofer and Steiner, 1995; Singh *et al.*, 2003). *T. lanuginosus* produces a low molecular weight xylanase which is responsible for cleaving the backbone of xylan which contain continuous xylose monomers linked by β-1,4-glycosidic bond producing a mixture of xylo-oligosaccharides and xylose (Dodd *et al.*, 2011; Hespelle and Whithead, 1990; Hrmova *et al.*, 1991; Lafond *et al.*, 2011; Maalej-Achouri *et al.*, 2012). Genome analysis confirmed the presence of this gene but only a single xylanase was identified. This was surprising as this strain produces copious amounts of this enzyme which is extracellular. This suggests that although copy numbers are important for high expression, they may not be the sole contributing factor in efficient degradation. Thus it maybe that gene regulation and transportation enables *T. lanuginosus* to achieve high production even with a single copy. This corroborated with the initial DNA microarray data (chapter 3) where one of the predicted genes was highly expressed (70-fold) when *T. lanuginosus* was grown on xylose compared to glucose while the xylanase itself resulted in just a 6-fold increase. This gene was suggested to be a xylose permease as its sequence showed some similarity with other sugar transporters.

Another enzyme involved in xylan degradation is xylosidase. This enzyme is responsible for converting xylo-oligosaccharide to the monomeric sugar, xylose, and usually enzyme assays show activity of less than a unit with this enzyme when tested in SSBP and other *Thermomyces* strains (Amaro-Reyes *et al.*, 2010; Banerjee *et al.*, 2010; Chen *et al.*, 2012a; Singh *et al.*, 2003; Singh *et al.*, 2000b) . Therefore further investigation of the expression of this enzyme is required. In the genome, two xylosidase genes (β and α) were identified in the fungal genome. The β-xylosidase was isolated and cloned giving proof of the enzyme and why this organism is well suited to its natural habitat where xylan comprises more than 30% of plant waste only second to cellulose (Bokhari *et al.*, 2010; Chen *et al.*, 2012b).

T. lanuginosus strains have been reported as cellulase-free organisms. Therefore, it was unexpected to identify not one but eight genes that might be related to cellulose degradation. Phylogenetic analysis revealed that three genes are very similar to cellulases from *T. reesei* which may be responsible for hydrolysing cellulose to glucose or cellobiose. This was partially supported by the phenotypic and DNA microarray data which showed that this fungus grows well on cellobiose as a sole carbon source. DNA microarray data alos revealed one of the glucanes being induced in xylose. This discovery may have a huge impact on the production of biofuels as conversion of cellulose to glucose is still the limiting factor due to enzyme cost and stability. Currently, the cellulases that are generally used are from a mesophilic fungus, *T. reesei*, hence a thermostable cellulase may of a better alternative for glucose production. The lack of evidence for activity of cellulase in *T. lanuginosus*, could be due to either very low activity or induction being difficult. However, cloning and characterisation to find out if the gene is functional or can be expressed is required.

T. lanuginosus genome encodes more than 50 ubiquinone/menaquinone methyltransferase genes which have been identified as being responsible for production of quinones. Menaquinone/ubiquinone (MK/ coenzyme Q10) plays an essential role in the electron transport chain and participates in aerobic cellular respiration, generating energy in the form of ATP. Though fungi are known to produce only ubiquinones, most of the methyltransferase genes found in *T. lanuginosus* were closely related to menaquinone genes which are generally

predominant in extremophilic prokaryotes. The high number of these genes may indicate that this system is used to produce surplus energy during optimum conditions, especially in the ever-changing conditions in a composting environment (Tiquia *et al*, 2002). Studies done on a *Bacillus subtilis* mutant which was unable to synthesis menaquinone, exhibited reduced O_2 uptake, together with a decreased level of cytochrome a, b and c and a decrease in sporulation (Farrand *et al*., 1973).

A most intriguing question is how some microorganisms thrive in conditions that in generally would not support life. The ability of *T. lanuginosus* to thrive in high temperatures (60°C, the upper limit for eukaryotic organism) is well-known, but the exact mechanisms are not well elucidated (Maheshwari *et al.*, 2000; Morgenstern *et al.*, 2012; Salar and Aneja, 2007). To all life forms, the integrity of genetic material is important for their survival, but the very nature of high temperature environments normally dictates the opposite. Thermophilic organisms use different strategies to protect DNA and one of the well-documented methods is higher GC content compared to mesophilic microorganisms (Kagawa *et al.*, 1984). *T. lanuginosus* GC content was calculated at 52.4% which was remarkable as also other fungi contain similar amounts, this being the upper limit for these organisms. Also, the GC content for the coding region was higher at above 56%, which suggested that the organism places emphasis on maintaining the integrity of important regions in the genome.

Also, the other reason that the GC content in *T. lanuginosus* is not significantly higher than other mesophilic fungi is that it may be using several strategies to thrive at high temperatures. Metabolic reconstruction revealed that *T. lanuginosus* is able to perform histone acetylation/deacetylation and methylation to assist in DNA condensation which is another strategy of stabilizing DNA (Nowak and Corces, 2004). Although some mesophilic fungi may be able to perform these processes, the presence of numerous methyltransferases makes *T. lanuginosus* unique and may imply that these genes may play an important role in thermal stability (Gostincar *et al.*, 2012).

One of the most vulnerable cell components in high temperature environments is the cell membrane as it contain lipids which may liquify at elevated temperatures. An increase in

temperature has been correlated to an increase in the proportion of saturated fatty acids incorporated into phospholipids, whereas at lower temperatures, a higher proportion of unsaturated fatty acids are incorporated. The metabolic reconstruction of *T. lanuginosus* confirmed that, this fungus can synthesis a variety of fatty acids of both saturated and unsaturated type. However, a study by Mehashwari *et al*. (2000) suggested that when this organism was grown at high temperature it produced more unsaturated fatty acid and may incorporate these to the cell membrane, but gene expression studies are required to support these observations.

One the most important aspects for surviving at high temperatures is the ability to prevent protein denaturation. Proteins are responsible for almost the entire cell metabolism and if protein structure and activity cannot be maintained then the organism will eventually die. Thus mechanisms for maintaining protein stability are absolutely essential for survival. *T. lanuginosus* appears to have several ways of achieving this and one of the less-known mechanisms is by methylation. Generally, in *T. lanuginosus*, gene duplication and ortholog presence is low, but with SAM-dependent methyltransferases, *T. lanuginosus* contains the highest number of these genes among fungi used in this study for comparison. Modifications performed by these proteins play a major role in methylation of biological molecules, including DNA, RNA, histones and other proteins. A study done on an unmodified protein purified from *E. coli* showed an increase in denaturation and aggregation events compared to the native version containing five lysine methylations (Febbraio *et al.*, 2004). This supports the hypothesis that amino acid methylation which later are incorporated into peptides improves protein stability in extremophiles.

Another mechanism for thermostability is the synthesis and accumulation of protective organic compounds, like sucrose, glycogen, and trehalose. The disaccharide trehalose, which accumulates dramatically during heat shock and stationary phase in many organisms, enhances thermotolerance and reduces aggregation of denatured proteins and protecting cells against oxygen radicals (Elbein *et al.*, 2003; Ferreira *et al.*, 2007). High metabolism of trehalose in *T. lanuginosus* was observed during phenotypic microarray analysis which may suggest that this compound is used in stabilising mechanisms. Also metabolic reconstruction showed that *T. lanuginosus* is able to synthesise trehalose. Ferreira et al (2007) demonstrated that *Pisolithus* sp.

accumulated trehalose and an increased trehalase activity in the mycelial mass of the fungus when exposed to elevated temperatures. *A. nidulans* accumulated trehalose to acquire tolerance to a variety of stresses in this filamentous fungus (Fillinger *et al.*, 2001a).

Temperature is an important factor that affects growth, reproduction and survivability of organisms. Thus, organisms are subjected to thermal stress due to change in temperatures that occur in nature. Another method to control thermal stress is by producing heat shock proteins (Hsps). Heat shock proteins generally function as intracellular chaperones for other proteins. They play an essential role in protein interactions such as folding and assisting in the establishment of proper protein conformation (shape) and in the prevention of unwanted protein aggregation. *T. lanuginosus* showed enhanced survival at higher temperature when previously exposed to heat shock treatment (Trent *et al.*, 1994). The genome of *T. lanuginosus* contains 19 genes coding for heat shock proteins, which was generally higher than most fungi. Also among the Hsp proteins identified in *T. lanuginosus* is the Hsp70 protein, which is involved in proper protein folding after synthesis. Hsp70 binds tightly to partially-synthesized peptide sequences (incomplete proteins) thus preventing them from aggregating and being rendered non-functional. Once the protein is completely synthesized Hsp70 will detach and the protein is then free to fold on its own, or to be transferred to other chaperones for further processing. Freitas et al. (2011) also observed that this Hsp accumulated in *A. nidulans* after prolonged heating. The high presence of these genes in *T. lanuginosus* in comparison to the its mesophilic counterparts gives clues nn how organisms maintain stability in stressful conditions.

The ubiquitin degradation system plays an important role in the degradation of denatured proteins following environmental stress such as nutrient limitation, heat shock and heavy metal exposure (Staszczak, 2008). Metabolic reconstruction indicated that the ubiquitin degradation pathway is functional in *T. lanuginosus*. Studies suggest that the degradation of denatured proteins in thermophilic fungi is faster compared to that of mesophilic fungi, thus increasing turnover rate of protein (Maheshwari *et al.*, 2000). Since *T. lanuginosus* is a thermophilic organism and grows on dead, woody material, this system may be essential to adaptation during rising temperature in composting materials.

The phenotypic and DNA microarray was used assess the metabolism of different carbon sources by *T. lanuginosus*. This organism grows better on xylose than glucose which is relatively rare for most filamentous fungi. Also, the metabolism of trehalose was similar to that of xylose owing to the importance of this molecule. Metabolic reconstruction confirmed that nearly 87 pathways are functional in *T. lanuginosus*. Although 119 gaps are found in these pathways more than 800 genes were poorly classified and more than 500 did not fit into any COG clusters. Thus, these gaps are probably due to limitations of genome annotation, currently. However, as more information becomes available these gaps can be easily closed due to these genes being assigned functions through characterization of similar sequences.

4.2 FUTURE PROSPECTS

Understanding the genetic processes behind adaptive phenotypic responses and adaptations of organisms is important in the application of these organisms for human and industrial processes. Although genome sequencing provides a huge amount of data and insight, further elucidation is still required. Similarly with *T. lanuginosus*, despite sequencing of the genome, more analysis is still needed e.g., continuing presence of active cellulase genes and the presence of a large group of methyltransferases. Further gene annotation, cloning and characterization may yet reveal further "secrets" locked within the genome data.

In order to obtain a more completely annotated genome and unravel more information about this fungus, the following research foci are recommended for further study:

(i) complete transcriptome and mitochondrial genome analysis needs to be done, including using cellulose and cellobiose as substrate for these studies;
(ii) suitable gap closing software can be used to reduce the number of scaffolds;
(iii) microarray studies can be used to study some of the pathways that are involved in thermostability and hydrolysis of carbohydrate compounds;
(iv) isolation and cloning of interesting genes like cellulase and transporter genes for heterologous expression can be undertaken in order to characterize these genes and their over-expression for industrial applications; and

(v) functional annotation of proteins current designate as unknown proteins.

The sequencing of the transcriptome and functional annotation of poorly characterized genes may lead to the indentification of biocatalysts with new applications. A lthough, currently very few tools are available to achieve this, it is with hope that rapid development of genomics may be new tools will be available in the future to be able to perform more complete annotations. Also as more thermophilic genomes from eukaryotic organisms become available, comparison studies can be done in order to acquire more information about these organisms. In conclusion, this work has paved the way for a better understanding of thermophilicity in eukaryotic organisms and their survival at elevated temperatures.

5. REFERENCES

Ariadne Genomics [Online]. Available: http://www.ariadnegenomics.com/products/pathway-studio/expression-analysis/algorithms/.

FUNGIpath [Online]. Available: http://embg.igmors.u-psud.fr/fungipath/.

Psort II protein analysis server.

ABAD, P., GOUZY, J., AURY, J.-M., CASTAGNONE-SERENO, P., DANCHIN, E. G. J., DELEURY, E., PERFUS-BARBEOCH, L., ANTHOUARD, V., ARTIGUENAVE, F., BLOK, V. C., CAILLAUD, M.-C., COUTINHO, P. M., DASILVA, C., DE LUCA, F., DEAU, F., ESQUIBET, M., FLUTRE, T., GOLDSTONE, J. V., HAMAMOUCH, N., HEWEZI, T., JAILLON, O., JUBIN, C., LEONETTI, P., MAGLIANO, M., MAIER, T. R., MARKOV, G. V., MCVEIGH, P., PESOLE, G., POULAIN, J., ROBINSON-RECHAVI, M., SALLET, E., SEGURENS, B., STEINBACH, D., TYTGAT, T., UGARTE, E., VAN GHELDER, C., VERONICO, P., BAUM, T. J., BLAXTER, M., BLEVE-ZACHEO, T., DAVIS, E. L., EWBANK, J. J., FAVERY, B., GRENIER, E., HENRISSAT, B., JONES, J. T., LAUDET, V., MAULE, A. G., QUESNEVILLE, H., ROSSO, M.-N., SCHIEX, T., SMANT, G., WEISSENBACH, J. & WINCKER, P. (2008) Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. Nat Biotech, 26**,** 909-915.

ABBOTT, J. C., AANENSEN, D. M., RUTHERFORD, K., BUTCHER, S. & SPRATT, B. G. (2005) WebACT—an online companion for the Artemis Comparison Tool. Bioinformatics, 21**,** 3665-3666.

ABNIZOVA, I., LEONARD, S., SKELLY, T., BROWN, A., JACKSON, D., GOURTOVAIA, M., QI, G., TE, R., FARUQUE, N., LEWIS, K. & COX, T. (2012) Analysis of context-dependent errors for illumina sequencing. Journal of Bioinformatics and Computational Biology, 10**,** 20.

AGRIOS, G. N. 2009. Plant Pathogens and Disease: General Introduction. *In:* EDITOR-IN-CHIEF: MOSELIO, S. (ed.) *Encyclopedia of Microbiology (Third Edition).* Oxford: Academic Press.

AL-REEDY, R. M., MALIREDDY, R., DILLMAN, C. B. & KENNELL, J. C. (2012) Comparative analysis of *Fusarium* mitochondrial genomes reveals a highly variable region that encodes an exceptionally large open reading frame. Fungal genetics and biology : FG & B, 49**,** 2.

ALLWOOD, J. W., DE VOS, R. C. H., MOING, A., DEBORDE, C., ERBAN, A., KOPKA, J., GOODACRE, R. & HALL, R. D. 2011. Chapter sixteen - Plant Metabolomics and Its Potential for Systems Biology Research: Background Concepts, Technology, and Methodology. *In:* DANIEL JAMESON, M. V. & HANS, V. W. (eds.) *Methods in Enzymology.* Academic Press.

ALSCHER, R. G., ERTURK, N. & HEATH, L. S. (2002) Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. Journal of Experimental Botany, 53**,** 1331-1341.

AMARO-REYES, A., GARCIA-ALMENDAREZ, B. E., VAZQUEZ-MANDUJANO, D. G., AMAYA-LLANO, S., CASTANO-TOSTADO, E., GUEVARA-GONZALEZ, R. G., LOERA, O. & REGALADO, C. (2010) Homologue expression of a β-xylosidase from native *Aspergillus niger*. Journal of Industrial Microbiology and Biotechnology, 38**,** 1311.

ANDERSEN, M. R., SALAZAR, M. P., SCHAAP, P. J., VAN DE VONDERVOORT, P., CULLEY, D., THYKAER, J., .FRISVAD, C. J., NIELSEN, K. F., ALBANG, R., ALBERMANN, K., BERKA, R. M., BRAUS, G., H., STROMEYER, S., CORROCHANO, L. M., DAI, Z., VAN DIJCK GERALD.HOFMANN, P., LASURE, L. L., MAGNUSON, J. K., MENKE, K., MEIJER, M., MEIJER, L. M., NIELSEN, J. B., NIELSEN, M. M., VAN OOYEN, A. J. J., PEL, H. J., POULSEN, L., SAMSON, A., STAM, H., TSANG, A., VAN DEN BRINK, J. M., ATKINS, A., AERTS, A., SHAPIRO, A., PANGILINAN, H., SALAMOV, A., LOU, Y., LINDQUIST, E., GRIMWOOD, J., GRIGORIEV, I. V., KUBICEK, C. P., MARTINEZ, D., VAN PEIJ, N.

N., ROUBOS, J. A., NIELSEN, J. & BAKER, S. (2011) Comparative genomics of citric-acid-producing *Aspergillus* ATCC 1015 versus enzyme-producing CBS 513.88. Genome Research, 21**,** 885-897.

ARANTES, V., MILAGRES, A. M. F., FILLEY, T. R. & GOODELL, B. (2011) Lignocellulosic polysaccharides and lignin degradation by wood decay fungi: the relevance of nonenzymatic Fenton-based reactions. Journal of Industrial Microbiology and Biotechnology 38**,** 541-555.

ARCHER, D., JEENES, D. & MACKENZIE, D. (1994) Strategies for improving heterologous protein production from filamentous fungi. Antonie van Leeuwenhoek, 65**,** 245-250.

ARCHER, D. B. (2000) Filamentous fungi as microbial cell factories for food use. . Current Opinion in Biotechnology, 11**,** 478–483.

ARGÜELLES, J. C. (2000) Physiological roles of trehalose in bacteria and yeasts: a comparative analysis. Archives of Microbiology, 174**,** 217-224.

AVRAMOFF, A., KHAN, W., EZRA, A., ELGART, A., HOFFMAN, A. & DOMB, A. J. (2012) Cyclosporin pro-dispersion liposphere formulation. Journal of Controlled Release, 160**,** 401-406.

BAKER, M. (2012) De novo genome assembly: what every biologist should know. Nat Meth, 9**,** 333-337.

BAKER, S. E., SCHADT, C. W., TOROK, T., TUSKAN, G., BENNETT, J., BERKA, R. M., BRIGGS, S. P., HEITMAN, J., TAYLOR, J., GILLIAN, T. B., WERNER-WASHBURNE, M., THYKAER, J., HIMMEL, M. E., ADNEY, W. S., BRETTIN, T. S., BROCKMAN, F. J., D'HAESELEER, P., MARTINEZ, A. D., MILLER, R. M. & ROKHSAR, D. S. (2008) Fungal genome sequencing and bioenergy. Fungal Biology Reviews, 22**,** 1-5.

BANERJEE, G., SCOTT-CRAIG, J. S. & WALTON, J. D. (2010) Improving Enzymes for Biomass Conversion: A Basic Research Perspective. BioEnergy Research, 3**,** 82-92.

BANSZKY, L., SIMONICS, T. & MARAZ, A. (2003) Sulphate metabolism of selenate-resistant Schizosaccharomyces pombe mutants. J Gen Appl Microbiol, 49**,** 271 - 278.

BAUER, I., GRAESSLE, S., LOIDL, P., HOHENSTEIN, K. & BROSCH, G. (2010) Novel insights into the functional role of three protein arginine methyltransferases in Aspergillus nidulans. Fungal Genetics and Biology, 47**,** 551-561.

BAUMANN, K., DATO, L., GRAF, A., B, FRASCOTTI, G., DRAGOSITS, M., PORRO, D., MATTANOVICH, D., FERRER, P. & BRANDUARDI, P. (2011) The impact of oxygen on the transcriptome of recombinant *S. cerevisiae* and *P. pastoris* - a comparative analysis. BMC Genomics, 12**,** 1-16.

BAXEVANIS, A. D. 2002. An Overview of Gene Identification: Approaches, Strategies, and Considerations. *Current Protocols in Bioinformatics.* John Wiley & Sons, Inc.

BCC 2012. Global Markets for Enzymes in Industrial Applications.

BEAUCHEMIN, K. A., COLOMBATTO, D., MORGAVI, D. & YANG, W. Z. (1996) Use of Exogenous Fibrolytic Enzymes to Improve Feed Utilization by Ruminants. Journal of Animal Science, 81**,** E37–E47.

BECHEM, E. E. T. (2012) Utilisation of organic and inorganic nitrogen sources by S*cleroderma sinnamariense* Mont. isolated from *Gnetum africanum* Welw. African Journal of Biotechnology Vol. 11**,** 9205-9213.

BEINERT, H., HOLM, R. H. & MÜNCK, E. (1997) Iron-Sulfur Clusters: Nature's Modular, Multipurpose Structures. Science, 277**,** 653-659.

BELENKY, P., BOGAN, K. & BRENNER, C. (2007) NAD+ metabolism in health and disease. Trends Biochemistry Science 32**,** 9-12.

BELLIS, M., PAGÈS, M. & ROIZÈS, G. (1987) A simple and ranid method for preparing yeast chromosomes for Pulse Field Gel Electrophoresis. Nucleic acids research, 15**,** 6749-6749.

BENNETT, N. B., RYAN, J., BEILY, P., VRASANKA, M., KREMINICKY, L., MACRIS, B. J., KEKOS, D., CHRISTAKOPOULOS, P., KATAPODIS, P., CLAEYSSENS, M., NERICKX, W., NTAUMA, P. & BHAT, M. K. (1998) Biochemical and catalytic properties of an endoxylanase purified from culture filtrate of Thermomyces lanuginosus ATCC 46882. Carbohydrate Research, 306**,** 445-455.

BERGER, F., M.H., R.-H. & ZIEGLER, M. (2004) The new life of a centenarian: signalling functions of NAD(P)". Trends Biochemistry Science 29**,** 111–118.

BERKA, R. M., GRIGORIEV IV, OTILLAR R, SALAMOV A, GRIMWOOD J, REID I, ISHMAEL N, JOHN T, DARMOND C, MOISAN M. C, HENRISSAT B, COUTINHO P. M, LOMBARD V, NATVIG DO, LINDQUIST E, SCHMUTZ J, LUCAS S, HARRIS P, POWLOWSKI J, BELLEMARE A, TAYLOR D, BUTLER G, DE VRIES R. P, ALLIJN I. E, VAN DEN BRINK J, USHINSKY S, STORMS R, POWELL A. J, PAULSEN I. T, ELBOURNE L. D, BAKER S. E, MAGNUSON J, LABOISSIERE S, CLUTTERBUCK A. J, MARTINEZ D, WOGULIS M, DE LEON A. L, REY M. W & A., T. (2011) Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. nature biotechnology, 29**,** 922-927.

BERNHARD, S. & BENJAMIN, M. (2011) Fungal arabinan and L-arabinose metabolism. Applied Microbiology and Biotechnology, 89**,** 1665-1673.

BINO, R. J., HALL, R. D., FIEHN, O., KOPKA, J., SAITO, K., DRAPER, J., NIKOLAU, B. J., MENDES, P., ROESSNER-TUNALI, U. & BEALE, M. H. (2004) Potential of metabolomics as a functional genomics tool. Trends in plant science, 9**,** 418-425.

BISSOON, S., CHRISTOV, L. & SINGH, S. (2002) Bleach boosting effects of purified xylanase from Thermomyces lanuginosus SSBP on bagasse pulp. Process Biochemistry, 37**,** 567-572.

BLÄTTEL, V., LARISIKA, M., PFEIFFER, P., NOWAK, C., EICH, A., ECKELT, J. & KÖNIG, H. (2011) Vinification β-1,3-Glucanase from *Delftia tsuruhatensis* Strain MV01 and Its Potential Application in Vinification. Applied and Environmantal Microbiology, 77**,** 983-990.

