Multi-Omics studies on the hypercellulolytic mutant *Penicillium Janthinellum* NCIM 1366 for elucidating its CAZyme profile and cellulase gene regulation

by

Athiraraj S R
10BB16A39012

A thesis submitted to the
Academy of Scientific & Innovative Research
for the award of the degree of

DOCTOR OF PHILOSOPHY
in
SCIENCE

Under the supervision of
Dr. Rajeev K. Sukumaran

CSIR-National Institute for Interdisciplinary Science and Technology, Thiruvananthapuram, India

Academy of Scientific and Innovative Research
AcSIR Headquarters, CSIR-HRDC campus
Sector 19, Kamla Nehru Nagar,
Ghaziabad, U.P. – 201 002, India

August 2025

राष्ट्रीय अंतर्विषयी विज्ञान तथा प्रौद्योगिकी संस्थान
वैज्ञानिक तथा औद्योगिक अनुसन्धान परिषद् | विज्ञान तथा प्रौद्योगिकी मंत्रालय, भारत सरकार
इंडस्ट्रियल इस्टेट पी.ओ, पाप्पनमकोड, तिरुवनंतपुरम, भारत - 695 019,
CSIR- National Institute For Interdisciplinary Science & Technology (CSIR-NIIST)
Council of Scientific and Industrial Research | Ministry of Science and Technology, Government of India
Industrial Estate PO, Pappanamcode, Thiruvananthapuram, India - 695019

CERTIFICATE

This is to certify that the work incorporated in this Ph.D. thesis entitled "**Multi-Omics studies on the hypercellulolytic mutant *Penicillium janthinellum* NCIM 1366 for elucidating its CAZyme profile and cellulase gene regulation**" submitted by **Ms. Athiraraj S R**, to the **Academy of Scientific and Innovative Research (AcSIR)**, in fulfilment of the requirements for the award of the **Degree of Doctor of Philosophy in Science,** embodies original research work carried out by the student. We further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research materials obtained from other sources and used in this research work have been duly acknowledged in the thesis. Images, illustrations, figures, tables etc., used in the thesis from other sources have also been duly cited and acknowledged.

06/08/2025

Athiraraj S R

(Student)

06/08/2025

Dr. Rajeev K Sukumaran

(Supervisor)

STATEMENTS OF ACADEMIC INTEGRITY

I, **Athiraraj S R**, a Ph.D. student of the Academy of Scientific and Innovative Research (AcSIR) with Registration No. **10BB16A39012**, hereby undertake that the thesis entitled **"Multi-Omics studies on the hypercellulolytic mutant *Penicillium janthinellum* NCIM 1366 for elucidating its CAZyme profile and cellulase gene regulation"**, has been prepared by me and that the document reports original work carried out by me and is free of any plagiarism in compliance with the UGC Regulations on "Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)" and the CSIR Guidelines for "Ethics in Research and in Governance (2020)".

Athiraraj S R

Date: 06/08/2025

Place: Thiruvananthapuram

It is hereby certified that the work done by the student, under my supervision, is plagiarism-free in accordance with the UGC Regulations on "Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)" and the CSIR Guidelines for "Ethics in Research and in Governance (2020)".

Dr. Rajeev K. Sukumaran

Date: 06/08/2025

Place: Thiruvananthapuram

DEDICATED TO MY FAMILY AND TEACHERS

Acknowledgements

I extend my heartfelt gratitude to my supervisor Dr. Rajeev K Sukumaran, for granting me the opportunity to pursue this research. I deeply appreciate the freedom to explore independently, paired with his expertise, clarity in breaking down challenging concepts and constructive feedback, which significantly contributed to my growth as a researcher. I am grateful for his unwavering patience and support during challenging times which gave me the resilience and encouragement to complete this work.

I thank Dr. C. Anandharamakrishnan, and Dr. A. Ajayaghosh (present and former Directors of CSIR-NIIST respectively) for providing the facilities for carrying out my research work in the Institute. I also thank Dr. Kaustabh Kumar Maiti and Dr. K. Madhavan Nampoothiri (Present & former HODs, Biosciences and Bioengineering Division, BBD) for all the support provided in the division.