BOCHNER, B. R., GADZINSKI, P. & PANOMITROS, E. (2001) Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome research, 11**,** 1246-1255.

BOKHARI, S. A. I., LATIF, F., AKHTAR, M. W. & RAJOKA, M. I. (2010) Characterization of a β-xylosidase produced by a mutant derivative of Humicola lanuginosa in solid state fermentation. Annals of microbiology, 60**,** 21-29.

BONETTA, L. (2010) Whole-Genome Sequencing Breaks the Cost Barrier. Cell, 141**,** 917-919.

BONNET, A., LAGARRIGUE, S., LIAUBET, L., ROBERT-GRANIE, C., SANCRISTOBAL, M. & TOSSER-KLOPP, G. (2009) Pathway results from the chicken data set using GOTM, Pathway Studio and Ingenuity softwares. BMC Proceedings, 3**,** S11.

BORGLIN, S., JOYNER, D., DEANGELIS, K. M., KHUDYAKOV, J., D'HAESELEER, P., JOACHIMIAK, M. P. & HAZEN, T. (2012) Application of phenotypic microarrays to environmental microbiology. Current Opinion in Biotechnology, 23**,** 41-48.

BP 2012. BP Statistical Review of World Energy

BRAAKSMA, A. & SCHAAP, D. J. (1996) Protein analysis of the common mushroom Agaricus bisporus. Postharvest Biology and Technology, 7**,** 119-127.

BRANDELLI, A., GEIMBA, M. P. & MANTOVANI, C. F. (2005) Enzymatic clarification of fruit juices by fungal pectin lyase. . Food Biotechnology 19**,** 173-181.

BROOIJMANS, R., SMIT, B., SANTOS, F., VAN RIEL, J., DE VOS, W. M. & HUGENHOLTZ, J. (2009) Heme and menaquinone induced electron transport in lactic acid bacteria. Microbial Cell Factories, 8.

BYUNG, H. P., TATIANA, V. K., MUSTAFA, H. S., MICHAEL, R. L. & EDWARD, C. U. (2010) CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology, 20**,** 1574–1584.

CANTAREL, B., KORF, I., ROBB, S., PARRA, S., ROSS, E., MOORE, B., HOLT, C., ALVARADO, A. & YANDELL, M. (2011) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research, 18**,** 188-196.

CANTAREL, B. L., COUTINHO, P. M., RANCUREL, C., BERNARD, T., V., L. & HENRISSAT, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Research 37**,** D233-D238.

CARVALHO, N. D. S. P., ARENTSHORST, M., WEENINK, X. O., PUNT, P. J., VAN DEN HONDEL, C. A. M. J. J. & RAM, A. F. J. (2011) Functional YFP-tagging of the essential GDP-mannose transporter reveals an important role for the secretion related small GTPase SrgC protein in maintenance of Golgi bodies in *Aspergillus niger*. Fungal Biology, 115**,** 253-264.

CARVER, T., BERRIMAN, M., TIVEY, A., PATEL, C., BÖHME, U., BARRELL, B. G., PARKHILL, J. & RAJANDREAM, M.-A. (2008) Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics, 24**,** 2672-2676.

CELESTINO, K. R. S., CUNHA, R. B. & FELIX, C. R. (2006) Characterization of a β-glucanase produced by *Rhizopus microsporus var. microsporus*, and its potential for application in the brewing industry. BMC Biochemistry, 7.

CHAMBERGO, F. S., BONACCORSI, E. D., FERREIRA, A. J., RAMOS, A. S., FERREIRA, J. R., ABRAHAO-NETO, J., FARAH, J. P. S. & EL-DORRY, H. (2002) Elucidation of the metabolic fate of glucose in the filamentous fungus Trichoderma reesei using expressed sequence tag (EST) analysis and cDNA microarrays. Journal of Biological Chemistry, 277**,** 13983-13988.

CHEN, X., JIANG, Z.-H., CHEN, S. & QIN, W. (2010) Microbial and Bioconversion Production of D-xylitol and Its Detection and Application. International Journal of Biological Sciences, 6**,** 834-844.

CHEN, Z., JIA, H., YANG, Y., YAN, Q., JIANG, Z. & TENG, C. (2012a) Secretory expression of a β-xylosidase gene from *Thermomyces lanuginosus* in *Escherichia coli* and

characterization of its recombinant enzyme. Letters in Applied Microbiology, 55**,** 330-337.

CHEN, Z., JIA, H., YANG, Y., YAN, Q., JIANG, Z. & TENG, C. (2012b) Secretory expression of a β-xylosidase gene from Thermomyces lanuginosus in Escherichia coli and characterization of its recombinant enzyme. Letters in applied microbiology, 55**,** 330-337.

CHIANG, Y.-M., LEE, K.-H., SANCHEZ, J. F., KELLER, N. P. & WANGA, C. C. C. (2009) Unlocking Fungal Cryptic Natural Products. Natural Product Communications, 4**,** 1505-1510.

CHRISTOPHER, L., BISSOON, S., SINGH, S., SZENDEFY, J. & SZAKACS, G. (2005) Bleach-enhancing abilities of Thermomyces lanuginosus xylanases produced by solid state fermentation. Process Biochemistry, 40**,** 3230-3235.

CLEMONS, P. A. (2004) Complex phenotypic assays in high-throughput screening. Current opinion in chemical biology, 8**,** 334-338.

COHEN, N., ABRAMOV, S., DROR, Y. & FREEMAN, A. (2001) In vitro enzyme evolution: the screening challenge of isolating the one in a million. Trends in Biotechnology, 19**,** 507-510.

CONESA, A., GÖTZ, S., GARCÍA-GÓMEZ, J. M., TEROL, J., TALÓN, M. & ROBLES, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 21**,** 3674-3676.

COOK, G. M., JANSSEN, P. H. & MORGAN, H. W. (1993) Simultaneous uptake and utilisation of glucose and xylose by< i> Clostridium thermohydrosulfuricum</i>. FEMS microbiology letters, 109**,** 55-61.

CORDA, D. & DI GIROLAMO, M. (2003) "NEW EMBO MEMBER'S REVIEW: Functional aspects of protein mono-ADP-ribosylation". . EMBO Journal 22**,** 1953-1958.

COWAN, D. A. (2004) The upper temperature for life – where do we draw the line? Trends in Microbiology, 12**,** 58-60.

CRISTIANO, C. N., MALALI, G., JOSHUA, S., MINFENG, X., FENG, C., DOUGLAS, E. B., YEONYEE, O., THOMAS, K. M. & A.DEAN., R. (2011) Diverse and tissue-enriched small RNAs in the plant pathogenic fungus, *Magnaporthe oryzae*. BMC Genomics, 12**,** 288.

CRUZ, C., EGSGAARD, H., TRUJILLO, C., AMBUS, P., REQUENA, N., MARTINS-LOUÇÃO, M. A. & JAKOBSEN, I. (2007) Enzymatic Evidence for the Key Role of Arginine in Nitrogen Translocation by Arbuscular Mycorrhizal Fungi. Plant Physiology, 144**,** 782-792.

DANIELS, N. M., GALLANT, A., PENG, J., COWEN, L. J., BAYM, M. & BERGER, B. (2013) Compressive genomics for protein databases. Bioinformatics, 29**,** i283-i290.

DE OLIVEIRA , J., MIGUEL.P.FERREIRA DE OLIVEIRA & DE GRAAFF, L., H. DE GRAAFF (2011) Proteomics of industrial fungi: trends and insights for biotechnology. Applied microbiology and Biotechnology, 89**,** 225-237.

DE OLIVEIRA, J. M. P. F. & DE GRAAFF, L. H. (2011) Proteomics of industrial fungi: trends and insights for biotechnology. Applied microbiology and biotechnology, 89**,** 225-237.

DEAN, R. A., TALBOT, N. J., EBBOLE, D. J., FARMAN, M. L., MITCHELL, T. K., ORBACH, M. J., THON, M., KULKARNI, R., XU, J.-R., PAN, H., READ, N. D., LEE, Y.-H., CARBONE, I., BROWN, D., OH, Y. Y., DONOFRIO, N., JEONG, J. S., SOANES, D. M., DJONOVIC, S., KOLOMIETS, E., REHMEYER, C., LI, W., HARDING, M., KIM, S., LEBRUN, M.-H., BOHNERT, H., COUGHLAN, S., BUTLER, J., CALVO, S., MA, L.-J., NICOL, R., PURCELL, S., NUSBAUM, C., GALAGAN, J. E. & BIRREN, B. W. (2005) The genome sequence of the rice blast fungus Magnaporthe grisea. Nature, 434**,** 980-986.

DESCORPS-DECLÈRE, S., LEMOINE, F., SCULO, Q., LESPINET, O. & LABEDAN, B. (2008) The multiple facets of homology and their use in comparative genomics to study the evolution of genes, genomes, and species. Biochimie, 90**,** 595-608.

DIEFENBACH, J. & KINDL, H. (2000) The membrane-bound DnaJ protein located at the cytosolic site of glyoxysomes specifically binds the cytosolic isoform 1 of Hsp70 but not other Hsp70 species. European Journal of Biochemistry, 267**,** 746-754.

DIGUISTINI, S., LIAO, N. Y., PLATT, D., ROBERTSON, G., SEIDEL, M., CHAN, S. K., DOCKING, T. R., BIROL, I., HOLT, R. A., HIRST, M., MARDIS, E., MARRA, M. A., HAMELIN, R. C., BOHLMANN, J., BREUIL, C. & JONES, S. J. M. (2009) De novo genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. Genome Biology (Online Edition), 10**,** R94.

DIGUISTINI, S., WANG, Y., LIAO, Y., TAYLOR, G., TANGUAYC, P., FEAUD, N., HENRISSATE, B., CHAN, S. K., HESSE-ORCEA, U., ALAMOUTI, S., TSUIF, C. K. M., DOCKING, R. T., LEVASSEURG, A., HARIDASA, S., ROBERTSON, G., BIROL, I., HOLT, C., MARRA, M. A., HAMELIN, R. C., HIRST, M., JONES, S. K., BOHLMANN, J. R., BREUIL, C. & JONES, S., J.M. (2010) Genome and transcriptome analyses of the mountain pine beetle fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen. PNAS, 26**,** 1-6.

DODD, D. & CANN, I., K.O. (2009) Enzymatic deconstruction of xylan for biofuel production. Global Change Biology Bioenergy, 1**,** 2-17.

DODD, D., KIYONARI, S., MACKIE, R. & CANN, I. (2010) Functional Diversity of Four Glycoside Hydrolase Family 3 Enzymes from the Rumen Bacterium *Prevotella bryantii*. Journal of Bacteriology, 192**,** 2335–2345.

DODD, D., MACKIE, R., I & CANN, I., K.O. (2011) Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. Molecular Microbiology, 79**,** 292-304.

DOERKS, T., VAN NOORT, V., MINGUEZ, P. & BORK, P. (2012) Annotation of the <italic>M. tuberculosis</italic> Hypothetical Orfeome: Adding Functional Information to More than Half of the Uncharacterized Proteins. PLoS ONE, 7**,** e34302.

DORRELL, N., CHAMPION, O. L. & WREN, B. W. 2002. Application of DNA microarrays for comparative and evolutionary genomics. *In:* BRENDAN WREN, N. D. (ed.) *Methods in Microbiology.* Academic Press.

DORSETT, Y. & TUSCHL, T. (2004) siRNAs: applications in functional genomics and potential as therapeutics. Nature Reviews Drug Discovery, 3**,** 318-329.

DOULL, J. L. & VINING, L. C. (1990) Physiology of antibiotic production in actinomycetes and some underlying control mechanisms. Biotechnology Advances, 8**,** 141-158.

DRUZHININA, I., S. , SCHMOLL, M., SEIBOTH, B. & KUBICEK, C., P. (2005) Global Carbon Utilization Profiles of Wild-Type, Mutant, Transformant Strains of *Hypocrea jecorina*. Applied and Environmantal Microbiology, 72.

DUFRESNE, M., LESPINET, O., DABOUSSI, M. J. & HUA-VAN, A. (2011) Genome-wide comparative analysis of pogo-like transposable elements in different *Fusarium* species. Journal of molecular evolution, 73**,** 230-243.

EKMAN, D. & ELOFSSON, A. (2010) Identifying and Quantifying Orphan Protein Sequences in Fungi. Journal of Molecular Biology, 396**,** 396-405.

ELBEIN, A. D., PAN, Y. T., PASTUSZAK, I. & CARROLL, D. (2003) New insight on trehalose: a multifunctional molecule. Glycobiologogy, 13**,** 17-27.

FARRAND, S. K. & TABER, H. W. (1973) Physiological Effects of Menaquinone Deficiency in *Bacillus subtilis*. Journal of Bacteriology 1973, 115(3):1035-1044., 115**,** 1035-1044.

FEBBRAIO, F., ANDOLFO, A. & TANFANI , F. (2004) Thermal stability and aggregation of *Sulfolobus solfataricus* β-glycosidase are dependent upon the N-ε-methylation of specific lysyl residues: critical role of in vivo post-translational modifications. Journal of Biological Chemistry, 279**,** 10185–10194.

FERREIRA, A. S., TÓTOLA, M. R. & BORGES, A. C. (2007) Physiological implications of trehalose in the ectomycorrhizal fungus Pisolithus sp. under thermal stress. Journal of Thermal Biology, 32**,** 34-41.

FILLINGER, S., CHAVEROCH, M. K., DIJCK, P. V., VRIES, R., EUIJTER, G., THEVELEIN, J. & D'ENFERT, C. (2001a) Trehalose is required for the acquisition of tolerance to a variety of stresses in the filamentous fungus *Aspergillus nidulans*. Microbiology and Molecular Biology Reviews, 147**,** 1851–1862.

FILLINGER, S., CHAVEROCHE, M.-K., VAN DIJCK, P., DE VRIES, R., RUIJTER, G., THEVELEIN, J. & D'ENFERT, C. (2001b) Trehalose is required for the acquisition of tolerance to a variety of stresses in the filamentous fungus Aspergillus nidulans. Microbiology, 147**,** 1851-1862.

FINN, R. D., MISTRY, J., TATE, J., COGGILL, P., HEGER, A., POLLINGTON, J. E., GAVIN, O. L., GUNASEKARAN, P., CERIC, G., FORSLUND, K., HOLM, L., SONNHAMMER, E. L. L., EDDY, S. R. & BATEMAN, A. (2010) The Pfam protein families database. Nucleic Acids Research, 38**,** D211-D222.

FLEMING, K., KELLEY, L. A., ISLAM, S. A., MACCALLUM, R. M., MULLER, A., PAZOS, F. & STERNBERG, M. J. (2006) The proteome: structure, function and evolution. Philos Trans R Soc Lond B Biol Sci. , 31**,** 441-451.

FOX, R. T. V. (1997) The present and future use of technology to detect plant pathogens to guide disease control in sustainable farming systems. Agriculture, Ecosystems & Environment, 64**,** 125-132.

FRANCIS, M., ANNEGRET, K., CLAUDE, M., RAFFAELLA, B., PEDRO, M. C., OLIVIER, J., BARBARA, M., EMMANUELLE, M., BENJAMIN, N., RICCARDO, P., BETTINA, P., ANDREA, R., ANTONELLA, A., JOELLE, A., VE´RONIQUE, A., SERGIO, A., FRANCOIS, A., JEAN-MARC, A., PAOLA, B., ANGELO, B., ANDREA, B., ANNICK, B., MARC, B., BRANDI, C., GERARD, C., ARNAUD, C., CORINNE, D. S., FRANCE, D., SEBASTIEN, D., STEFANO, G., BENOıT, H., MIRCO, I., ., BENOıT, M., ANTONIETTA, M., MICHELE, M., GIOVANNI, P., HADI, Q., CLAUDIA, R., ROBERTA, R., RICHARD, S., VILBERTO, S., EMILIE, T., ARTURO, R. V., ALESSANDRA, Z., ELISA, Z., BERNARD, H., MARC-HENRI, L., FRANCESCO, P.,

PAOLA, B., SIMONE, O. & PATRICK, W. (2010) Perigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature, 464.

FREITAS, J. S., SILVA, E., JULIANA LEAL, E. M., GRAS, D. E., MARTINEZ-ROSSI, N. M., DOS SANTOS, L. D., PALMA, M. S. & ROSS, A. (2011) Transcription of the Hsp30, Hsp70, and Hsp90 heat shock protein genes is modulated by the PalA protein in response to acid pH-sensing in the fungus *Aspergillus nidulans*. Cell Stress and Chaperones 16**,** 565–571.

FRIDLYAND, J., SNIJDERS, A. M., PINKEL, D., ALBERTSON, D. G. & JAIN, A. N. (2004) Hidden Markov models approach to the analysis of array CGH data. Journal of Multivariate Analysis, 90**,** 132-153.

FULLWOOD, M., WEI, C., ET, L. & Y, R. (2009) Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. Genome Research, 19**,** 521-32.

GALAGAN, J., HENN, M., MA, L.-J., CUOMO, C. A. & BIRREN, B. (2005b) Genomics of the fungal kingdom: Insights into eukaryotic biology. Genome Research, 15**,** 1620-1631.

GALAGAN, J. E., CALVO, S. E., CUOMO, C., MA, L., WORTMAN, J. R., BATZOGLOU, S., LEE, S., BASTU¨RKMEN, M., SPEVAK, C. C., CLUTTERBUCK, J., KAPITONOV, V., JURKA, J., SCAZZOCCHIO, C., M., F., BUTLER, J., PURCELL, S., HARRIS, S., BRAUS, G. H., DRAHT, O., BUSCH, S., D'ENFERT, C., BOUCHIER, C., GOLDMAN, G. H., BELL-PEDERSEN, D., GRIFFITHS-JONES, S., DOONAN, J. H., YU, J., VIENKEN, K., PAIN, A., FREITAG, M., SELKER, E. U., ARCHER, D. B., PEN˜ALVA, M. A., OAKLEY, B. R., MOMANY, M., TANAKA, T., T., K., ASAI, K., MACHIDA, M., NIERMAN, W. C., DENNING, D. W., CADDICK, M., HYNES, M., PAOLETTI, M., FISCHER, R., MILLER, B., DYER, P., SACHS, M. S., OSMANI, S. A. & BIRREN, B. W. (2005a) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature, 438**,** 1105-1115.

GALAGAN, J. E., PURCELL, S., REHMAN, B., ELKINS, T., ENGELS, R., WANG, S., NIELSEN, C. B., BUTLER, J., ENDRIZZI, M., QUI, D., IANAKIEV, P., CALVO, S. E., BELL-PEDERSEN, D., NELSON, M. A., WERNER-WASHBURNE, M.,

SELITRENNIKOFF, C. P., KINSEY, J. A., BRAUN, E. L., ZELTER, A., SCHULTE, U., KOTHE, G. O., JEDD, G., BORKOVICH, K. A., MEWES, W., STABEN, C., MARCOTTE, E., GREENBERG, D., ROY, A., FOLEY, K., NAYLOR, J., STANGE-THOMANN, N., BARRETT, R., GNERRE, S., SELKER, E. U., KAMAL, M., KAMVYSSELIS, M., MAUCELI, E., BIELKE, C., RUDD, S., FRISHMAN, D., KRYSTOFOVA, S., RASMUSSEN, C., METZENBERG, R. L., PERKINS, D. D., READ, N. D., KROKEN, S., COGONI, C., MACINO, G., CATCHESIDE, D., LI, W., PRATT, R. J., OSMANI, S. A., DESOUZA, C. P. C., GLASS, L., ORBACH, M. J., JAFFE, D., BERGLUND, J. A., VOELKER, R., YARDEN, O., PLAMANN, M., SEILER, S., DUNLAP, J., RADFORD, A., ARAMAYO, R., NATVIG, D. O., ALEX, L. A., FITZHUGH, W., MANNHAUPT, G., EBBOLE, D. J., FREITAG, M., PAULSEN, I., SACHS, M. S., LANDER, E. S., NUSBAUM, C., BIRREN, B., MA, L.-J. & SMIRNOV, S. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. Nature, 422**,** 859-868.

GALANTE, Y., DE CONTI, A. & MONTEVERDI, R. (1998) Application of Trichoderma enzymes in the food and feed industries. Trichoderma and Gliocladium**,** 327 - 342.

GANCEDO, C. & FLORES, C. L. (2004) The importance of a functional trehalose biosynthetic pathway for the life of yeasts and fungi. FEMS yeast research, 4**,** 351-359.

GAO, X., HAN, J., LU, Z., LI, Y. & HE, C. (2013) De novo assembly and characterization of spotted seal Phoca largha transcriptome using Illumina paired-end sequencing. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 8**,** 103-110.

GARG, A. K., KIM, J.-K., OWENS, T. G., RANWALA, A. P., DO CHOI, Y., KOCHIAN, L. V. & WU, R. J. (2002) Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proceedings of the National Academy of Sciences, 99**,** 15898-15903.

GAZI, A. V., KYRIACOU, A., KOTSOU, M. & LASARIDI, K. E. (2007) Microbial Community Dynamics and Stability Assessment During Green Waste Composting. Global NEST Journal, 9**,** 35-41.

GHARIZADEH, B., AKHRAS, M., NOURIZAD, N., GHADERI, M., YASUDA, K., NYRÉN, P. & POURMAND, N. (2006) Methodological improvements of pyrosequencing technology. Journal of Biotechnology, 124**,** 504-511.

GHARIZADEH, B., ERIKSSON, J., NOURIZAD, N., NORDSTRÖM, T. & NYRÉN, P. (2004) Improvements in Pyrosequencing technology by employing Sequenase polymerase. Analytical Biochemistry, 330**,** 272-280.

GHORAI, S., BANIK, S. P., VERMA, D., CHOWDHURY, S., MUKHERJEE, S. & KHOWALA, S. (2009) Fungal biotechnology in food and feed processing. Food Research International, 42**,** 577-587.

GIAEVER, G., CHU, A. M., NI, L., CONNELLY, C., RILES, L., VERONNEAU, S., DOW, S., LUCAU-DANILA, A., ANDERSON, K., ANDRE, B., ARKIN, A. P., ASTROMOFF, A., EL BAKKOURY, M., BANGHAM, R., BENITO, R., BRACHAT, S., CAMPANARO, S., CURTISS, M., DAVIS, K., DEUTSCHBAUER, A., ENTIAN, K.-D., FLAHERTY, P., FOURY, F., GARFINKEL, D. J., GERSTEIN, M., GOTTE, D., GULDENER, U., HEGEMANN, J. H., HEMPEL, S., HERMAN, Z., JARAMILLO, D. F., KELLY, D. E., KELLY, S. L., KOTTER, P., LABONTE, D., LAMB, D. C., LAN, N., LIANG, H., LIAO, H., LIU, L., LUO, C., LUSSIER, M., MAO, R., MENARD, P., OOI, S. L., REVUELTA, J. L., ROBERTS, C. J., ROSE, M., ROSS-MACDONALD, P., SCHERENS, B., SCHIMMACK, G., SHAFER, B., SHOEMAKER, D. D., SOOKHAI-MAHADEO, S., STORMS, R. K., STRATHERN, J. N., VALLE, G., VOET, M., VOLCKAERT, G., WANG, C.-Y., WARD, T. R., WILHELMY, J., WINZELER, E. A., YANG, Y., YEN, G., YOUNGMAN, E., YU, K., BUSSEY, H., BOEKE, J. D., SNYDER, M., PHILIPPSEN, P., DAVIS, R. W. & JOHNSTON, M. (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature, 418**,** 387-391.

GILLES, A., MEGLECZ, E., PECH, N., FERREIRA, S., MALAUSA, T. & MARTIN, J.-F. (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC genomics, 12**,** 245.

GOFFEAU, A., BARRELL, B. G., BUSSEY, H., DAVIS, R. W., DUJON, B., FELDMANN, H., GALIBERT, F., HOHEISEL, J. D., JACQ, C., JOHNSTON, M., LOUIS, E. J., MEWES, H. W., MURAKAMI, Y., PHILIPPSEN, P., TETTELIN, H. & OLIVER, S. G. (1996) Life with 6000 genes. SCIENCE, 274**,** 563-567.

GOODWIN, S. B., M'BAREK, S. B., DHILLON, B., WITTENBERG, A. H. J., CRANE, C. F., HANE, J. K., FOSTER, A. J., VAN DER LEE, T. A. J., GRIMWOOD, J., AERTS, A., ANTONIW, J., BAILEY, A., BLUHM, B., BOWLER, J., BRISTOW, J., VAN DER BURGT, A., CANTO-CANCHA, B., CHURCHILL, A. C. L., CONDE-FERRÃ EZ, L., COOLS, H. J., COUTINHO, P. M., CSUKAI, M., DEHAL, P., DE WIT, P., DONZELLI, B., VAN DE GEEST, H. C., VAN HAM, R. C. H. J., HAMMOND-KOSACK, K. E., HENRISSAT, B., KILIAN, A., KOBAYASHI, A. K., KOOPMANN, E., KOURMPETIS, Y., KUZNIAR, A., LINDQUIST, E., LOMBARD, V., MALIEPAARD, C., MARTINS, N., MEHRABI, R., NAP, J. P. H., PONOMARENKO, A., RUDD, J. J., SALAMOV, A., SCHMUTZ, J., SCHOUTEN, H. J., SHAPIRO, H., STERGIOPOULOS, I., TORRIANI, S. F. F., TU, H., DE VRIES, R. P., WAALWIJK, C., WARE, S. B., WIEBENGA, A., ZWIERS, L. H., OLIVER, R. P., GRIGORIEV, I. V. & KEMA, G. H. J. (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS genetics, 7**,** e1002070.

GOSTINCAR, G., GUNDE-CIMERMAN, N. & TURK, M. (2012) Genetic resources of extremotolerant fungi: A method for identification of genes conferring stress tolerance. Bioresource Technology, 111**,** 360–367.

GOUGH, J., KARPLUS, K., HUGHEY, R. & CHOTHIA, C. (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. Journal of Molecular Biology, 313**,** 903-919.

GOUKA, R. J., PUNT, P. J. & VAN DEN HONDEL, C. A. M. J. J. (1997) Efficient production of secreted proteins by Aspergillus: progress, limitations and prospects. Applied Microbiology and Biotechnology, 47**,** 1-11.

GREMEL, G., DORRER, M. & SCHMOLL, M. (2008) Sulphur metabolism and cellulase gene expression are connected processes in the filamentous fungus Hypocrea jecorina (anamorph Trichoderma reesei). BMC microbiology, 8**,** 174.

GROSSETÊTE, S., LABEDAN, B. & LESPINET, O. (2010) FUNGIpath: a tool to assess fungal metabolic pathways predicted by orthology. GMC Genomics, 11**,** 81.

GU, Y., DING, Y., REN, C., SUN, Z., RODIONOV, D. A., ZHANG, W., YANG, S., YANG, C. & JIANG, W. (2010) Reconstruction of xylose utilization pathway and regulons in *Firmicutes*. BMC Genomics, 11.

GUILLOU, V., PLOURDE-OWOBI, L., PARROU, J. L., GOMA, G. & FRANÇOIS, J. (2004) Role of reserve carbohydrates in the growth dynamics of *Saccharomyces cerevisiae.* FEMS Yeast Research, 4**,** 773-787.

GUO, R.-F., SHI, B.-S., LI, D.-C., MA, W. & WEI, Q. (2008) Purification and Characterization of a Novel Thermostable Chitinase from< i> Thermomyces lanuginosus</i> SY2 and Cloning of Its Encoding Gene. Agricultural Sciences in China, 7**,** 1458-1465.

HANE, J. K., ROUXEL, T., HOWLETT, B. J., KEMA, G. H. J., GOODWIN, S. B. & OLIVER, R. P. (2011) A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. Genome Biology, 12**,** 45-57.

HANNEMANN, F., BICHET, A., EWEN, K. M. & BERNHARDT, R. (2007) Cytochrome P450 systems—biological variations of electron transport chains. Biochimica et Biophysica Acta (BBA)-General Subjects, 1770**,** 330-344.

HARIDAS, S., BREUILL, C., BOHLMANN, J. & HSIANG, T. (2011) A biologist's guide to de novo genome assembly using next-generation sequence data: A test with fungal genomes. . Journal of Microbiological Methods., 86**,** 368-375.

HARTL, L., ZACH, S. & SEIDL-SEIBOTH, V. (2012) Fungal chitinases: diversity, mechanistic properties and biotechnological potential. Applied microbiology and biotechnology, 93**,** 533-543.

HASAN, F., SHAH, A. A. & HAMEED, A. (2006) Industrial applications of microbial lipases. Enzyme and Microbial Technology, 39**,** 235-251.

HEATON, E. A., DOHLEMAN, F. G. & LONG, S. P. (2008) Meeting us biofuel goals with less land:potential of miscanthus. Global Change Biology Bioenergy, 14**,** 2000-2014.

HEDEGAARD, J., BICCIATO, S., BONNET, A., RAMIREZ BOO, M., BUITENHUIS, B., COLLADO-ROMERO, M., NAGSTRUP CONLEY, L., SAN CRISTOBAL, M., FERRAR, I., GROENEN, M., HORNSHOJ, H., HULSEGGE, I., JIANG, L., ARCE JIMENEZ, C., JIMENEZ-MARIN, A., KOMMADATH, A., LAGARRIGUE, S., LEUNISSEN, J., LIAUBET, L., NEERINCX, P., NIE, H., GARRIDO PAVON, J., PRICKETT, D., REBEL, J., ROBERT-GRANIE, C., SKARMAN, A., SMITS, M., SORENSEN, P., TOSSER-KLOPP, G., POEL, J. & WATSON, M. (2009) Methods for interpreting lists of affected genes obtained in a DNA microarray experiment. BMC Proceedings, 3**,** S5.

HELLER, M. J. (2002) DNA microarray technology: devices, systems, and applications. Annual review of biomedical engineering, 4**,** 129-153.

HESPELLE, B. & WHITHEAD, T. R. (1990) Physiology and Genetics of Xylan Degradation by Gastrointestinal Tract Bacteria Journal of Dairy Science 73**,** 3013-3022.

HIBBETT, D. S., BINDER, M., BISCHOFF, J. F., BLACKWELL, M., CANNON, P. F., ERIKSSON, O. E., HUHNDORF, S., JAMES, T., KIRK, P. M., LÜCKING, R., THORSTEN LUMBSCH, H., LUTZONI, F., MATHENY, P. B., MCLAUGHLIN, D. J., POWELL, M. J., REDHEAD, S., SCHOCH, C. L., SPATAFORA, J. W., STALPERS, J. A., VILGALYS, R., AIME, M. C., APTROOT, A., BAUER, R., BEGEROW, D., BENNY, G. L., CASTLEBURY, L. A., CROUS, P. W., DAI, Y.-C., GAMS, W., GEISER, D. M., GRIFFITH, G. W., GUEIDAN, C., HAWKSWORTH, D. L., HESTMARK, G., HOSAKA, K., HUMBER, R. A., HYDE, K. D., IRONSIDE, J. E.,

KÕLJALG, U., KURTZMAN, C. P., LARSSON, K.-H., LICHTWARDT, R., LONGCORE, J., MIĄDLIKOWSKA, J., MILLER, A., MONCALVO, J.-M., MOZLEY-STANDRIDGE, S., OBERWINKLER, F., PARMASTO, E., REEB, V., ROGERS, J. D., ROUX, C., RYVARDEN, L., SAMPAIO, J. P., SCHÜßLER, A., SUGIYAMA, J., THORN, R. G., TIBELL, L., UNTEREINER, W. A., WALKER, C., WANG, Z., WEIR, A., WEISS, M., WHITE, M. M., WINKA, K., YAO, Y.-J. & ZHANG, N. (2007) A higher-level phylogenetic classification of the Fungi. Mycological Research, 111**,** 509-547.

HRMOVA, M., PETRAKOVA, E. & BIELY, P. (1991) Induction of cellulose and xylan-degrading enzyme systems in *Aspergillus terreus* by homo- and heterodisaccharides composed of glucose and xylose. Journal of General Microbiology, 137**,** 541-547.

HULL, D., WOLSTENCROFT, K., STEVENS, R., GOBLE, C., POCOCK, M. R., LI, P. & OINN, T. (2006) Taverna: a tool for building and running workflows of services. Nucleic Acids Research, 34**,** W729-W732.

IGARASHI, K., TANI, T., RIE, K. & MASAHIRO, S. (2003) Family 3 β-glucosidase from cellulose-degrading culture of the white-rot fungus Phanerochaete chrysosporium is a glucan 1,3-β-glucosidase. Journal of Bioscience and Bioengineering, 95**,** 572-576.

ITO, M., BABA, T., MORI, H. & MORI, H. (2005) Functional analysis of 1440 *Escherichia coli* genes using the combination of knock-out library and phenotype microarrays. Metabolic engineering, 7**,** 318-327.

JAMES, J. C. (2006) Understanding the Pathogenic Fungus *Penicillium marneffei*: A Computational Genomics Perspective.

JOSEPH-HORNE, T. & HOLLOMON, D. W. (2000) Functional diversity within the mitochondrial electron transport chain of plant pathogenic fungi. Pest management science, 56**,** 24-30.

JOUBERT, A., SIMONEAU, P., CAMPION, C., BATAILLA, S. N., IACOMIA, V. B., POUPARD, P., FRANASOIS, J. M., GEORGEAULT, S., SELLIER, E. &

GUILLEMETTE, T. (2011) Impact of the unfolded protein response on the pathogenicity of the necrotrophic fungus *Alternaria brassicicola*. Molecular Microbiology, 79**,** 1305-1324.

KAGAWA, Y., NOJIMA, H., NUKIWA, N., ISHIZUKA, M., NAKAJIMA, T., YASUHARA, T., TANAKA, T. & OSHIMA, T. (1984) High guanine plus cytosine content in the third letter of codons of an extreme thermophile. DNA sequence of the isopropylmalate dehydrogenase of Thermus thermophilus. Journal of Biological Chemistry, 259**,** 2956-2960.

KASHEFI, K. (2003) Extending the upper-temperature limit for life. SCIENCE, 30`**,** 934.

KAZUMI, T., YOSHINO, G., ISHIDA, Y., IWATANI, I., MORITA, S., TATEIWA, M. & KASUGA, M. (1995) Long-term efficacy and tolerability of pravastatin in hypercholesterolemia in patients with non-insulin-dependent diabetes mellitus. Diabetes Research and Clinical Practice, 27**,** 61-68.

KERN, A., TILLEY, E., HUNTER, I. S., LEGIŠA, M. & GLIEDER, A. (2007) Engineering primary metabolic pathways of industrial micro-organisms. Journal of Biotechnology, 129**,** 6-29.

KIM, C., LILLEHOJ, H., HONG, Y. & KEELER, C. (2008a) Comparison of transcriptional changes associated with E. acervulina and E. maxima infections using cDNA microarray technology. Dev Biol (Basel), 132**,** 121 - 130.

KIM, J. C., SANDS, J. S., MULLAN, B. P. & PLUSKE, J. R. (2008b) Performance and total-tract digestibility responses to exogenous xylanase and phytase in diets for growing pigs. Animal Feed Science and Technology, 142**,** 163-172.

KIRK, O., BORCHERT, T. V. & FUGLSANG, C. C. (2002) Industrial enzyme applications. Current Opinion in Biotechnology, 13**,** 345-351.

KIRSANOW, K. & BURGER, J. (2012) Ancient human DNA. Annals of Anatomy - Anatomischer Anzeiger, 194**,** 121-132.

KO, H. G., PARK, H. G., PARK, S. H., CHOI, C. W., KIM, S. H. & PARK, W. M. (2005) Comparative study of mycelial growth and basidiomata formation in seven different species of the edible mushroom genus Hericium. Bioresource Technology, 96**,** 1439-1444.

KRAUSE, J., FU, Q., GOOD, J. M., VIOLA, B., SHUNKOV, M. V., DEREVIANKO, A. P. & PAABO, S. (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature, 464**,** 894-897.

KRONSTAD, J. W., ATTARIAN, R., CADIEUX, B., CHOI, J., D'SOUZA, C. A., GRIFFITHS, E. J., GEDDES, J. M. H., HU, G., JUNG, W. H., KRETSCHMER, M., SAIKIA, S. & WANG, J. (2011) Expanding fungal pathogenesis: *Cryptococcus* breaks out of the opportunistic box. Nature reviews. Microbiology, 9**,** 193-203.

KUBICEK, C. P., HERRERA-ESTRELLA, A., SEIDL-SEIBOTH, V., MARTINEZ, D. A., DRUZHININA, I. S., THON, M., ZEILINGER, S., CASAS-FLORES, S., HORWITZ, B. A., MUKHERJEE, P. K., MUKHERJEE, M., KREDICS, L., ALCARAZ, L. D., AERTS, A., ANTAL, Z., ATANASOVA, L., CERVANTES-BADILLO, M. G., CHALLACOMBE, J., CHERTKOV, O., MCCLUSKEY, K., COULPIER, F., DESHPANDE, N., VON DOHREN, H., EBBOLE, D. J., ESQUIVEL-NARANJO, E. U., FEKETE, E., FLIPPHI, M., GLASER, F., GOMEZ-RODRIGUEZ, E. Y., GRUBER, S., HAN, C., HENRISSAT, B., HERMOSA, R., HERNANDEZ-ONATE, M., KARAFFA, L., KOSTI, I., LE CROM, S., LINDQUIST, E., LUCAS, S., LUBECK, M., LUBECK, P. S., MARGEOT, A., METZ, B., MISRA, M., NEVALAINEN, H., OMANN, M., PACKER, N., PERRONE, G., URESTI-RIVERA, E. E., SALAMOV, A., SCHMOLL, M., SEIBOTH, B., SHAPIRO, H., SUKNO, S., TAMAYO-RAMOS, J. A., TISCH, D., WIEST, A., WILKINSON, H. H., ZHANG, M., COUTINHO, P. M., KENERLEY, C. M., MONTE, E., BAKER, S. E. & GRIGORIEV, I. V. (2011) Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. Genome Biology 12**,** 40.

KUMAR, K. S., MANIMARAN, A., PERMAUL, K. & SINGH, S. (2009c) Production of β-xylanase by a Thermomyces lanuginosus MC 134 mutant on corn cobs and its application

in biobleaching of bagasse pulp. Journal of Bioscience and Bioengineering, 107**,** 494-498.

KUMAR, S. & BLAXTER, M. L. (2010) Comparing *de novo* assemblers for 454 transcriptome data. BMC Genomics, 11**,** 571.

KUNAMNENI, A., PERMAUL, K. & SINGH, S. (2005) Amylase production in solid state fermentation by the thermophilic fungus Thermomyces lanuginosus. Journal of Bioscience and Bioengineering, 100**,** 168-171.

KUTTANPILLAI, S., KUMAR, K. P. & SINGH, S. (2009a) Inducible character of b-xylanase in a hyperproducing mutant of *Thermomyces lanuginosus*. Engineering in Life Sciences, 9**,** 298–302.

KUTTANPILLAI, S. K., AYYACHAMY, M., KUGEN, P. & SUREN, S. (2009b) Production of β-xylanase by a *Thermomyces lanuginosus* MC 134 mutant on corn cobs and its application in biobleaching of bagasse pulp. Journal of Bioscience and Bioengineering, 107**,** 494–498.

LA GRANGE, D. C., DEN HAAN, R. & VAN ZYL, W. H. (2010) Engineering cellulolytic ability into bioprocessing organisms. Applied Microbiology and Biotechnology, 87**,** 1195-1208.

LAFOND, M., TAUZIN, A., DESSEAUX, V., BONNIN, E., AJANDOUZ, E.-H. & GIARDINA, T. (2011) GH10 xylanase D from *Penicillium funiculosum*: biochemical studies and xylooligosaccharide production. Microbial cell factories, 10**,** 20-20.

LAMOOLPHAK, W., GOTO, M., SASAKI, M., SUPHANTHARIKA, M., MUANGNAPOH, C., PROMMUAG, C. & SHOTIPRUK, A. (2006) Hydrothermal decomposition of yeast cells for production of proteins and amino acids. Journal of Hazardous Materials, 137**,** 1643-1648.

LANGE, L. (2010) The importance of fungi for a more sustainable future on our planet. Fungal Biology Reviews, 24**,** 90-92.

LAPIDUS, A. L. 2009. Genome Sequence Databases: Sequencing and Assembly. *In:* EDITOR-IN-CHIEF: MOSELIO, S. (ed.) *Encyclopedia of Microbiology (Third Edition).* Oxford: Academic Press.

LASSEN, K. S., SCHULTZ, H., HEEGAARD, N. H. H. & HE, M. (2012) A novel DNAseq program for enhanced analysis of Illumina GAII data: a case study on antibody complementarity-determining regions. New BIOTECHNOLOGY, 29**,** 271-278.

LATGÉ, J.-P. & CALDERONE, R. (2002) Host–microbe interactions: fungi invasive human fungal opportunistic infections. Current Opinion in Microbiology, 5**,** 355-358.

LAVOIE, H., HOGUES, H. & WHITEWAY, M. (2009) Rearrangements of the transcriptional regulatory networks of metabolic pathways in fungi. Current Opinion in Microbiology, 12**,** 655-663.

LEA, A. & IRINA, S. D. (2010) Global nutrient profiling by Phenotype MicroArrays: a tool complementing genomic and proteomic studies in conidial fungi. Journal of Biomedicine and Biotechnology, 11**,** 151-168.

LI, S., YANG, X., YANG, S., ZHU, M. & WANG, M. (2012) Technology Prospecting on Enzymes: Application, Marketing and Engineering. Computational and structural Biochenology, 3**,** 1-11.

LINKS, M. G., HOLUB, E., JIANG, R. H. Y., SHARPE, A. G., HEGEDUS, D., BEYNON, E., SILLITO, D., CLARKE, W. E., UZUHASHI, S. & BORHAN, M. H. (2011) De novo sequence assembly of Albugo candida reveals a small genome relative to other biotrophic oomycetes. BMC genomics, 12, 12**,** 503-503.

LITTLEJOHN, K. A., HOOLEY, P. & COX, P. W. (2012) Bioinformatics predicts diverse Aspergillus hydrophobins with novel propertie. Food Hydrocolloids, 27**,** 503-516.

LIU, E. T. (2005) Genomic technologies and the interrogation of the transcriptome. Mechanisms of Ageing and Development, 126**,** 153-159.

LIU, E. T., GOEL, S., DESAI, K. & VOORHOEVE, M. 2010. Chapter 2 - Genomic Technologies for Systems Biology. *In:* EDISON, T. L. & DOUGLAS, A. L. (eds.) *Systems Biomedicine.* San Diego: Academic Press.

LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. (2012) Comparison of Next-Generation Sequencing Systems. Journal of Biomedicine and Biotechnology, 2012**,** 11.

LUKE, K. A., HIGGINS, C. L. & WITTUNG-STAFSHEDE, P. (2007) Thermodynamic stability and folding of proteins from hyperthermophilic organisms. FEBS Journal, 274**,** 4023-4033.

MA, L., VAN DER DOES, H. C., BORKOVICH, K. A., COLEMAN, J. J., DABOUSSI, M., DI PIETRO, A., DUFRESNE, M., FREITAG, M., GRABHERR, M., HENRISSAT, B., HOUTERMAN, P. M., KANG, S., SHIM, W., WOLOSHUK, C., XIE, X., XU, J., ANTONIW, J., BAKER, S. E., BLUHM, B. H., BREAKSPEAR, A., BROWN, D. W., BUTCHKO, R. A. E., CHAPMAN, S., COULSON, R., COUTINHO, P. M., DANCHIN, E. G. J., DIENER, A., GALE, L. R., GARDINER, D. M., GOFF, S., HAMMOND-KOSACK, K. E., HILBURN, K., HUA-VAN, A., JONKERS, W., KAZAN, K., KODIRA, C. D., KOEHRSEN, M., KUMAR, L., LEE, Y., LI, L., MANNERS, J. M., MIRANDA-SAAVEDRA, D., MUKHERJEE, M., PARK, G., PARK, J., PARK, S., PROCTOR, R. H., REGEV, A., RUIZ-ROLDAN, M. C., SAIN, D., SAKTHIKUMAR, S., SYKES, S., SCHWARTZ, D. C., TURGEON, B. G., WAPINSKI, I., YODER, O., YOUNG, S., ZENG, Q., ZHOU, S., GALAGAN, J., CUOMO, C. A., KISTLER, H. C. & REP, M. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. NATURE, 464**,** 367-373.

MAALEJ-ACHOURI, I., GUERFALI, M., ROMDHANE, I. B.-B., GARGOURI, A. & BELGHITH, H. (2012) The effect of Talaromyces thermophilus cellulase-free xylanase and commercial laccase on lignocellulosic components during the bleaching of kraft pulp. International Biodeterioration & Biodegradation, 75**,** 43-48.

MACCALLUM, I., PRZYBYLSKI, D., GNERRE, S., BURTON, J., SHLYAKHTER, I., GNIRKE, A., MALEK, J., MCKERNAN, K., RANADE, S., SHEA, T. P., WILLIAMS, L., YOUNG, S., NUSBAUM, C. & JAFFE, D. B. (2009) ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. Genome biology, 10**,** R103-R103.

MACH-AIGNER, A. R., PUCHER, M. E. & MACH, R. L. (2010) D-Xylose as a Repressor or Inducer of Xylanase Expression in *Hypocrea jecorina (Trichoderma reesei)*. Applied and Environmental Microbiology, 76**,** 1770-1776.

MACH, R. L. & ZEILINGER, S. (2003) Regulation of gene expression in industrial fungi: *Trichoderma*. Applied Microbiology and Biotechnology 60.

MACHIDA M, ASAI K, SANO M, TANAKA T, KUMAGAI T, TERAI G, KUSUMOTO K, ARIMA T, AKITA O, KASHIWAGI Y, ABE K, GOMI K, HORIUCHI H, KITAMOTO K, KOBAYASHI T, TAKEUCHI M, DENNING DW, GALAGAN JE, NIERMAN WC, YU J, ARCHER DB, BENNETT JW, BHATNAGAR D, CLEVELAND TE, FEDOROVA ND, GOTOH O, HORIKAWA H, HOSOYAMA A, ICHINOMIYA M, IGARASHI R, IWASHITA K, JUVVADI PR, KATO M, KATO Y, KIN T, KOKUBUN A, MAEDA H, MAEYAMA N, MARUYAMA J, NAGASAKI H, NAKAJIMA T, ODA K, OKADA K, PAULSEN I, SAKAMOTO K, SAWANO T, TAKAHASHI M, TAKASE K, TERABAYASHI Y, WORTMAN JR, YAMADA O, YAMAGATA Y, ANAZAWA H, HATA Y, KOIDE Y, KOMORI T, KOYAMA Y, MINETOKI T, SUHARNAN S, TANAKA A, ISONO K, KUHARA S, OGASAWARA N & H., K. (2005) Genome sequencing and analysis of *Aspergillus oryzae*. Nature, 438**,** 1157-1161.

MADLALA, A., BISSOON, S., SINGH, S. & CHRISTOV, L. (2001) Xylanase-induced reduction of chlorine dioxide consumption during elemental chlorine-free bleaching of different pulp types. Biotechnology Letters, 23**,** 345-351.

MAGNUSON, J. & LASURE., L. L. (2004) Organic Acid Production by Filamentous Fungi. Advances in Fungal Biotechnology for Industry, Agriculture, and Medicine

MAHESHWARI, R., BHARADWAJ, G. & MAHAKINGESGWARA, K. B. (2000) Thermophilic Fungi: Their Physiology and Enzymes. Microbiology and Molecular Biology Reviews, 64**,** 461-488.

MANDELS, M. & ANDREOTTI, R. (1978) Problems and challenges in the cellulose to cellulase fermentation. Proc Biochem, 13**,** 6 - 13.

MANIMARAN, A., KUTTAN-PILLAI, S. K., PERMAUL, K. & SINGH, S. (2009) Hyper production of cellulase-free xylanase by *Thermomyces lanuginosus* SSBP on bagasse pulp and its application in biobleaching. Applied Microbiology and Biotechnology, 81**,** 887-893.

MARCET-HOUBEN, M., MARCEDDU, G. & GABALDON, T. (2009) Phylogenomics of the oxidative phosphorylation in fungi reveals extensive gene duplication followed by functional divergence. BMC Evolutionary Biology, 9**,** 295.

MARCO, A. V. D. B., RICHARD, A., KAJ, A., JONATHAN, H. B., JEAN-MARC, D., ARNOLD, J. M. D., CARLOS, G. E., NATALIE, D. F., DIANA, M. H., WILBERT, H. M. H., VINITA, J., JAN, A. K. W. K., ANDRIY, K., JUAN, F. M., WILLIAM, C. N., JEROEN, G. N., JACK, T. P., JOHANNES, A. R., IDA, J. V. D. K., NOEL, N. M. E. V. P., MARTEN, V., HANS, V. D., CHRISTIAN, W., JENNIFER, W. & ROEL, A. L. B. (2008) Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. Nature biotechnology, 26.

MARJAMAA, K., TOTH, K., BROMANN, P. A., SZAKACS, G. & KRUUS, K. (2013) Novel Penicillium cellulases for total hydrolysis of lignocellulosics. Enzyme and Microbial Technology, 52**,** 358-369.

MARQUEZ, A., ZAVALA-PAJRAMO, M. G., LAPEZ-ROMERO, E. & CAMACHO, H. C. (2011) Biotechnological potential of pectinolytic complexes of fungi. Biotechnology letters, 33**,** 859-868.

MARTIN, F., KOHLER, A., MURA, C., BALESTRINI, R., M., C. P., JAILLON, O., MONTANINI, B., MORIN, E., NOEL, B., PERCUDANI, R., PORCEL, B., RUBINI,

A., AMICUCCI, A., AMSELEM, J., ANTHOUARD, V., ARCIONI, S., ARTIGUENAVE, F., AURY, J., BALLARIO, P., BOLCHI, A., BRENNA, A., BRUN, A., BUE´E, M., CANTAREL, B., CHEVALIER, G., COULOUX, A., DA SILVA, C., DENOEUD, F., DUPLESSIS, S., GHIGNONE, S., HILSELBERGER, B., IOTTI, M., MARCAIS, B., MELLO, A., MIRANDA, M., PACIONI, G., QUESNEVILLE, H., RICCIONI, C., RUOTOLO, R., SPLIVALLO, R., STOCCHI, V., TISSERANT, E., VISCOMI, A. R., ZAMBONELLI, A., ZAMPIERI, E., HENRISSAT, B., LEBRUN, M., PAOLOCCI, F., BONFANTE, P., OTTONELLO, S. & WINCKER, P. (2010) Perigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature, 464**,** 1033-1038.

MARTIN, W., RUJAN, T., RICHLY, E., HANSEN, A., CORNELSEN, S., LINS, T., LEISTER, D., STOEBE, B., HASEGAWA, M. & PENNY, D. (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proceedings of the National Academy of Sciences, 99**,** 12246-12251.

MARTINEZ, D., CULLEN, D., DANCHIN, E. G. J., GRIGORIEV, I. V., HARRIS, P., JACKSON, M., KUBICEK, C. P., HAN, C. S., HO, I., LARRONDO, L. F., DE LEON, A. L., BERKA, R. M., MAGNUSON, J. K., MERINO, S., MISRA, M., NELSON, B., PUTNAM, N., ROBBERTSE, B., SALAMOV, A. A., SCHMOLL, M., TERRY, A., THAYER, N., HENRISSAT, B., WESTERHOLM-PARVINEN, A., SCHOCH, C. L., YAO, J., BARABOTE, R., BARBOTE, R., NELSON, M. A., DETTER, C., BRUCE, D., KUSKE, C. R., XIE, G., SALOHEIMO, M., RICHARDSON, P., ROKHSAR, D. S., LUCAS, S. M., RUBIN, E. M., DUNN-COLEMAN, N., WARD, M., BRETTIN, T. S., ARVAS, M., BAKER, S. E., CHAPMAN, J., CHERTKOV, O. & COUTINHO, P. M. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nature biotechnology, 26**,** 553-560.

MARUYAMA-NAKASHITA, A., NAKAMURA, Y., YAMAYA, T. & TAKAHASHI, H. (2004) Regulation of high-affinity sulphate transporters in plants: towards systematic analysis of sulphur signalling and regulation. J Exp Bot, 55**,** 1843 - 1849.

MASUREKAR, P. 2009. Antibiotic Production. *In:* EDITOR-IN-CHIEF: MOSELIO, S. (ed.) *Encyclopedia of Microbiology (Third Edition).* Oxford: Academic Press.

MCHUNU, N. P., SINGH, S. & PERMAUL, K. (2009) Expression of an alkalo-tolerant fungal xylanase enhanced by directed evolution in Pichia pastoris and Escherichia coli. Journal of Biotechnology, 141**,** 26-30.

MCLAUGHLIN, D. J., HIBBETT, D. S., LUTZONI, F., SPATAFORA, J. W. & VILGALYS, R. (2009) The search for the fungal tree of life. Trends in Microbiology, 17**,** 488-497.

MELO, S. C. O., PUNGARTNIK, C., CASCARDO, J. C. M. & BRENDEL, M. (2006) Rapid and efficient protocol for DNA extraction and molecular identification of the basidiomycete *Crinipellis perniciosa*. Genetics and Molecular Research, 5**,** 851-855.

MILLER, W., GUILLORY, J., STINSON, J., SESHAGIRI, S. & SCHUSTERS (2013) Comparison of Sequencing Platforms for Single Nucleotide Variant Calls in a Human Sample. PLoS One, 8**,** e55089.

MINOCHE, A. E., DOHM, J. C. & HIMMELBAUER, H. (2012) Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. Genome Biology, 12**,** R112

MISHRA, K. K., PAL, R. S., ARUNKUMAR, R., CHANDRASHEKARA, C., JAIN, S. K. & BHATT, J. C. (2013) Antioxidant properties of different edible mushroom species and increased bioconversion efficiency of Pleurotus eryngii using locally available casing materials. Food Chemistry, 138**,** 1557-1563.

MOHAMMADA, S. H., JAHIMA, J. M., NOPIAHB, Z., MURADC, A. M. A., BAKARC, F. D. A. & ILLIASD, R. (2012) Preliminary Study on Diverse Carbon Utilization by Transformant *Aspergillus niger*. International Journal on Advanced Science E n g i n e e r i n g Information Technology, 2**,** 11-15.

MORGENSTERN, I., POWLOWSKI, J., ISHMAEL, N., DARMOND, C., MARQUETEAU, S., MOISAN, M., QUENNEVILLE, G. E. & TSANG, A. (2012) A molecular phylogeny of thermophilic fungi. Fungal Biology, 116**,** 489-502.

MOROZOVA, O. & MARRA, M. A. (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics, 92**,** 255-264.

NAGANAGOUDA, K., SALIMATH, P. V. & MULIMANI, V. H. (2009) Purification and Characterization of Endo-β-1,4 Mannanase from *Aspergillus niger gr* for Application in Food Processing Industry. Journal of Microbiology and Biotechnology, 19**,** 1184–1190.

NARA, T., HSHIMOTO, T. & AOKI, T. (2000) Evolutionary implications of the mosaic pyrimidine-biosynthetic pathway in eukaryotes. Gene, 257**,** 209-222.

NEVALAINEN, K. M. H., TE'O, V. S. J. & BERGQUIST, P. L. (2005) Heterologous protein expression in filamentous fungi. Trends in Biotechnology, 23**,** 468-474.

NIEHAUS, F., BERTOLDO, C., KAEHLER, M. & ANTRANIKIAN, G. (1999) Extremophiles as a source of novel enzymes for industrial application. Applied Microbiology and Biotechnology, 51**,** 711-729.

NIERMAN, W. C., PAIN A, ANDERSON M.J, WORTMAN J.R, KIM H.S, ARROYO J, BERRIMAN M, ABE K, ARCHER D.B, BERMEJO C, BENNETT J, BOWYER P, CHEN D, COLLINS M, COULSEN R, DAVIES R, DYER PS, FARMAN M, FEDOROVA N, FEDOROVA N, FELDBLYUM TV, FISCHER R, FOSKER N, FRASER A, GARCÍA J.L, GARCÍA M.J, GOBLE A, GOLDMAN G.H, GOMI K, GRIFFITH-JONES S, GWILLIAM R, HAAS B, HAAS H, HARRIS D, HORIUCHI H, HUANG J, HUMPHRAY S, JIMÉNEZ J, KELLER N, KHOURI H, KITAMOTO K, KOBAYASHI T, KONZACK S, KULKARNI R, KUMAGAI T, LAFON A, LATGÉ J.P, LI W, LORD A, LU C, MAJOROS W.H, MAY G.S, MILLER B.L, MOHAMOUD Y, MOLINA M, MONOD M, MOUYNA I, MULLIGAN S, MURPHY L, O'NEIL S, PAULSEN I, PEÑALVA M.A, PERTEA M, PRICE C, PRITCHARD B.L, QUAIL M.A, RABBINOWITSCH E, RAWLINS N, RAJANDREAM M.A, REICHARD U, RENAULD H, ROBSON G.D, RODRIGUEZ DE CÓRDOBA S, RODRÍGUEZ-PEÑA J.M, RONNING C.M, RUTTER S, SALZBERG S.L, SANCHEZ M, SÁNCHEZ-FERRERO J.C, SAUNDERS D, SEEGER K, SQUARES R, SQUARES S, TAKEUCHI M, TEKAIA F, TURNER G, VAZQUEZ DE ALDANA C.R, WEIDMAN J, WHITE O,

WOODWARD J, YU JH, FRASER C, GALAGAN J.E, ASAI K, MACHIDA M, HALL N, BARRELL B & D.W., D. (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. Nature, 438**,** 22-29.

NIGAM, P. S. & SINGH, A. (2011) Production of liquid biofuels from renewable resources. Progress in Energy and Combustion Science, 37**,** 52-68.

NIKITIN, A., EGOROV, S., DARASELIA, N. & MAZO, I. (2003) Pathway studio—the analysis and navigation of molecular networks. Bioinformatics, 19**,** 2155-2157.

NOWAK, S. & CORCES, V. (2004) Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. TRENDS in Genetics, 20.

NOWROUSIAN, M., KUO, H.-C., OSIEWACZ, H. D., PAGGELLER, S., READ, N. D., SEILER, S., SMITH, K. M., ZICLER, D., KACK, U., FREITAG, M., STAJICH, J. E., CHU, M., ENGH, I., ESPAGNE, E., HALLIDAY, K., KAMEREWERED, J., KEMPKEN, F. & KNAB, B. (2010) De novo assembly of a 40 Mb eukaryotic genome from short sequence reads: *Sordaria macrospora*, a model organism for fungal morphogenesis. PLoS genetics, 6**,** e1000891.

OKHUOYA, J. A. & ETUGO, J. E. (1993) Studies of the cultivation of Pleurotus tuberregium (FR) sing. An edible mushroom. Bioresource Technology, 44**,** 1-3.

OLĘDZKA, G., DĄBROWSKI, S. & KUR, J. (2003) High-level expression, secretion, and purification of the thermostable aqualysin I from Thermus aquaticus YT-1 in Pichia pastoris. Protein Expression and Purification, 29**,** 223-229.

OVERBEEK, R., LARSEN, N., PUSCH, G. D., D'SOUZA, M., JR, E. S., KYRPIDES, N., FONSTEIN, M., MALTSEV, N. & SELKOV, E. (2000) WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. Nucleic Acids Research, 28**,** 123-125.

PARIZA, M. W. & COOK, M. (2010) Determining the safety of enzymes used in animal feed. Regulatory Toxicology and Pharmacology, 56**,** 332-342.

PATIST, A. & ZOERB, H. (2005) Preservation mechanisms of trehalose in food and biosystems. Collect. Surf. B: Biointerfaces 40**,** 107-112.

PEARCE, C. (2012) Review of Fungi: Biology and Applications. Journal of Natural Products, 75**,** 2274-2274.

PECYNA, M. & BIZUKOJC, M. (2011) Lovastatin biosynthesis by Aspergillus terreus with the simultaneous use of lactose and glycerol in a discontinuous fed-batch culture. Journal of Biotechnology, 151**,** 77-86.

PETER BOSCH, O. W., ELISABETH JOELSSON, MATS GALBE, GUIDO ZACCHI (2010) Impact of dual temperature profile in dilute acid hydrolysis of spruce for ethanol production. Biotechnology for Biofuels**,** 15.

PETERSON, R., GRINYER, J. & NEVALAINEN, H. (2011) Extracellular hydrolase profiles of fungi isolated from koala faeces invite biotechnological interest. Mycological Progress, 10**,** 207-218.

PROSAD, B. S., PAL, S., CHOWDHURY, S., GHORAI, S. & KHOWALA, S. (2011) Evidence of an Alternative Route of Cellobiase Secretion in the Presence of Brefeldin A in the Filamentous Fungus *Termitomyces clypeatus*. Journal of Microbiology and Biotechnology, 21**,** 411-420.

PUCHART, V., KATAPODIS, P. & BIELY, P. (1999) Production of xylanases, mannanases, and pectinases by the themophilic *fungus Thermomyces* lanuginosus. Enzyme Microbial Technology 24**,** 355–361.

PURI, M., ABRAHAM, R. E. & BARROW, C. J. (2012) Biofuel production: Prospects, challenges and feedstock in Australia. Renewable and Sustainable Energy Reviews, 16**,** 6022-6031.

PURKARTHOFER, H., SINNER, M. & STEINER, W. (1993) Cellulase-free xylanase from *Thermomyces lanuginosus*: optimization of production in submerged and solid-state culture. Enzyme. Microb. Technol, 15**,** 677–682.

PURKARTHOFER, H. & STEINER, W. (1995) Induction of endo-β-xylanase in the fungus Thermomyces lanuginosus. Enzyme and Microbial Technology, 17**,** 114-118.

QIANG, G., KA, I. J., SHENG-HUA, Y., YONGJUN, Z., GUOHUA, X., YANFANG, S., ZHIBING, D., XIAO, H., XUE-QIN, X., GANG, Z., GUOXIONG, P., ZHIBING, L., WEI, H., BING, W., WEIGUO, F., SIBAO, W., YI, Z., LI-JUN, M., RAYMOND, J. S. L., GUO-PING, Z., YAN, P., MING-GUANG, F., YUXIAN, X. & CHENGSHU, W. (2011) Genome Sequencing and Comparative Transcriptomics of the Model Entomopathogenic Fungi *Metarhizium anisopliae* and *M. acridum*. PLoS Genetics, 7**,** 1-18.

QING, Q. & WYMAN, C. (2011) Hydrolysis of different chain length xylooliogmers by cellulase and hemicellulase. Bioresource Technology 102**,** 1359-1366.

RAO, M. B., TANKSALE, A. M., GHATGE, M. S. & DESHPANDE, V. V. (1998) Molecular and Biotechnological Aspects Molecular and Biotechnological Aspects. Microbiology and Molecular Biology Reviews, 62**,** 597-635.

REIS, F. S., BARROS, L., MARTINS, A. & FERREIRA, I. C. F. R. (2012) Chemical composition and nutritional value of the most widely appreciated cultivated mushrooms: An inter-species comparative study. Food and Chemical Toxicology, 50**,** 191-197.

REP, M. & KISTLER, H. C. (2010) The genomic organization of plant pathogenicity in *Fusarium* species. Current Opinion in Plant Biology, 13**,** 420-426.

ROJAS, C. & MANSUR, E. (1995) Ecuador: Informaciones Generales Sobre Productos Non Madereros en Ecuador. In Memoria, Consulta De Expertos Sobre Productos Forestales. Madereros Para America Latina Y el Caribe**,** 208-223.

RONCERO, M. B., TORRES, A. L., COLOM, J. F. & VIDAL, T. (2000) Effects of xylanase treatment on fibre morphology in totally chlorine free bleaching (TCF) of Eucalyptus pulp. Process Biochemistry, 36**,** 45-50.

ROPARS, J., CRUAUD, C., LACOSTE, S. & DUPONT, J. (2012) A taxonomic and ecological overview of cheese fungi. International Journal of Food Microbiology, 155**,** 199-210.

RUEL, K. & JOSELEAU, J. P. (1991) Involvement of an extracellular glucan sheath during degradation of populous wood by *Phanerochaete chrysporium*. Appl. Environ. Microbiol., 57**,** 374-384.

SALAR, R. K. & ANEJA, K. R. (2007) Thermophilic Fungi: Taxonomy and Biogeography. Journal of Agricultural Technology, 3**,** 77-107.

SAMBROOK, J. & RUSSELL, D. W. 2006. *The condensed protocols from molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY.

SAMOLSKI, I., DE LUIS, A., VIZCAANO, J. A., MONTE, E. & SUAREZ, M. B. (2009) Gene expression analysis of the biocontrol fungus *Trichoderma harzianum* in the presence of tomato plants, chitin, or glucose using a high-density oligonucleotide microarray. BMC microbiology, 9**,** 217-217.

SAXENA, R., GHOSH, P., GUPTA, R., DAVIDSON, W. S., BRADOO, S. & GULATI, R. (1999) Microbial lipases: potential biocatalysts for the future industry. Current Science, 77**,** 101-115.

SCHOLZ, R., HANSEN, W. & THURMAN, R. G. (1973) Interaction of Mixed-Function Oxidation with Biosynthetic Processes. European Journal of Biochemistry, 38**,** 64-72.

SCOTT, D., NANCY, Y. L., DARREN, P., GORDON, R., MICHAEL, S., SIMON, K. C., RODERICK, D. T., INANC BIROL, ROBERT, A. H., MARTIN, H., ELAINE, M., MARCO, A. M., RICHARD, C. H., JÖRG, B. & COLETTE, B. S. J. M. J. (2009) *De novo* genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. Genome Biology, 10.

SEIBOTH, B., HARTL, L., PAIL, M. & KUBICEK, C. P. (2003) D-xylose Metabolism in *Hypocrea jecorina*: Loss of the Xylitol Dehydrogenase Step Can Be Partially Compensated for by lad1-Encoded l-Arabinitol-4-Dehydrogenase. Eukaryotic Cell 2**,** 867-875.

SHARMA, M., CHADHA, B. S., KAUR, M., GHATORA, S. K. & SAINI, H. S. (2008) Molecular characterization of multiple xylanase producin thermophilic thermotolerant

fungi isolated from composting materials. The Society for Applied Microbiology, 46**,** 526–535.

SHRADDHA, S., SHEKHER, R., SEHGAL, S., KAMTHANIA, M. & KUMAR, A. (2011) Laccase:Microbial Sources, Production, Purification, and Potential Biotechnological Applications. Enzyme Research, 2011.

SINGER, G. A. & HICKEY, D. A. (2003) Thermophilic prokaryotes have characteristic patterns of codon usage, amino acid composition and nucleotide content. Gene
317**,** 39-47.

SINGH, S., MADLALA, A. M. & PRIOR, B. A. (2003) *Thermomyces lanuginosus*: properties of strains and their hemicellulases. FEMS Microbiology Reviews, 27**,** 3-16.

SINGH, S., PILLAY, B., DILSOOK, V. & PRIOR, B. A. (2000b) Production and properties of hemicellulases by a *Thermomyces lanuginosus* strain. Journal of Applied Microbiology, 88**,** 975-982.

SINGH, S., REDDY, P., HAARHOFF, J., BIELY, P., JANSE, B., PILLAY, B., PILLAY, D. & PRIOR , B. A. (2000a) Relatedness of *Thermomyces lanuginosus* strains producing a thermostable xylanase. Journal of Biotechnology, 81**,** 119–128.

SOMERVILLE, C. (2007) Biofuels. CURRENT BIOLOGY**,** R115-R119.

SONENBERG, N. & HINNEBUSCH, A. G. (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell, 136**,** 731-745.

SPENCER-MARTINS, I. (1994) Transport of sugars in yeasts: Implications in the fermentation of lignocellulosic materials. Bioresource Technology, 50**,** 51-57.

STAJICH, J., HARRIS, T., BRUNK, B., BRESTELLI, J., FISCHER, S., HARB, O., KISSINGER, J., LI, W., NAYAK, V. & PINNEY, D. (2012) FungiDB: an integrated functional genomics database for fungi. Nucleic Acids Research, 40**,** D675–D681.

STAJICH, J. E., BERBEE, M. L., BLACKWELL, M., HIBBETT, D. S., JAMES, T. Y., SPATAFORA, J. W. & TAYLOR, J. W. (2009) The Fungi. CURRENT BIOLOGY, 19**,** R840-R845.

STASZCZAK, M. (2008) The role of the ubiquitin-proteasome system in the response of the ligninolytic fungus *Trametes versicolor* to nitrogen deprivation Fungal Genetics Biolilogy, 45**,** 328-337.

STEPHENS, D. E., RUMBOLD, K., PERMAUL, K., PRIOR, B. A. & SINGH, S. (2007) Directed evolution of the thermostable xylanase from Thermomyces lanuginosus. Journal of Biotechnology, 127**,** 348-354.

STERNBACH, G. & VARON, J. (1992) Alexander fleming: The spectrum of penicillin. The Journal of Emergency Medicine, 10**,** 89-91.

STUKENBROCK, E., H., JORGENSEN, F., G., ZALA, F., G., HANSEN, T. T., MCDONALD, A. & SCHIERUP, M. K. (2010) Whole-Genome and Chromosome Evolution Associated with Host Adaptation and Speciation of the Wheat Pathogen *Mycosphaerella graminicola*. PLoS Genetics, 6**,** 1-14.

SUKHARNIKOV, L. O., CANTWELL, B. J., PODAR, M. & ZHULIN, I. B. (2011) Cellulases: ambiguous nonhomologous enzymes in a genomic perspective. Trends in Biotechnology, 29**,** 473-479.

SUTHERS, P. F., DASIKA, M. S., KUMAR, V. S., DENISOV, G., GLASS, J. I. & MARANAS, C. D. (2009) A Genome-Scale Metabolic Reconstruction of *Mycoplasma genitalium* PS189. PLoS Comput Biol, 5**,** e1000285.

SVASTITS-DÜCSŐ, L., NGUYEN, Q. D., LEFLER, D. D. & REZESSY-SZABÓ, J. M. (2009) Effects of galactomannan as carbon source on production of α-galactosidase by Thermomyces lanuginosus: Fermentation, purification and partial characterisation. Enzyme and Microbial Technology, 45**,** 367-371.

TAHERZADEH, M. J. & KARIMI, K. (2008) Pretreatment of Lignocellulosic Wastes to Improve Ethanol and Biogas Production: A Review. International Journal of Molecular Sciences, 9**,** 1621-1651.

TANG, J.-C., KANAMORI, T., INOURE, Y., YASUTA, T., YOSHIDA, S. & KATAYAMA, A. (2004) Changes in the microbial community structure during thermophilic composting of manure as detected by the quinone profile method. Process Biochemistry, 39**,** 1999–2006.

TANG, J.-C. & KATAYAMA, A. (2005) Relating quinone profile to the aerobic biodegradation in thermophilic composting processes of cattle manure with various bulking agents. World Journal of Microbiology and Biotechnology, 21**,** 1249–1254.

TATUSOV, R. L., GALPERIN, M. Y., NATALE, D. A. & KOONIN, E. V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research, 28**,** 33-36.

TEH, K.-Y. & LUTZ, A. E. (2010) Thermodynamic analysis of fermentation and anaerobic growth of baker's yeast for ethanol production. Journal of Biotechnology, 147**,** 80-87.

TIQUIA, S. M., WAN, J. H. C. & TAM, N. F. Y. (2002) Microbial population dynamics andenzyme activities during composting. . Compost Science and Utilization., 10**,** 150-161.

TOMA, F. M., ISMAEL, H. M. & Q, A. N. (2013) Survey and Identification of Mushrooms in Erbil Governorate. Research Journal of Environmental and Earth Sciences 5**,** 262-266.

TRENT, J. D., GABRIELSEN, M., JENSEN, B., NEUHARD, J. & OLSEN, J. (1994) Acquired thermotolerance and heat shock proteins in thermophiles from the three phylogenetic domains. Journal of Bacteriology, 176**,** 6148–6152.

VAN DIJCK, P., COLAVIZZA, D., SMET, P. & THEVELEIN, J. M. (1995) Differential importance of trehalose in stress resistance in fermenting and nonfermenting Saccharomyces cerevisiae cells. Applied and Environmental Microbiology, 61**,** 109-115.

VENTER, J. C., ADAMS, M. D., MYERS, E. W., LI, P. W., MURAL, R. J., SUTTON, G. G., SMITH, H. O., YANDELL, M., EVANS, C. A. & HOLT, R. A. (2001) The sequence of the human genome. science, 291**,** 1304-1351.

VENTER, J. C., REMINGTON, K., HEIDELBERG, J. F., HALPERN, A. L., RUSCH, D., EISEN, J. A., WU, D., PAULSEN, I., NELSON, K. E., NELSON, W., FOUTS, D. E., LEVY, S., KNAP, A. H., LOMAS, M. W., NEALSON, K., WHITE, O., PETERSON, J., HOFFMAN, J., PARSONS, R., BADEN-TILLSON, H., PFANNKOCH, C., ROGERS, Y.-H. & SMITH, H. O. (2004) Environmental Genome Shotgun Sequencing of the Sargasso Sea. Science, 304**,** 66-74.

VIEILLE, C. & ZEIKUS, G. J. (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. . Microbiology and Molecular Biology Reviews, 65**,** 1-43.

VILGALYS, R. J. & JOHNSON, J. L. (1987) Extensive genetic divergence associated with speciation in filamentous fungi. Proceedings of the National Academy of Sciences, 84**,** 2355-2358.

VON EIFF, C., MCNAMARA, P., BECKER, K., BATES, D., LEI, X.-H., ZIMAN, M., BOCHNER, B. R., PETERS, G. & PROCTOR, R. A. (2006) Phenotype microarray profiling of Staphylococcus aureus menD and hemB mutants with the small-colony-variant phenotype. Journal of bacteriology, 188**,** 687-693.

WAINWRIGHT, M. (2008) Some highlights in the history of fungi in medicine – A personal journey. Fungal Biology Reviews, 22**,** 97-102.

WANG, B., GUO, G., WANG, C., LIN, Y., WANG, X., ZHAO, M., GUO, Y., HE, M., ZHANG, Y. & PAN, L. (2010) Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. Nucleic Acids Research, 38**,** 5075-5087.

WARD, O. P. (2012) Production of recombinant proteins by filamentous fungi. Biotechnology Advances, 30**,** 1119-1139.

WEI, Z. & LI, H. (2007) A Markov random field model for network-based analysis of genomic data. Bioinformatics, 23**,** 1537-1544.

WIATER, A., JANUSZ, S. & MALGORZATA, P. (2008) Mutanase Induction in *Trichoderma harzianum* by Cell Wall of *Laetiporus sulphureus* and its Application for Mutan Removal from Oral Biofilms. Journal of Microbiology and Biotechnology, 8**,** 1335–1341.

WIEMKEN, A. (1990) Trehalose in yeast, stress protectant rather than reserve carbohydrate. Antonie van Leeuwenhoek, 58**,** 209-217.

WOOD, Z. A., POOLE, L. B. & KARPLUS, P. A. (2003) Peroxiredoxin Evolution and the Regulation of Hydrogen Peroxide Signaling. Science, 300**,** 650-653.

WRIGHT, C., KAFKEWITZ, D. & SOMBERG, W. E. (1983) Eucaryote thermophily: role of lipids in the growth of *Talaromyces thermophilus*. Journal of Bacteriology, 156**,** 493-497.

WU, D., RAYMOND, J., WU, M., CHATTERJI, S., REN, Q., GRAHAM, J. E., BRYANT, D. A., ROBB, F., COLMAN, A., TALLON, L. J., BADGER, J. H., MADUPU, R., WARD, N. L. & EISEN, J. A. (2009) Complete Genome Sequence of the Aerobic CO-Oxidizing Thermophile *Thermomicrobium roseum*. PLoS One, 4**,** 1-6.

WU, J., LIU, B., CHENG, F., RAMCHIARY, N., CHOI, S.-R., LIM, Y. P. & WANG, X. (2012) Sequencing of chloroplast genome using whole cellular DNA and Solexa sequencing technology. Frontier in Plant Sciences, 3**,** 1.

XU, X., NICHOLSON, P. & RITIENI, A. (2007) Effects of fungal interactions among Fusarium head blight pathogens on disease development and mycotoxin accumulation. International Journal of Food Microbiology, 119**,** 67-71.

XUAN, J., YU, Y., QING, T., GUO, L. & SHI, L. (2012) Next-generation sequencing in the clinic: Promises and challenges. Cancer Letters, In Press.

YAMANE, Y.-I., FUJITA, J., IZUWA, S., FUKUCHI, K., SHIMIZU, R.-I., HIYOSHI, A., FUKUDA, H., MIKAMI, S., KIZAKI, Y. & WAKABAYASHI, S. (2002) Properties of cellulose-degrading enzymes from Aspergillus oryzae and their contribution to material

utilization and alcohol yield in sake mash fermentation. Journal of Bioscience and Bioengineering, 93**,** 479-484.

YU, H., ZENG, G., HUANG, H., XI, X., WANG, R., HUANG, D., HUANG, G. & LI, J. (2007) Microbial community succession and lignocellulose degradation during agricultural waste composting. Biodegradation, 18**,** 793–802.

ZHANG, J., CHIODINI, R., BADR, A. & ZHANG, G. (2011) The impact of next-generation sequencing on genomics. Journal of Genetics and Genomics, 38**,** 95-109.

ZHANG, R., LI, X. & FADEL, J. G. (2002) Oyster mushroom cultivation with rice and wheat straw. Bioresource Technology, 82**,** 277-284.

ZHAO, Z., LIU, H., WANG, C. & XU, J.-R. (2013) Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC genomics, 14**,** 274.

ZHENG, W. & KATHARIOU, S. (1995) Differentiation of epidemic-associated strains of *Listeria monocytogenes* by restriction fragment length polymorphism in a gene region essential for growth at low temperatures (4 degrees C). Applied Environmental Microbiology, 61**,** 4310-4.

ZUCCARO, A., LAHRMANN, U., GALDENER, U., LANGEN, G., PFIFFI, S., BIEDENKOPF, D., WONG, P., SAMANS, B., GRIMM, C., BASIEWICZ, M., MURAT, C., MARTIN, F. & KOGEL, K. H. Endophytic life strategies decoded by genome and transcriptome analyses of the mutualistic root symbiont *Piriformospora indica*. PLoS pathogens, 7**,** e1002290.

6. APPENDIX:

Table S1: Paralog families found in *T. lanuginosus*

Paralog family	Paralog sub-family		Size
SAM-dependent methylases			84
F-box domain proteins			27
Transporters		**112**	
	Sugar and organic MFS transporters		50
	Mitochondria carriers		31
	ABC transporters		21
	Aminoacid permeases		10
Dehydrogenases		**95**	
	Short chain dehydrogenases		37
	Aldehyde dehydrogenases		23
	Acyl-CoA dehydrogenases		7
	Alcohol dehydrogenases		28
WD-40 repeats proteins			48
Helicases		**37**	
	DAED box helicases		19
	ATP-dependent helicases		18
Cytochrome p450			10
Acyl-CoA synthases			15
TPR repeat proteins			16
Protein folding		**35**	
	Ubiquitin ligases		16
	CCT chaperons		11
	Prolyl isomerases		8

Table S2: Carbohydrate related proteins found in *T. lanuginosus*

Sequence ID	Query-AA	Subject-AA	Organism	CAZy Families
TL-CAZY 207	542	858	*Hypoxylon haematostroma CBS 255.63*	CBM1
TL-CAZY 143	324	1862	*Teredinibacter turnerae T7901*	CBM10|CBM2
TL-CAZY 50	309	290	*Streptosporangium roseum DSM 43021*	CBM12
TL-CAZY 192	387	1090	*Streptosporangium roseum DSM 43021*	CBM12
TL-CAZY 10	710	777	*Frankia alni ACN14a*	CBM13
TL-CAZY 91	1575	640	*Streptomyces coelicolor A3(2)*	CBM13
TL-CAZY 106	670	634	*Frankia alni ACN14a*	CBM13
TL-CAZY 118	455	554	*Stigmatella aurantiaca DW4/3-1*	CBM13
TL-CAZY 196	642	910	*Catenulispora acidiphila DSM 44928*	CBM13
TL-CAZY 204	479	657	*Leptosphaeria maculans v23.1.3*	CBM18
TL-CAZY 3	389	471	*Amycolatopsis mediterranei U32*	CBM2
TL-CAZY 139	882	471	*Amycolatopsis mediterranei U32*	CBM2
TL-CAZY 181	1124	552	*Micromonospora aurantiaca ATCC 27029*	CBM2
TL-CAZY 182	1124	552	*Micromonospora sp. L5*	CBM2
TL-CAZY 152	1195	688	*Xenopus (Silurana) tropicalis N6 (Nigerian 6th generation inbred)*	CBM20
TL-CAZY 172	617	630	*Thermomyces lanuginosus*	CBM20|GH15
TL-CAZY 159	591	559	*Aspergillus niger CBS 513.88*	CBM21
TL-CAZY 43	475	446	*Streptosporangium roseum DSM 43021*	CBM32
TL-CAZY 219	124	365	*Spodoptera litura NPV G2*	CBM33
TL-CAZY 33	475	648	*Aspergillus niger CBS 513.88*	CBM48
TL-CAZY 96	484	471	*Aspergillus niger CBS 513.88*	CBM48
TL-CAZY 209	411	495	*Ostreococcus lucimarinus CCE9901*	CBM48
TL-CAZY 218	638	1026	*Haliangium ochraceum DSM 14365*	CBM48
TL-CAZY 88	737	699	*Emericella nidulans*	CBM48|GH13
TL-CAZY 183	733	615	*Paenibacillus sp. JDR-2*	CBM50
TL-CAZY 113	199	262	*Podospora anserina S mat+*	CBM52
TL-CAZY 157	840	940	*Jatropha curcas Palawan*	CBM57
TL-CAZY 155	795	541	*Saccharopolyspora erythraea NRRL 2338*	CBM6
TL-CAZY 31	288	301	*Aspergillus niger CBS 513.88*	CE1
TL-CAZY 97	380	694	*Micromonas sp. RCC299*	CE1
TL-CAZY 23	723	645	*Pyrococcus horikoshii*	CE10
TL-CAZY 124	345	355	*Schizosaccharomyces pombe 972h-*	CE10
TL-CAZY 144	539	502	*Bacillus subtilis NRRL B8079*	CE10
TL-CAZY 195	759	592	*Schizosaccharomyces pombe 972h-*	CE10
TL-CAZY 58	399	1058	*Ostreococcus tauri OTTH0595*	CE11
TL-CAZY 104	336	726	*Podospora anserina S mat+*	CE12
TL-CAZY 78	373	NA	*uncultured soil bacterium*	CE3
TL-CAZY 154	308	320	*Aspergillus niger CBS 513.88*	CE4
TL-CAZY 200	347	472	*Aspergillus niger CBS 513.88*	CE4|CBM18

TL-CAZY 16	460	430	*Aspergillus oryzae RIB40*	CE9
TL-CAZY 180	372	743	*Acidiphilium cryptum JF-5*	CE9
TL-CAZY 149	1433	736	*Deinococcus geothermalis DSM 11300*	GH1
TL-CAZY 178	479	502	*Talaromyces emersonii*	GH1
TL-CAZY 133	313	556	*Xanthomonas campestris pv. campestris str. 8004*	GH103
TL-CAZY 83	355	380	*Beutenbergia cavernae DSM 12333*	GH109
TL-CAZY 198	401	384	*Spirosoma linguale DSM 74*	GH109
TL-CAZY 41	253	239	*Paecilomyces sp. J18 / Paecilomyces thermophila J18*	GH11
TL-CAZY 8	331	347	*Penicillium chrysogenum Wisconsin 54-1255*	GH12
TL-CAZY 112	477	629	*Sinorhizobium meliloti*	GH12
TL-CAZY 45	1527	1547	*Penicillium chrysogenum Wisconsin 54-1255*	GH13
TL-CAZY 67	601	575	*Geobacillus sp. Y4.1MC1*	GH13
TL-CAZY 103	589	775	*Penicillium chrysogenum Wisconsin 54-1255*	GH13
TL-CAZY 127	494	506	*Thermomyces lanuginosus*	GH13
TL-CAZY 140	603	610	*Penicillium minioluteum HI-4*	GH13
TL-CAZY 14	608	591	*Penicillium chrysogenum Wisconsin 54-1255*	GH16
TL-CAZY 46	349	483	*Aspergillus nidulans FGSC A4*	GH16
TL-CAZY 70	396	417	*Aspergillus nidulans FGSC A4*	GH16
TL-CAZY 90	634	NA	*Paecilomyces sp. J18 / Paecilomyces thermophila J18*	GH16
TL-CAZY 164	401	422	*Aspergillus oryzae RIB40*	GH16
TL-CAZY 173	658	670	*Aspergillus oryzae RIB40*	GH16
TL-CAZY 61	432	446	*Penicillium chrysogenum Wisconsin 54-1255*	GH16|CBM18
TL-CAZY 69	302	317	*Aspergillus oryzae RIB40*	GH17
TL-CAZY 77	527	537	*Coccidioides posadasii*	GH17
TL-CAZY 85	645	685	*Aspergillus oryzae RIB40*	GH17
TL-CAZY 163	447	463	*Aspergillus oryzae RIB40*	GH17
TL-CAZY 7	348	600	*Clostridium cellulolyticum H10 ATCC 35319*	GH18
TL-CAZY 18	354	980	*Aspergillus nidulans FGSC A4*	GH18
TL-CAZY 52	322	785	*Leptosphaeria maculans v23.1.3*	GH18
TL-CAZY 65	1060	255	*Pantoea dispersa*	GH18
TL-CAZY 108	225	722	*Clostridium difficile CD196*	GH18
TL-CAZY 109	225	722	*Clostridium difficile R20291*	GH18
TL-CAZY 110	225	725	*Clostridium difficile 630*	GH18
TL-CAZY 137	436	455	*Thermomyces lanuginosus SY2*	GH18
TL-CAZY 142	409	446	*Aspergillus fumigatus ATCC 13073*	GH18
TL-CAZY 208	434	378	*Penicillium chrysogenum Wisconsin 54-1255*	GH18
TL-CAZY 214	670	938	*Aspergillus nidulans FGSC A4*	GH18
TL-CAZY 114	383	1705	*Ostreococcus tauri OTTH0595*	GH2
TL-CAZY 134	851	867	*Podospora anserina S mat+*	GH2
TL-CAZY 22	854	1388	*Aspergillus oryzae RIB40*	GH28
TL-CAZY 212	389	393	*Aspergillus nidulans FGSC A4*	GH28

TL-CAZY 55	914	974	*Aspergillus oryzae RIB40*	GH3
TL-CAZY 63	839	852	*Aspergillus oryzae RIB40*	GH3
TL-CAZY 72	845	869	*Aspergillus oryzae RIB40*	GH3
TL-CAZY 121	829	841	*Penicillium chrysogenum Wisconsin 54-1255*	GH3
TL-CAZY 220	858	870	*Talaromyces emersonii*	GH3
TL-CAZY 221	777	792	*Aspergillus oryzae RIB40*	GH3
TL-CAZY 54	756	771	*Aspergillus oryzae RIB40*	GH31
TL-CAZY 75	960	979	*Aspergillus oryzae RIB40*	GH31
TL-CAZY 117	900	913	*Thermomyces lanuginosus*	GH31
TL-CAZY 222	674	681	*Penicillium chrysogenum Wisconsin 54-1255*	GH31
TL-CAZY 170	593	1100	*Aspergillus oryzae RIB40*	GH32
TL-CAZY 188	606	1103	*Leishmania major strain Friedlin*	GH32
TL-CAZY 202	1012	1021	*Talaromyces emersonii*	GH35
TL-CAZY 125	750	763	*Aspergillus nidulans FGSC A4*	GH36
TL-CAZY 171	948	958	*Penicillium chrysogenum Wisconsin 54-1255*	GH36
TL-CAZY 19	777	744	*Aspergillus oryzae RIB40*	GH37
TL-CAZY 73	1093	1101	*Aspergillus fumigatus YJ-407*	GH38
TL-CAZY 189	3856	2318	*uncultured marine bacterium 1n22*	GH4
TL-CAZY 25	338	348	*Penicillium herquei IFO 4674*	GH43
TL-CAZY 84	1105	1096	*Aspergillus niger*	GH47
TL-CAZY 119	913	864	*Aspergillus oryzae RIB40*	GH47
TL-CAZY 131	511	523	*Aspergillus oryzae RIB40*	GH47
TL-CAZY 132	511	524	*Penicillium citrinum*	GH47
TL-CAZY 201	587	608	*Aspergillus oryzae RIB40*	GH47
TL-CAZY 205	631	604	*Aspergillus fumigatus YJ-407*	GH47
TL-CAZY 206	631	604	*Aspergillus fumigatus YJ-407*	GH47
TL-CAZY 136	861	563	*Aspergillus oryzae RIB40*	GH5
TL-CAZY 151	760	762	*Aspergillus niger CBS 513.88*	GH5
TL-CAZY 185	419	430	*Penicillium chrysogenum Wisconsin 54-1255*	GH5
TL-CAZY 190	469	482	*Penicillium chrysogenum Wisconsin 54-1255*	GH5
TL-CAZY 187	931	947	*Aspergillus fumigatus*	GH55
TL-CAZY 17	272	291	*Aspergillus niger CBS 513.88*	GH61
TL-CAZY 38	369	441	*Aspergillus oryzae RIB40*	GH61
TL-CAZY 186	1388	835	*Aspergillus fumigatus YJ-407*	GH63
TL-CAZY 62	1075	1079	*Talaromyces emersonii*	GH65
TL-CAZY 94	471	563	*Aspergillus nidulans FGSC A4*	GH72
TL-CAZY 101	443	471	*Penicillium chrysogenum Wisconsin 54-1255*	GH72
TL-CAZY 57	535	551	*Aspergillus oryzae RIB40*	GH72|CBM43
TL-CAZY 4	580	1318	*Aspergillus nidulans FGSC A4*	GH75
TL-CAZY 123	467	469	*Aspergillus niger CBS 513.88*	GH76
TL-CAZY 60	966	749	*Aspergillus oryzae RIB40*	GH81
TL-CAZY 165	491	787	*Schistosoma mansoni*	GH84
TL-CAZY 216	569	1064	*Trypanosoma brucei gambiense DAL972 Dal*	GH85

			927 clone 1 (MHOM/CI/86/DAL972)	
TL-CAZY 168	763	774	*Aspergillus oryzae RIB40*	GH92
TL-CAZY 100	874	611	*Schistosoma mansoni*	GH99
TL-CAZY 34	248	314	*Aspergillus nidulans FGSC A4*	GT1
TL-CAZY 35	188	217	*Aspergillus oryzae RIB40*	GT1
TL-CAZY 64	1358	1457	*Aspergillus oryzae RIB40*	GT1
TL-CAZY 98	158	470	*Pueraria montana var. lobata*	GT1
TL-CAZY 215	371	868	*Aspergillus nidulans FGSC A4*	GT1
TL-CAZY 37	396	410	*Aspergillus oryzae RIB40*	GT15
TL-CAZY 66	402	415	*Aspergillus oryzae RIB40*	GT15
TL-CAZY 89	496	515	*Aspergillus niger CBS 513.88*	GT15
TL-CAZY 12	1840	1870	*Aspergillus oryzae RIB40*	GT2
TL-CAZY 13	1840	1870	*Aspergillus oryzae*	GT2
TL-CAZY 30	1736	1773	*Aspergillus oryzae RIB40*	GT2
TL-CAZY 51	699	761	*Coccidioides immitis*	GT2
TL-CAZY 53	752	766	*Penicillium chrysogenum Wisconsin 54-1255*	GT2
TL-CAZY 56	2363	1350	*Bacteroides thetaiotaomicron VPI-5482*	GT2
TL-CAZY 79	398	334	*Lotus japonicus GIFU*	GT2
TL-CAZY 80	398	334	*Lotus japonicus GIFU*	GT2
TL-CAZY 81	398	334	*Lotus japonicus GIFU*	GT2
TL-CAZY 92	432	727	*Cyanothece sp. PCC 7822*	GT2
TL-CAZY 93	419	428	*Aspergillus oryzae RIB40*	GT2
TL-CAZY 99	468	523	*Chitinophaga pinensis DSM 2588*	GT2
TL-CAZY 129	802	798	*Thermobaculum terrenum ATCC BAA-798*	GT2
TL-CAZY 130	906	924	*Aspergillus fumigatus*	GT2
TL-CAZY 135	921	928	*Aspergillus niger CBS 513.88*	GT2
TL-CAZY 147	584	1078	*Thauera sp. MZ1T*	GT2
TL-CAZY 166	598	1013	*Bacillus pseudofirmus OF4*	GT2
TL-CAZY 176	245	258	*Aspergillus oryzae RIB40*	GT2
TL-CAZY 191	830	820	*Thermomonospora curvata DSM 43183*	GT2
TL-CAZY 194	1562	2289	*Pinctada fucata*	GT2
TL-CAZY 197	1204	1215	*Penicillium chrysogenum Wisconsin 54-1255*	GT2
TL-CAZY 211	1174	1024	*Aspergillus oryzae RIB40*	GT2
TL-CAZY 6	935	941	*Aspergillus oryzae RIB40*	GT20
TL-CAZY 128	526	522	*Aspergillus oryzae RIB40*	GT20
TL-CAZY 141	1126	494	*Penicillium chrysogenum Wisconsin 54-1255*	GT20
TL-CAZY 158	524	493	*Aspergillus niger CBS 513.88*	GT20
TL-CAZY 213	946	956	*Penicillium chrysogenum Wisconsin 54-1255*	GT20
TL-CAZY 36	2023	618	*Aspergillus nidulans FGSC A4*	GT21
TL-CAZY 32	552	556	*Aspergillus oryzae RIB40*	GT22
TL-CAZY 111	802	725	*Penicillium chrysogenum Wisconsin 54-1255*	GT22
TL-CAZY 169	610	1001	*Aspergillus nidulans FGSC A4*	GT22
TL-CAZY 203	1484	1500	*Aspergillus oryzae RIB40*	GT24

TL-CAZY 15	400	389	*Neurospora crassa OR74A*	GT25
TL-CAZY 59	241	688	*Aspergillus niger CBS 513.88*	GT25
TL-CAZY 5	399	1873	*Schistosoma mansoni*	GT27|CBM13
TL-CAZY 145	403	1873	*Schistosoma mansoni*	GT27|CBM13
TL-CAZY 174	730	724	*Aspergillus nidulans FGSC A4*	GT3
TL-CAZY 199	497	940	*Neisseria meningitidis alpha275*	GT30
TL-CAZY 116	435	439	*Aspergillus oryzae RIB40*	GT31
TL-CAZY 160	449	1129	*Vitis vinifera*	GT31
TL-CAZY 184	452	1129	*Vitis vinifera*	GT31
TL-CAZY 71	398	792	*Aspergillus nidulans FGSC A4*	GT32
TL-CAZY 86	302	372	*Aspergillus niger CBS 513.88*	GT32
TL-CAZY 87	302	371	*Aspergillus oryzae RIB40*	GT32
TL-CAZY 193	415	413	*Aspergillus niger CBS 513.88*	GT32
TL-CAZY 20	478	462	*Penicillium chrysogenum Wisconsin 54-1255*	GT33
TL-CAZY 74	489	845	*Aspergillus nidulans FGSC A4*	GT34
TL-CAZY 126	563	474	*Aspergillus niger CBS 513.88*	GT34
TL-CAZY 179	323	331	*Aspergillus oryzae RIB40*	GT34
TL-CAZY 138	881	894	*Aspergillus niger CBS 513.88*	GT35
TL-CAZY 24	894	872	*Aspergillus oryzae RIB40*	GT39
TL-CAZY 29	775	784	*Aspergillus niger CBS 513.88*	GT39
TL-CAZY 153	740	754	*Aspergillus awamori*	GT39
TL-CAZY 1	639	707	*Penicillium chrysogenum Wisconsin 54-1255*	GT4
TL-CAZY 42	552	566	*Aspergillus oryzae* RIB40	GT4
TL-CAZY 47	476	493	*Aspergillus oryzae* RIB40	GT4
TL-CAZY 49	1322	667	*Desulfovibrio desulfuricans subsp. desulfuricans str.* ATCC 27774	GT4
TL-CAZY 107	494	501	*Aspergillus nidulans* FGSC A4	GT4
TL-CAZY 115	252	661	*Frankia alni ACN14a*	GT4
TL-CAZY 120	529	874	*Solibacter usitatus Ellin6076*	GT4
TL-CAZY 175	378	647	*Cellulomonas flavigena DSM 20109*	GT4
TL-CAZY 40	1354	NA	*Gallionella capsiferriformans ES-2*	GT41
TL-CAZY 148	1645	1596	*Aspergillus nidulans FGSC A4*	GT41
TL-CAZY 122	1314	1376	*Vitis vinifera*	GT47
TL-CAZY 39	2496	1915	*Coccidioides posadasii SILVEIRA*	GT48
TL-CAZY 95	512	1331	*Dictyostelium discoideum AX4*	GT49
TL-CAZY 2	429	453	*Aspergillus oryzae RIB40*	GT50
TL-CAZY 26	478	885	*Vitis vinifera*	GT50
TL-CAZY 167	1542	1732	*Magnaporthe grisea 70-15*	GT55
TL-CAZY 27	532	604	*Aspergillus nidulans FGSC A4*	GT57
TL-CAZY 28	532	606	*Aspergillus oryzae RIB40*	GT57
TL-CAZY 44	501	517	*Aspergillus oryzae RIB40*	GT57
TL-CAZY 11	424	456	*Neurospora crassa OR74A*	GT58
TL-CAZY 105	637	621	*Aspergillus oryzae RIB40*	GT59

TL-CAZY 9	377	390	*Aspergillus oryzae RIB40*	GT62
TL-CAZY 21	482	490	*Aspergillus oryzae RIB40*	GT62
TL-CAZY 48	457	490	*Aspergillus oryzae RIB40*	GT62
TL-CAZY 102	428	424	*Aspergillus niger CBS 513.88*	GT62
TL-CAZY 82	733	755	*Aspergillus oryzae RIB40*	GT66
TL-CAZY 146	2230	381	*Leishmania braziliensis*	GT67
TL-CAZY 177	444	479	*Aspergillus niger CBS 513.88*	GT76
TL-CAZY 162	844	661	*Dictyostelium discoideum AX3*	GT77
TL-CAZY 156	229	325	*Aspergillus oryzae RIB40*	GT8
TL-CAZY 161	600	724	*Penicillium chrysogenum Wisconsin 54-1255*	GT8
TL-CAZY 210	412	399	*Aspergillus oryzae RIB40*	GT8
TL-CAZY 217	203	316	*Schistosoma mansoni*	GT8
TL-CAZY 76	451	713	*Streptomyces bingchenggensis BCW-1*	GT87
TL-CAZY 68	619	609	*Aspergillus oryzae RIB40*	GT90
TL-CAZY 150	404	832	*Caenorhabditis briggsae AF16*	GT92

Metabolic Pathways

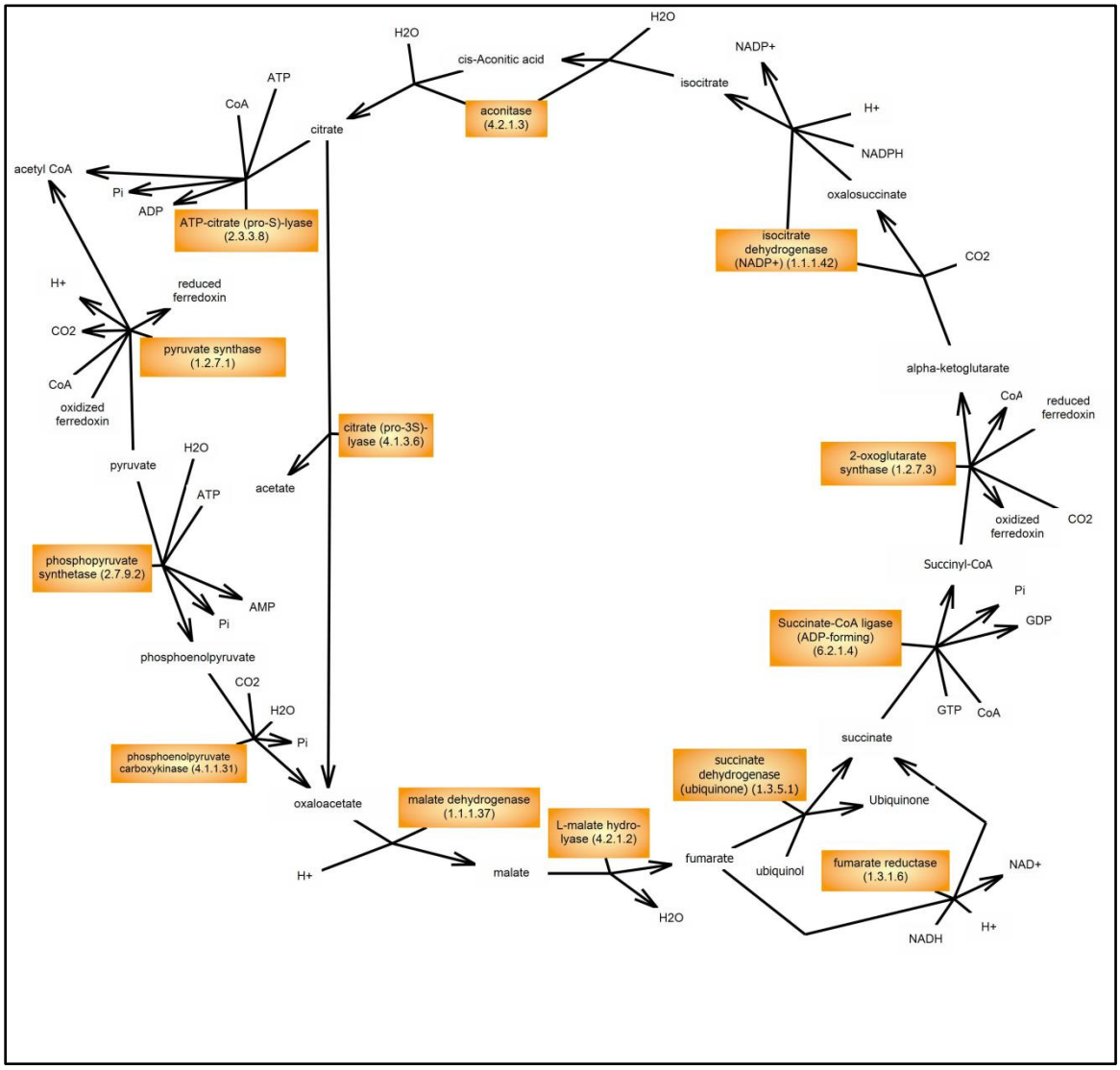

Figure S1: The tricarboxylic acid cycle (TCA) in *T. lanuginosus*.

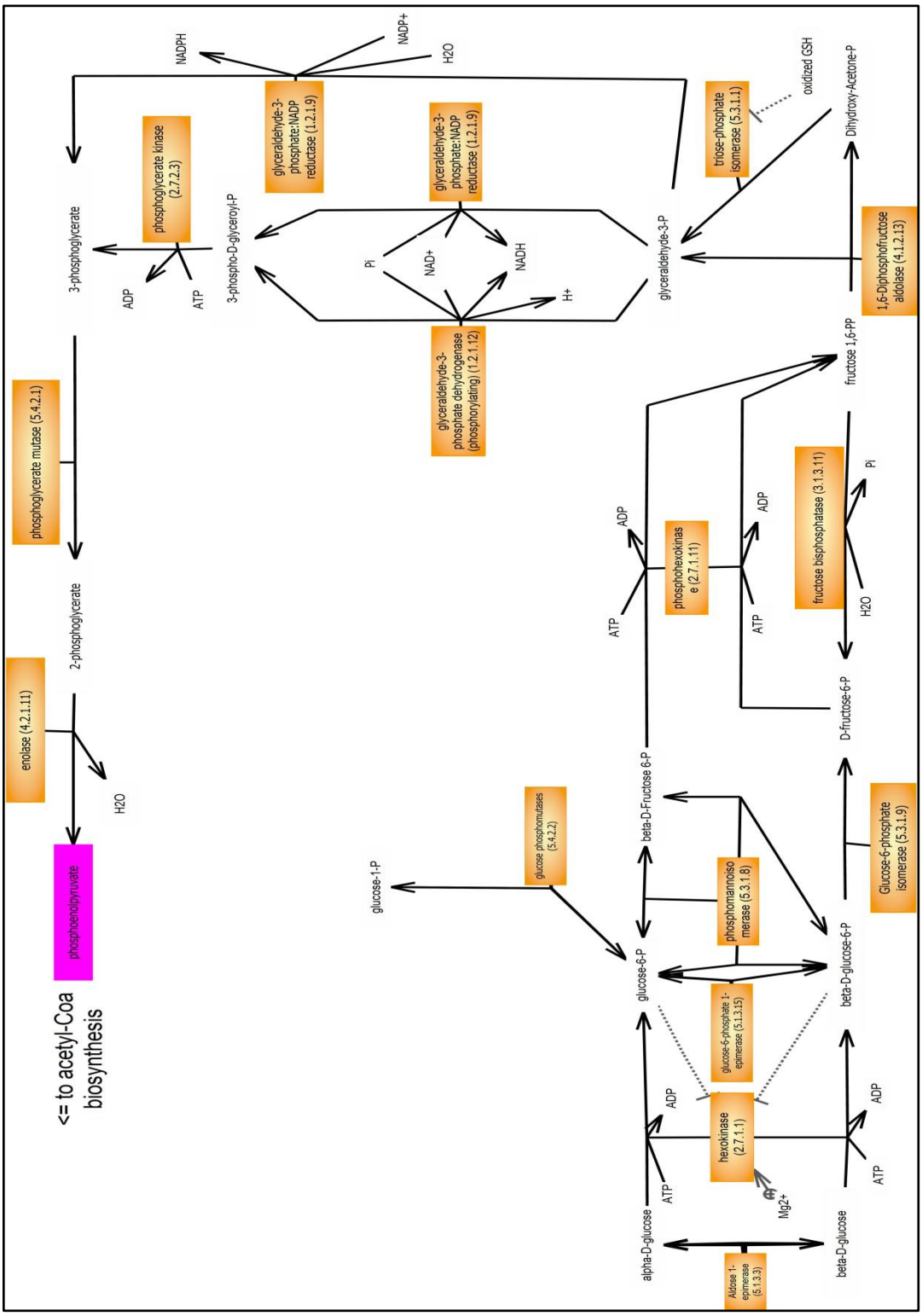

Figure S2: Glycolysis pathway in *T. lanuginosus* showing the enzymatic reactions in the conversion of glucose to acetyl-CoA.

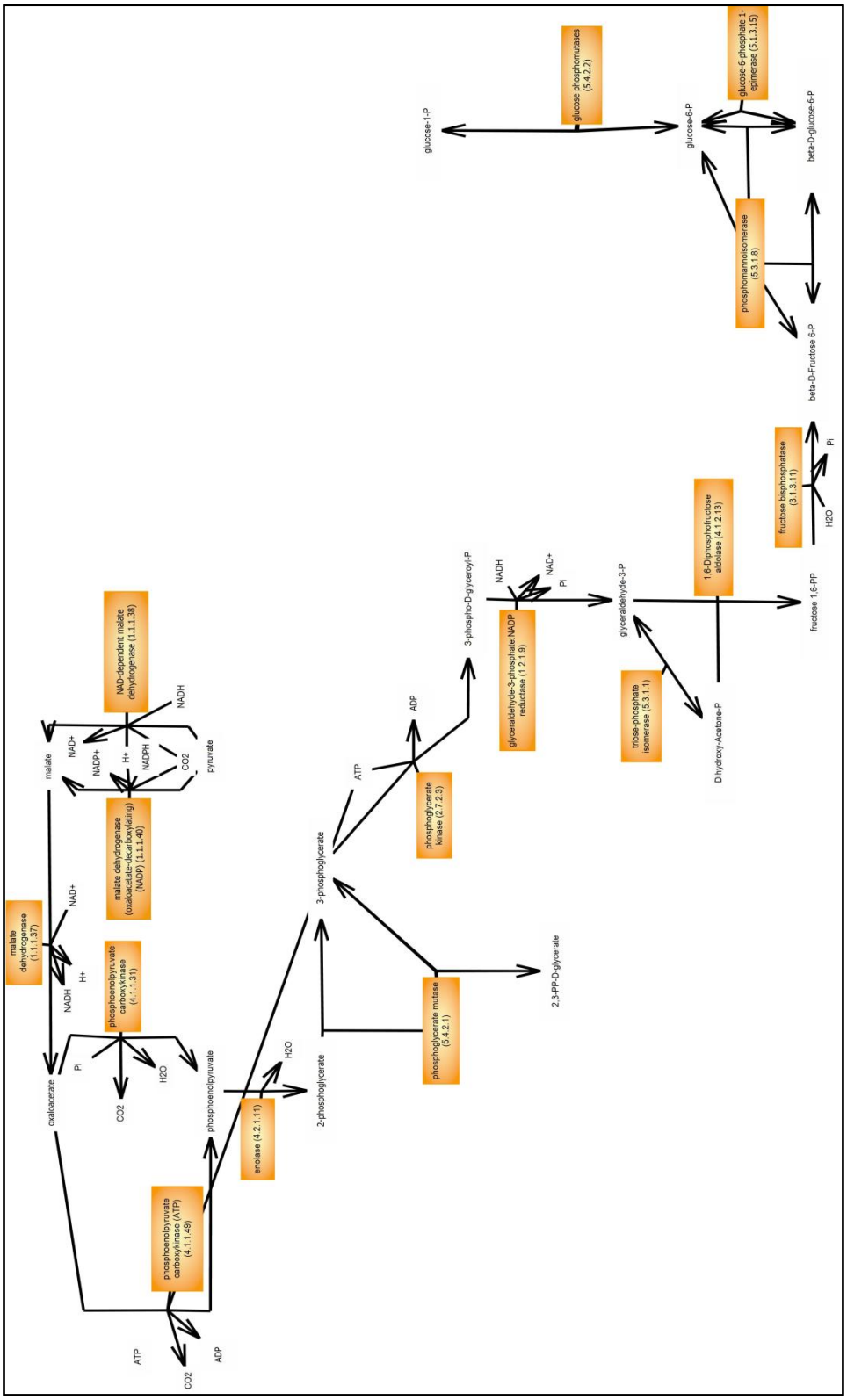

Figure S3: Predicted gluconogenesis pathway in *T. lanuginosus*.

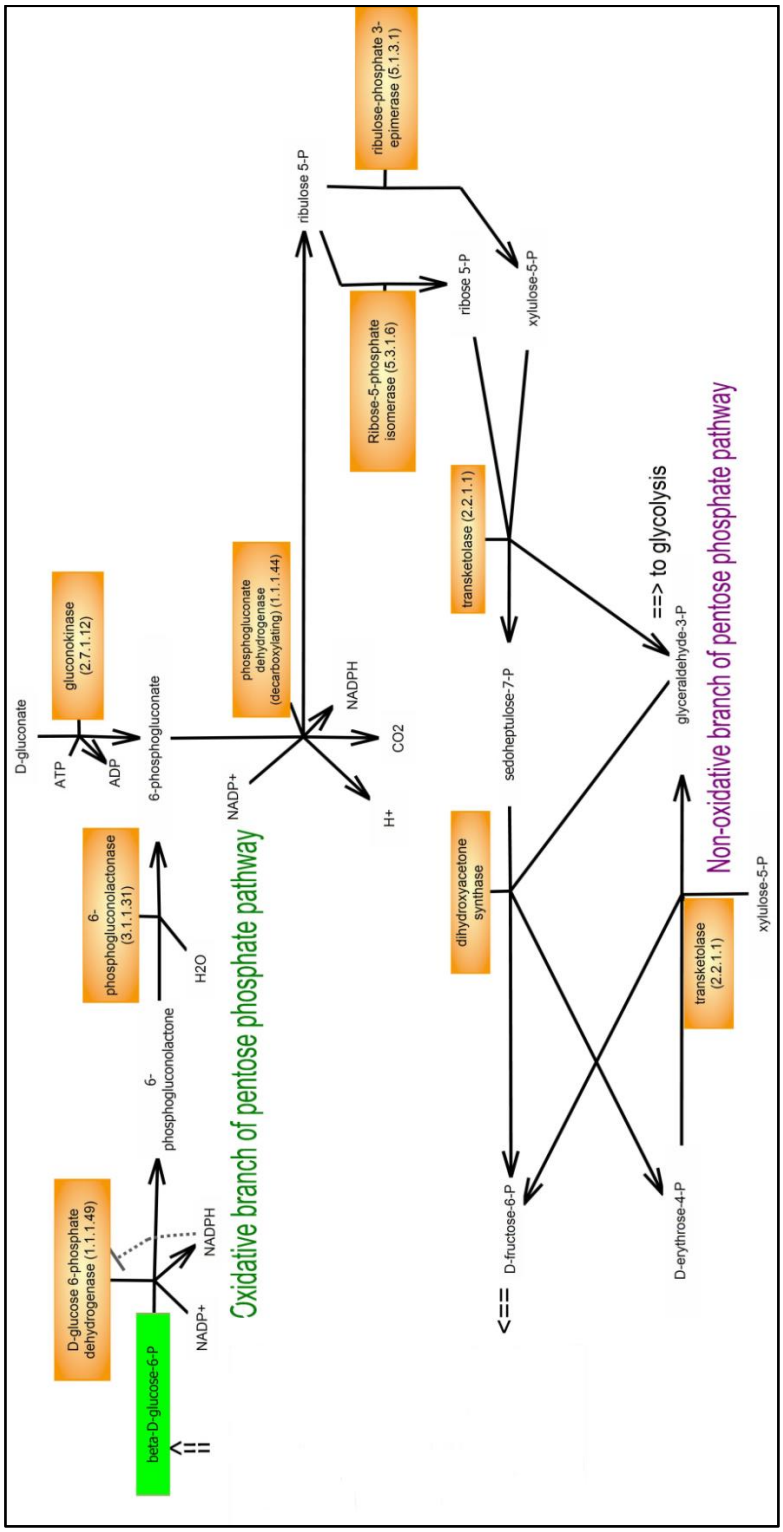

Figure S4: Predicted pentose phosphate shunt pathway in *T. lanuginosus*.

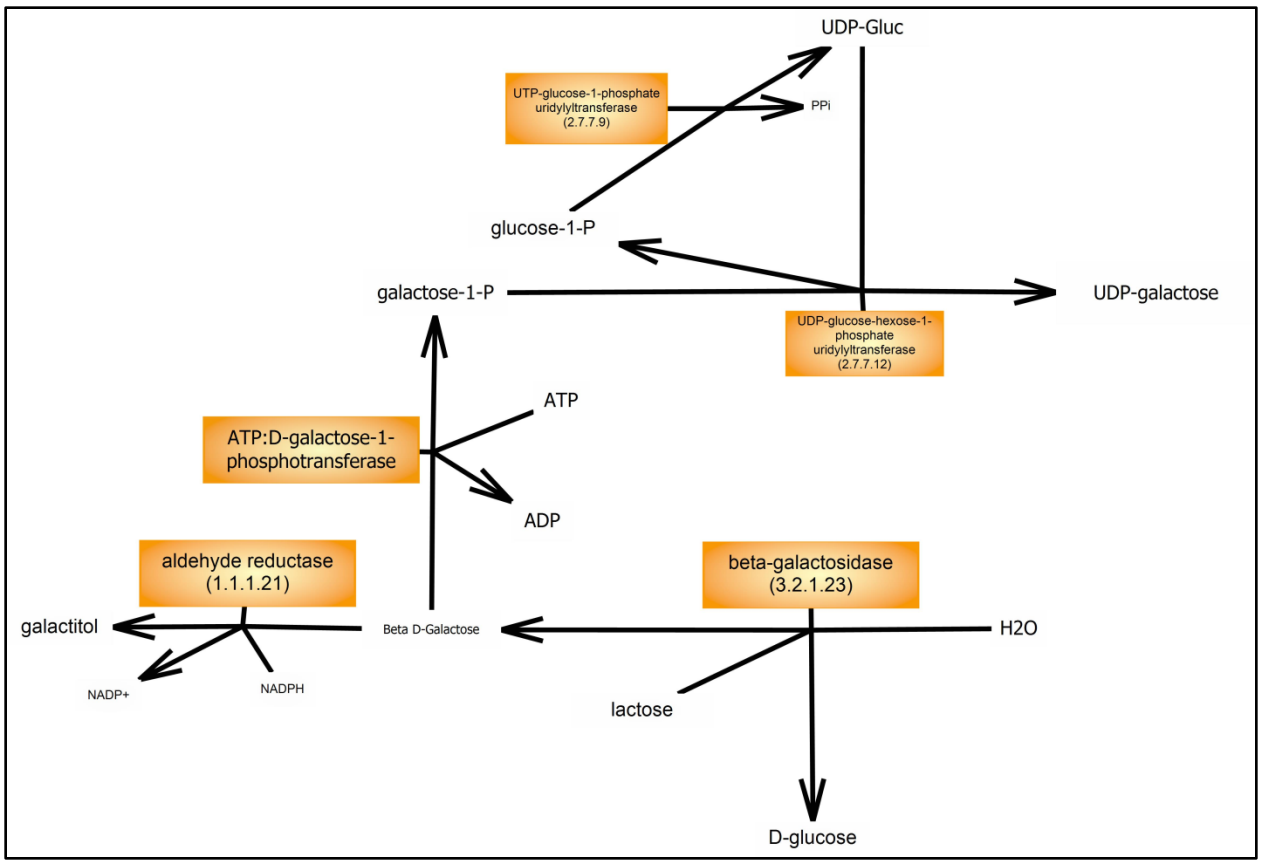

Figure S5: Predicted galactose metabolism in *T. lanuginosus*.

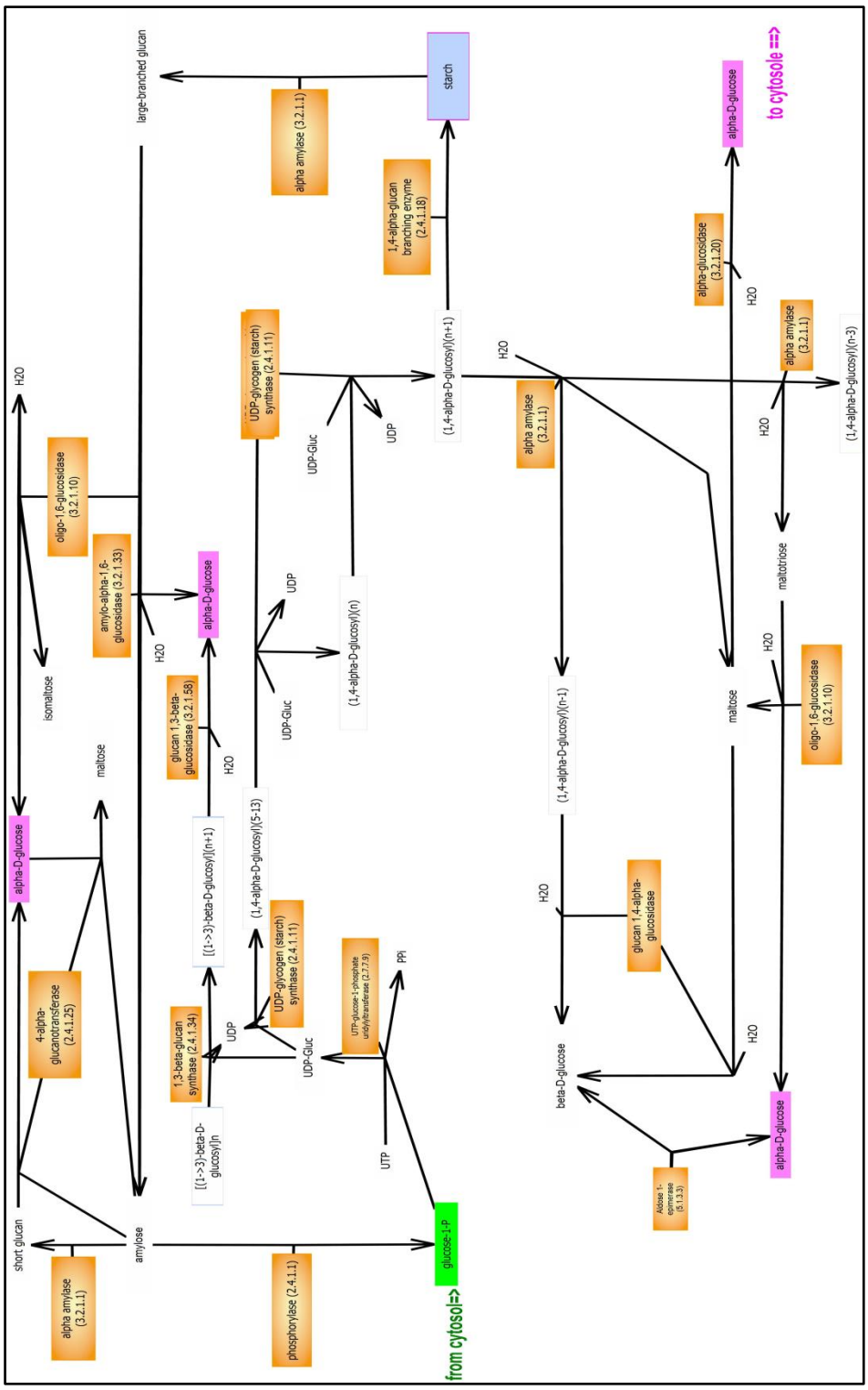

Figure S6: Predicted starch and glycogen metabolism in *T. lanuginosus*.

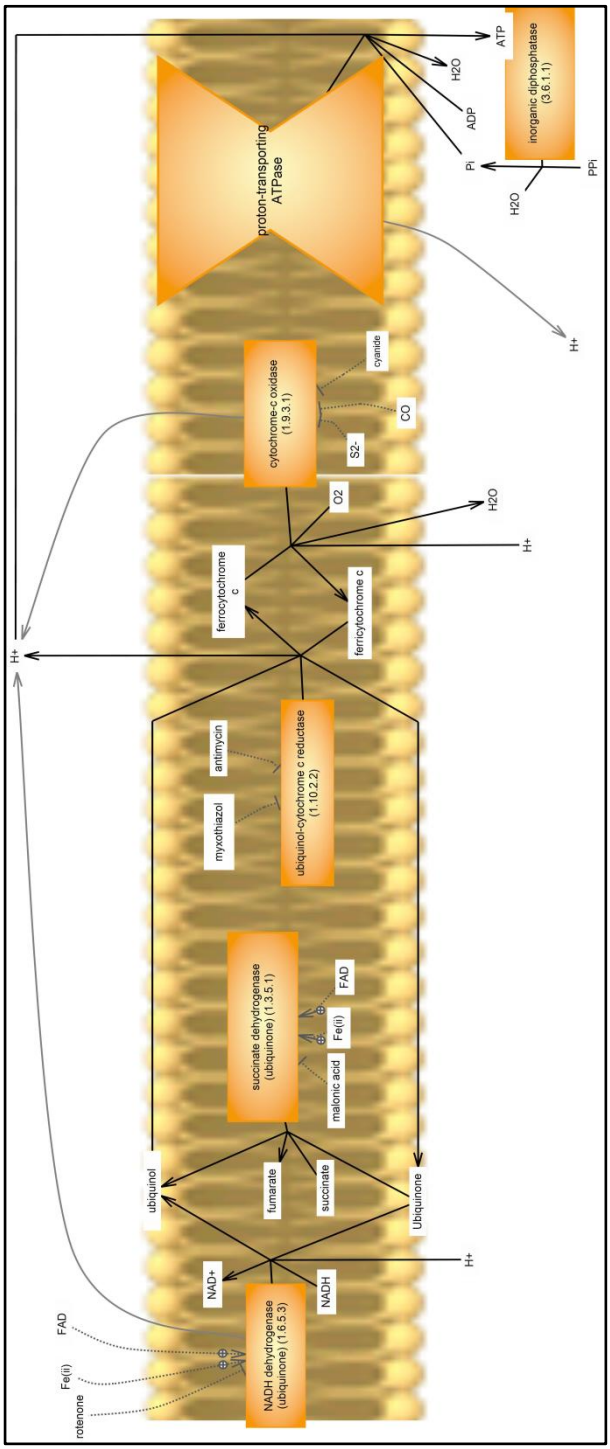

Figure S6: Predicted respiratory chain and oxidative phosphorylation pathway in *T. lanuginosus*.

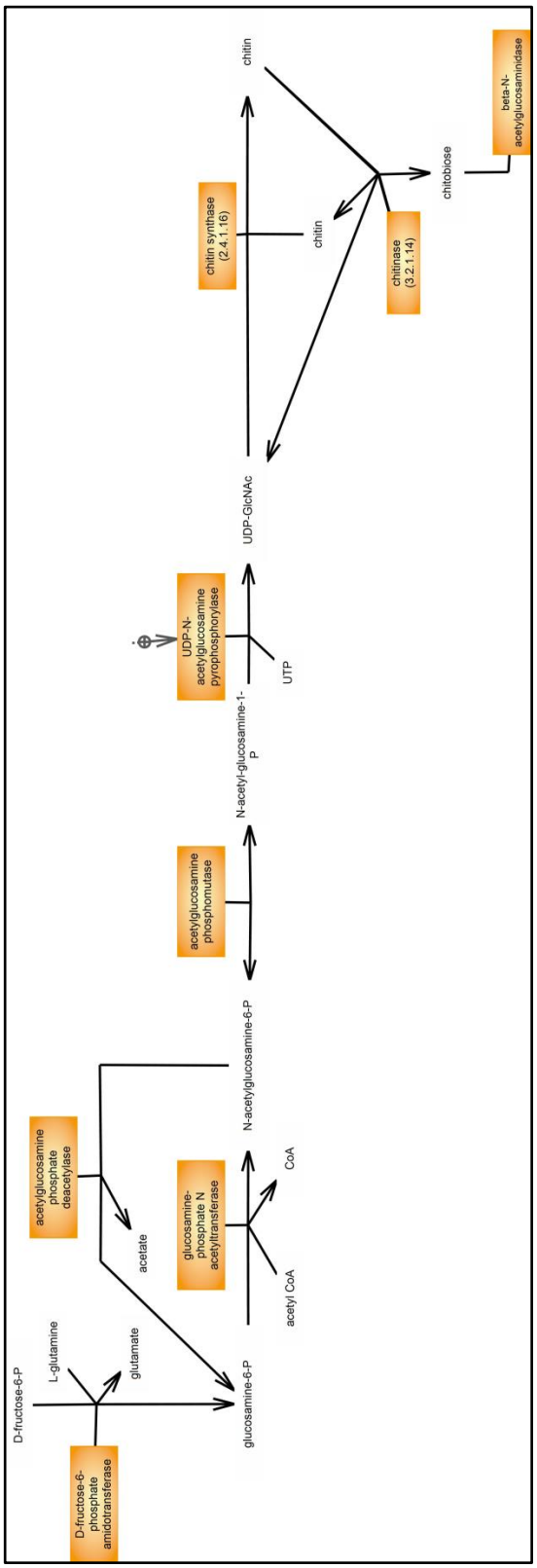

Figure S7: Predicted amino-sugar biosynthesis pathway in *T. lanuginosus*.

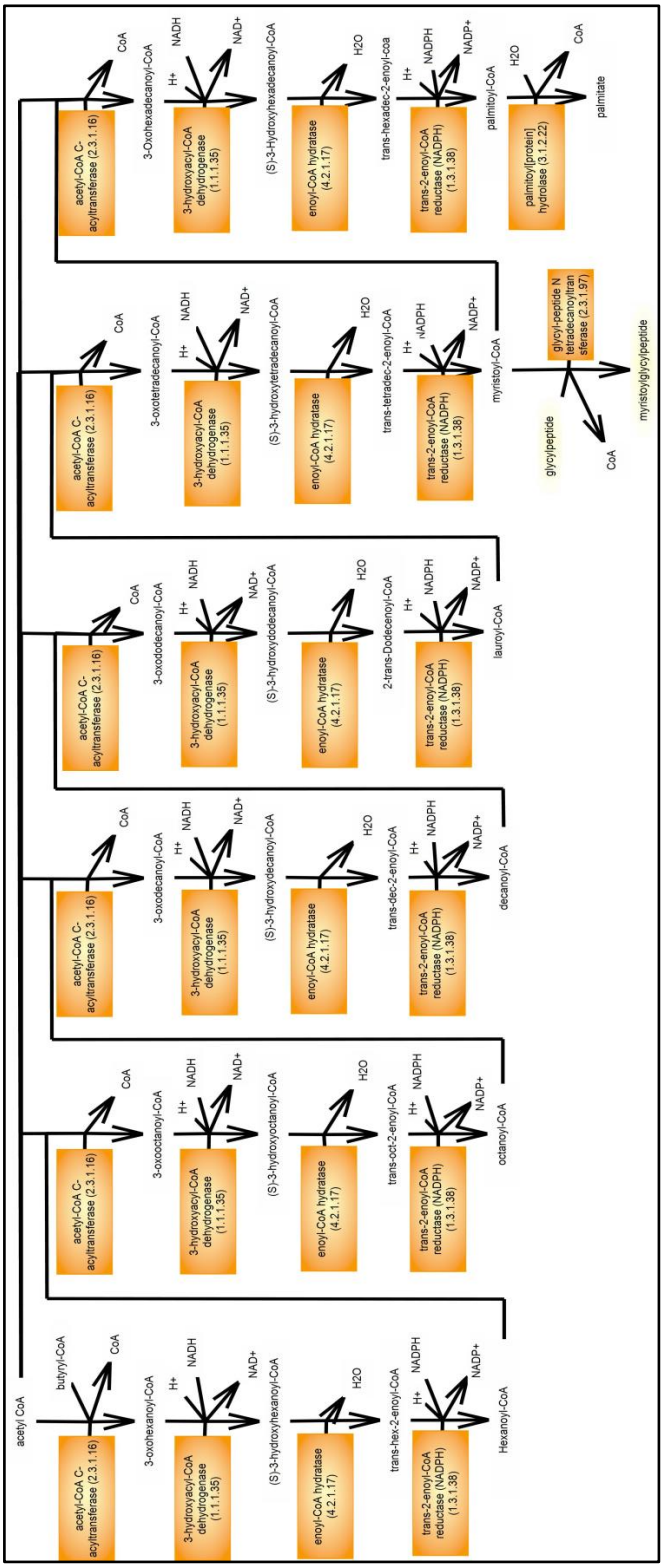

Figure S8: Predicted fatty acids synthesis pathway in *T. lanuginosus*.

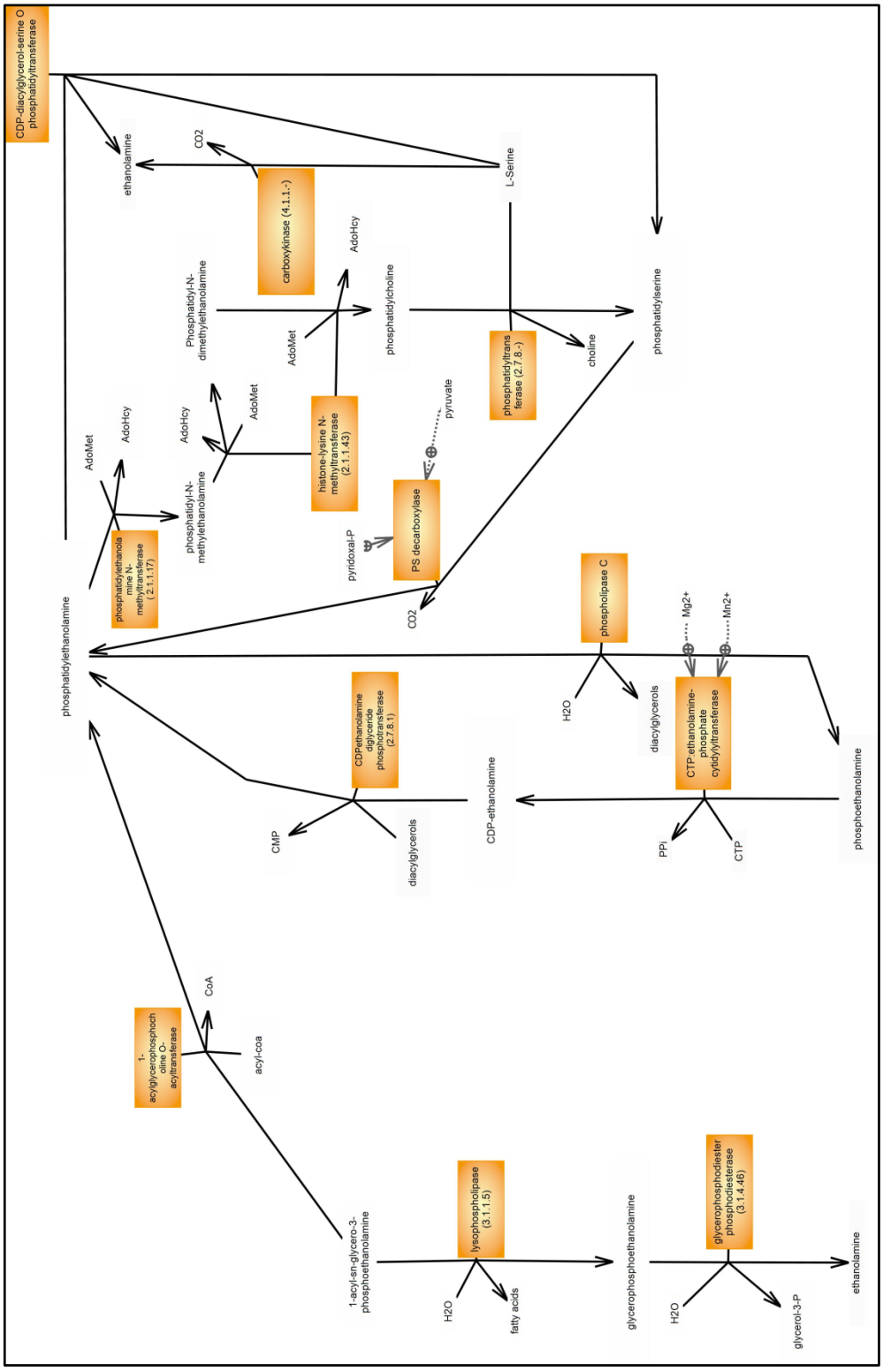

Figure S9: Predicted Phosphatidyl ethanolamine metabolic pathway in *T. lanuginosus*.

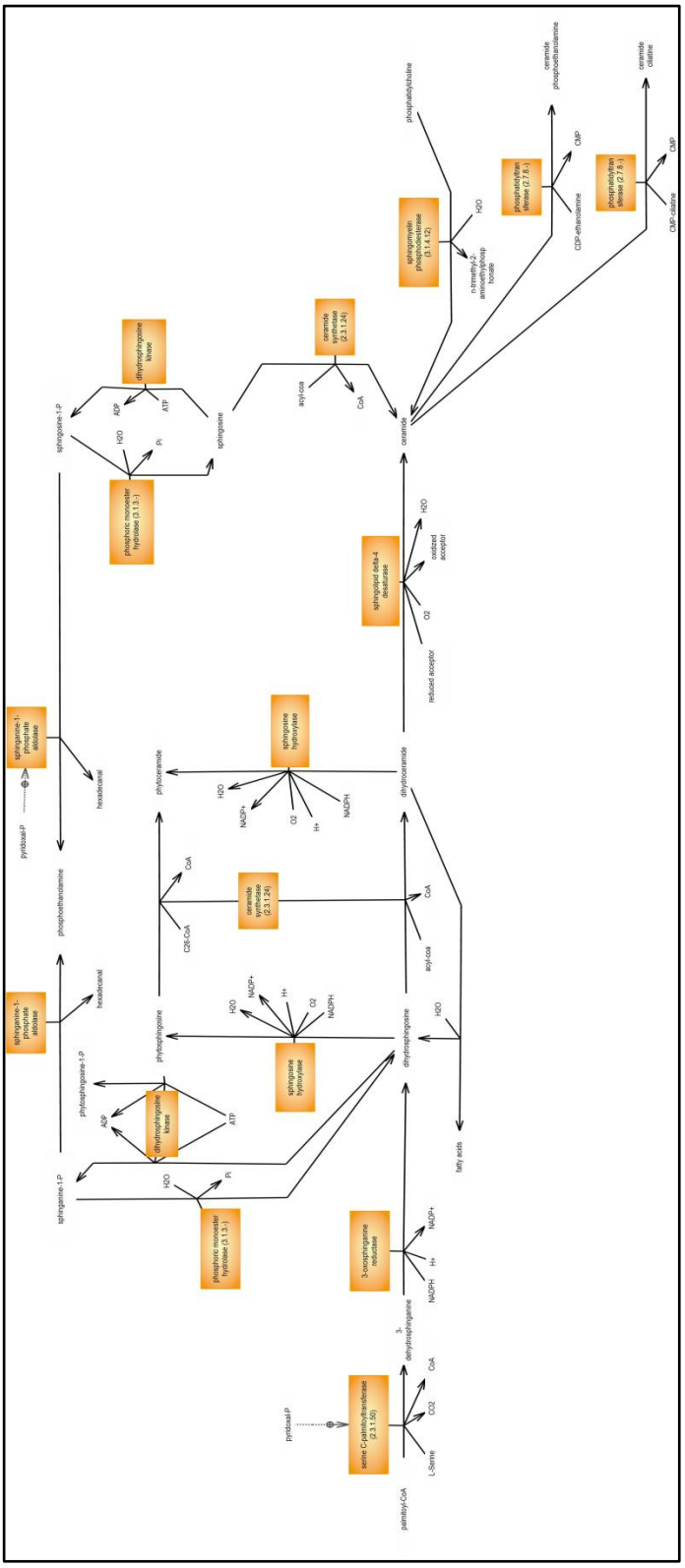

Figure S10: Predicted sphingolipid metabolism in *T. lanuginosus*.

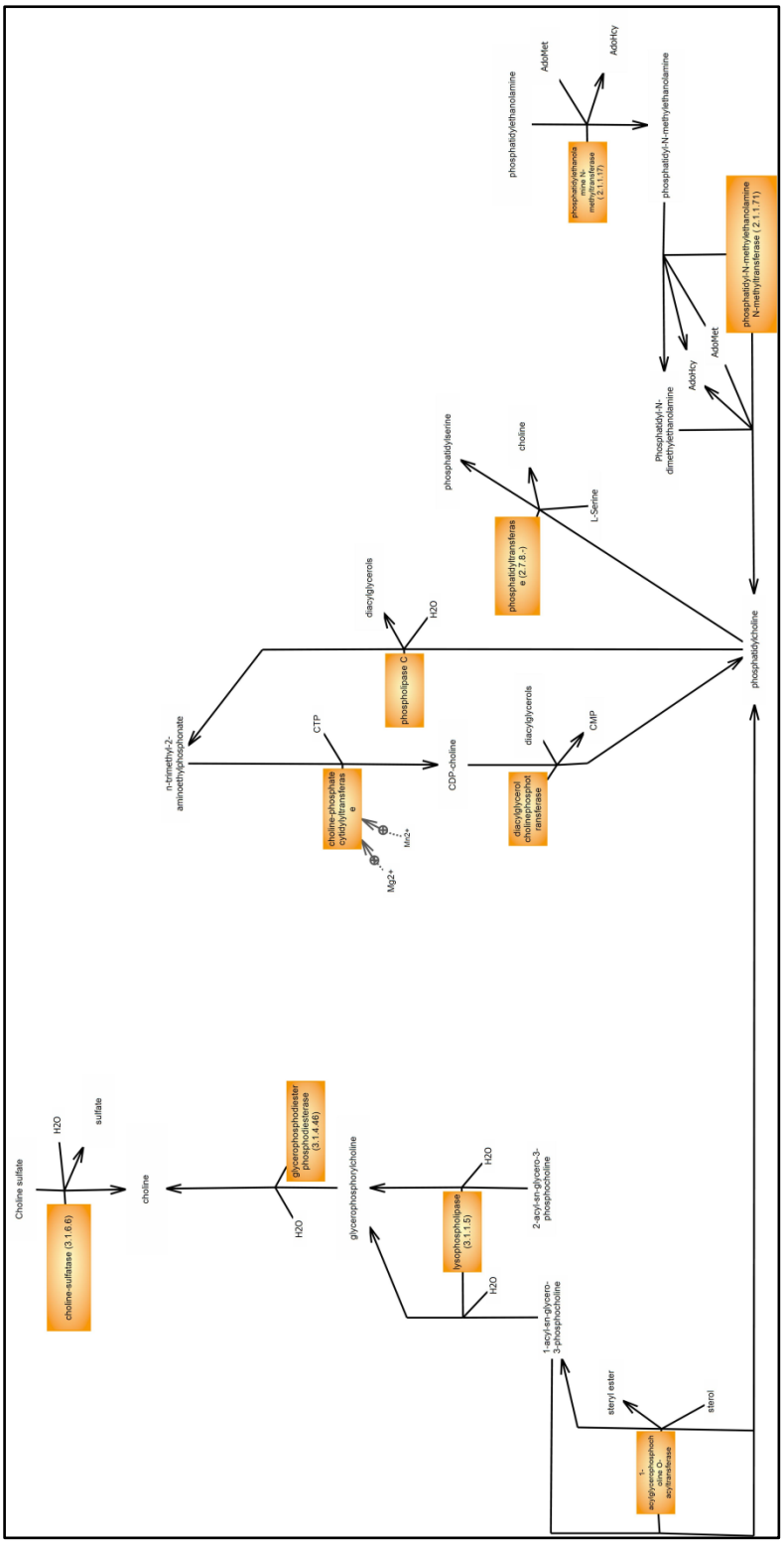

Figure S11: Predicted choline metabolism in *T. lanuginosus*.

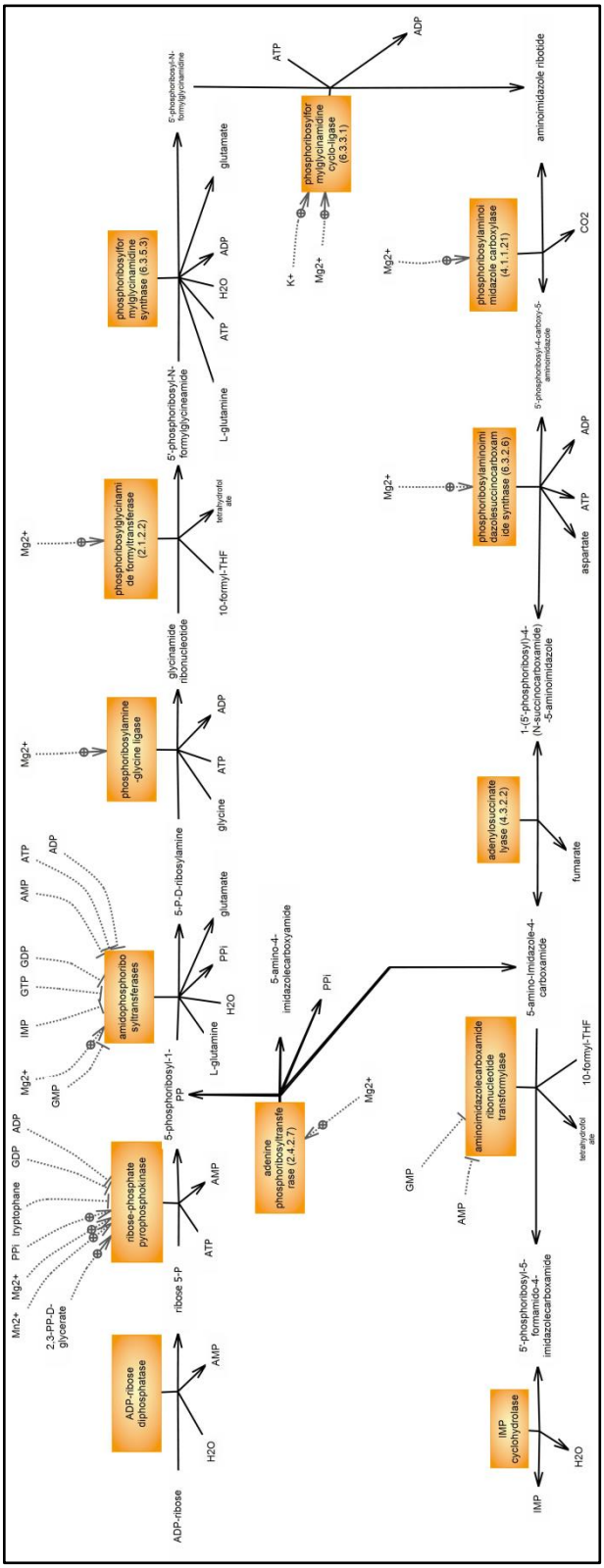

Figure S12: Predicted purine precursor biosynthesis pathway in *T. lanuginosus*.

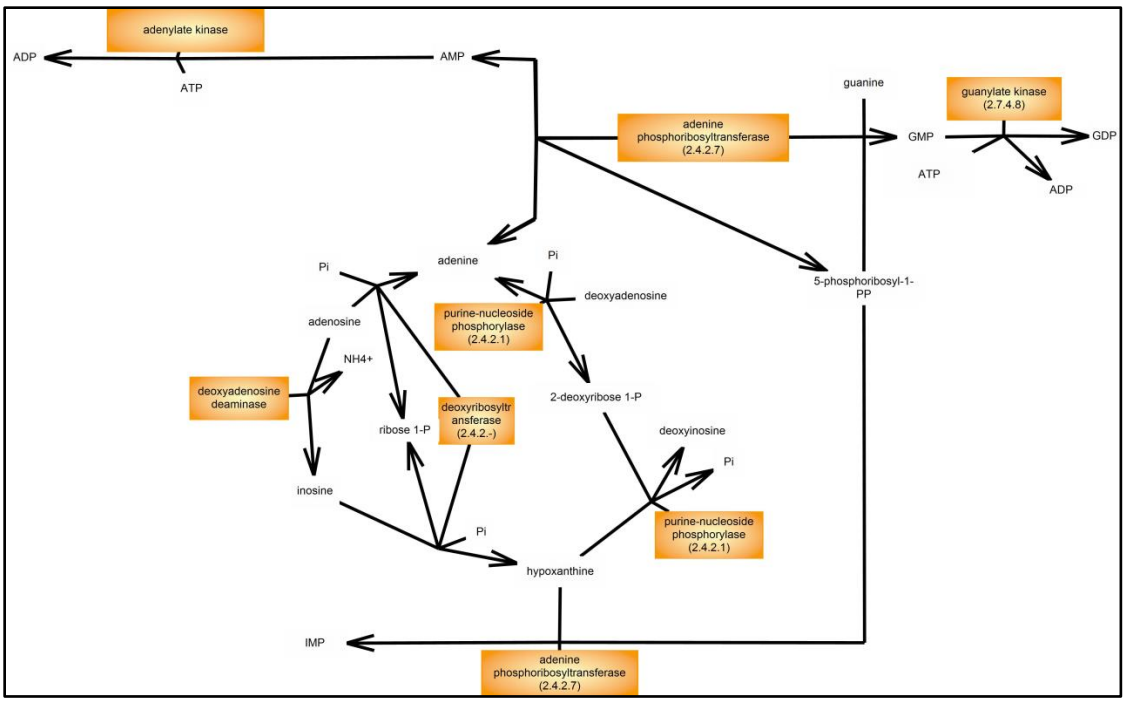

Figure S13: Predicted purine salvage cycle in *T. lanuginosus*.

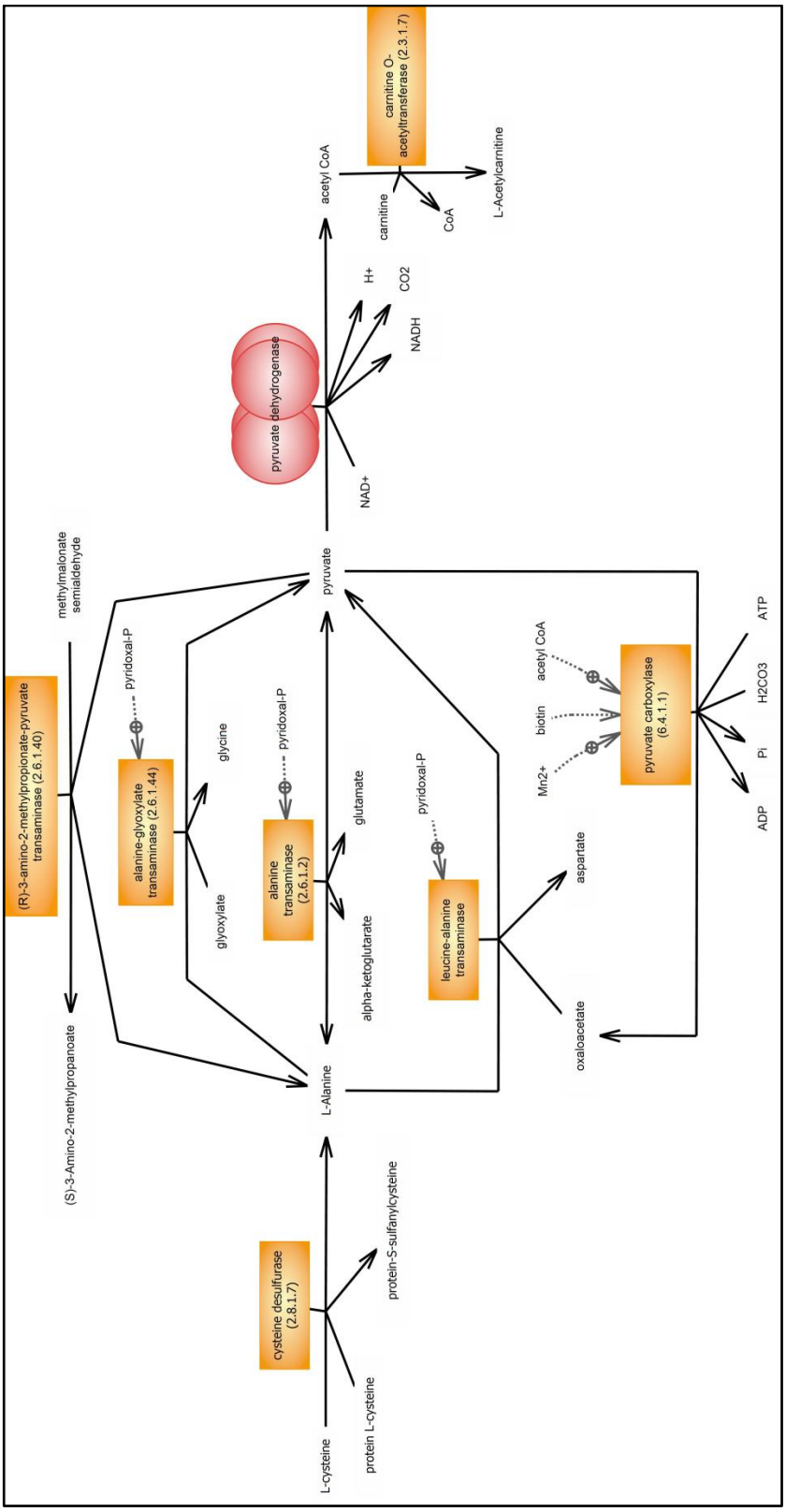

Figure S14: Predicted alanine biosynthesis in *T. lanuginosus*.

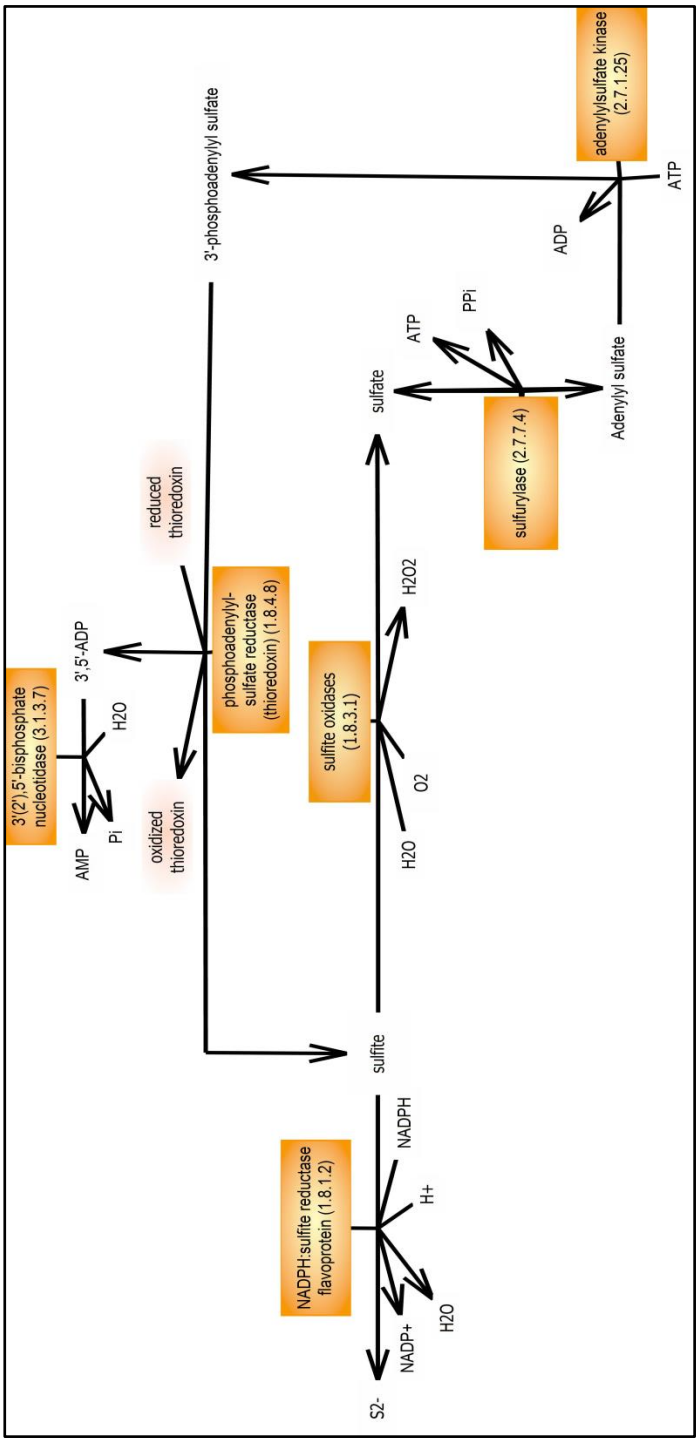

Figure S15: Predicted sulphate metabolism in *T. lanuginosus.*

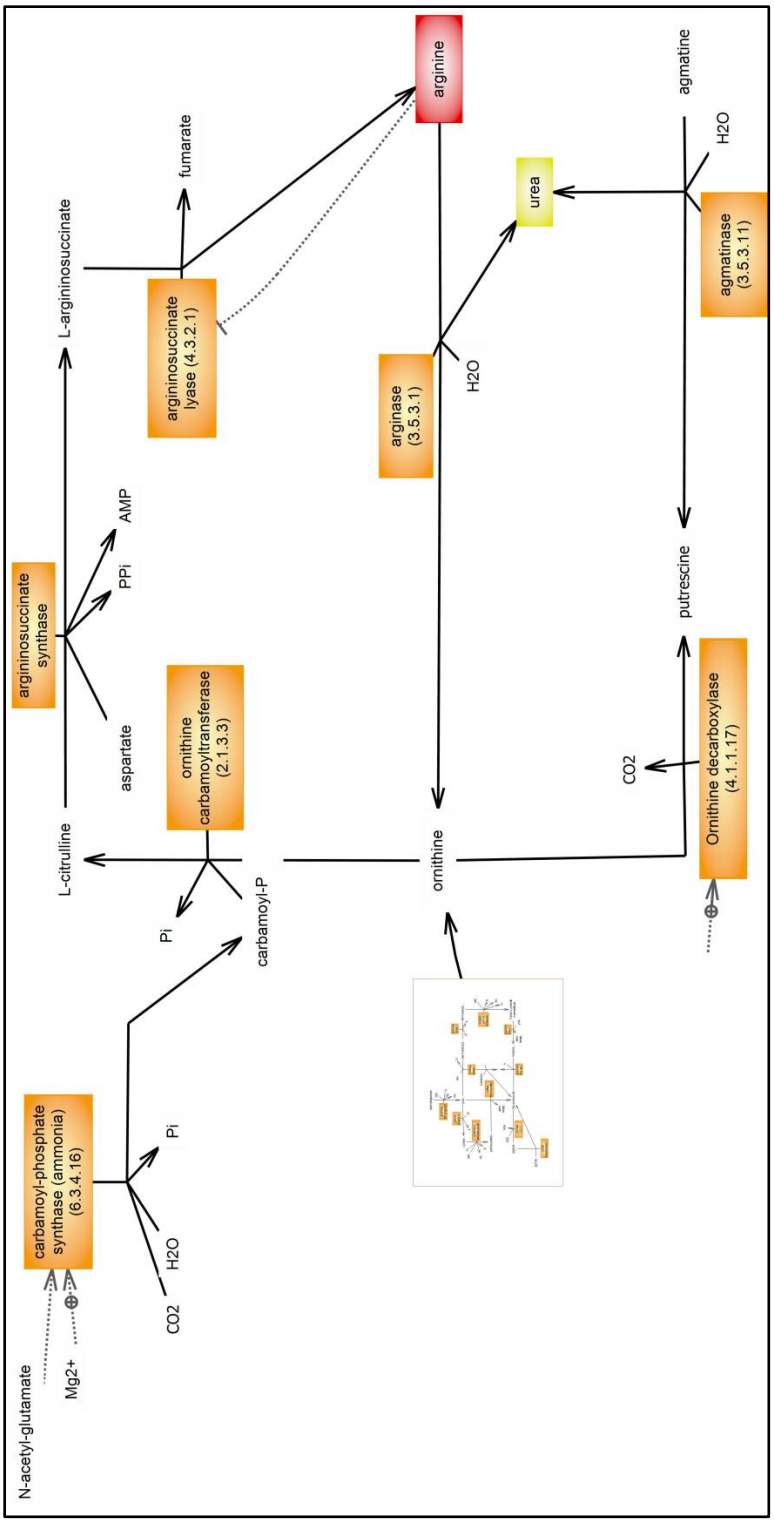

Figure S16: Predicted urea cycle in *T. lanuginosus*.

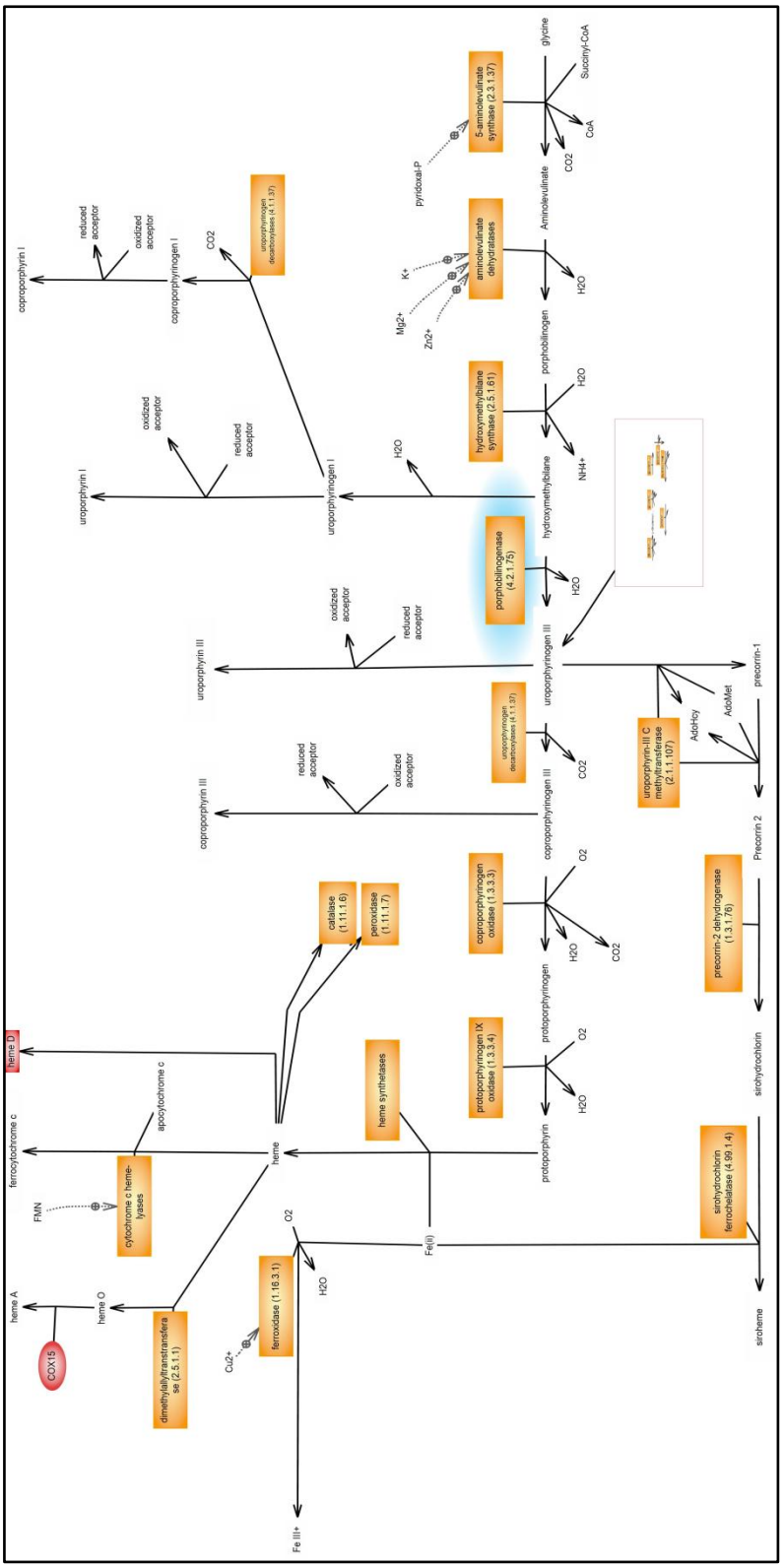

Figure S17: Predicted heme biosynthesis pathway in *T. lanuginosus.*