I am grateful to the support provided by Dr. Jayamurthy P., Dr. Karunakaran V., Dr. Suresh C. H., Dr. Luxmi Varma and Dr. Muthu Arumugam as the AcSIR coordinators. I would like to acknowledge the Doctoral Advisory Committee (DAC) members- Dr. Binod Parameswaran, Dr. N. Ramesh Kumar, and Dr. Jayamurthy P. for their suggestions and support during evaluations.

I would like to acknowledge the financial support from Council for Scientific and Industrial Research (CSIR), India in the form of Junior/Senior Research Fellowship.

I am also grateful to all the scientists, and staff of BBD, for providing me with valuable suggestions and support during review meetings and in my research activities. I am thankful to Dr. Amith Abraham for all the support and guidance during the initial phase of the study. I thank all the current and former members of the Biofuels Research Group, Meena, Reshma, Prajeesh, Anoop, Meera, Aswathi, Adarsh, Valan, Athulya, Neetha, Biji Aravind, Ayman, Sabeela, Dileep and Merin, for fostering a friendly and collaborative working environment. I am grateful to my friends at NIIST, especially Anju, Reshu, Roopa and Amala for all the emotional support during my research journey.

I am forever grateful to my parents, Sreeja and Raju, and my sister Anupama for their unconditional love and support throughout this journey. My heartfelt thanks go to my husband, Arunghosh, whose selfless sacrifice of his own career enabled me to complete this work. The cheerful presence of my baby, Milan made all the difference during the most challenging times. I thank God almighty for giving me this strength, patience, courage and a wonderful group of people around me to face all these challenges boldly.

Athiraraj S R

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA	:	Auxiliary Activities
BGL	:	Beta-glucosidase
CAZymes	:	Carbohydrate Active enZymes
CBH	:	Cellobiohydrolase
CCD	:	Central Composite Design
CE	:	Carbohydrate Esterases
CMCase	:	Carboxymethyl Cellulase
EG	:	Endoglucanase
FPU	:	Filter Paper Unit
FC	:	Fold Change
FPKM	:	Fragments Per Kilobase of transcript per Million
GH	:	Glycoside Hydrolases
GT	:	Glycosyl Transferases
HPLC	:	High-performance Liquid Chromatography
IU/mL	:	International Units Per Milliliter
KEGG	:	Kyoto Encyclopedia of Genes and Genomes
LC-MS/MS	:	Liquid Chromatography Tandem Mass Spectrometry
LPMO	:	Lytic Polysaccharide Mono-Oxygenases
MW	:	Mandels & Weber Medium
NaOH	:	Sodium hydroxide
NCIM	:	National Collection of Industrial Microorganisms
NREL	:	National Renewable Energy Laboratory
PDA	:	Potato Dextrose Agar
Pj	:	*Penicillium janthinellum NCIM-1366*
PL	:	Polysaccharide Lyases
SDS – PAGE	:	Sodium Dodecyl-Sulfate Polyacrylamide Gel Electrophoresis
w/w	:	weight per weight
w/v	:	weight in volume

Chapter1

Introduction and Review of Literature

1.1 Introduction

Cellulases are a group of enzymes that play vital role in converting cellulose to glucose making the essential for in the global carbon cycle. Active research on these enzymes began in the early 1950s, driven by their potential to degrade lignocellulosic biomass (Reese and Mandels, 1984). According to the Carbohydrate-Active enZYmes database (CAZy), cellulases are included in the Glycoside Hydrolase (GH) family. They are produced by wide range of organisms including bacteria, fungi and protozoans. Among these, filamentous fungi are the most prominent industrial producers, attributed to their high secretion of extracellular proteins (Chen, 2014). Based on the mode of action, cellulases are categorized into three major groups-endo-(1,4)-β-D-glucanases (EC 3.2.1.4), exo-(1,4)-β-D-glucanases (EC 3.2.1.91), and β-glucosidases (EC 3.2.1.21) (Schülein, 1988). Other enzymes such as hemicellulases, lytic polysaccharide monooxygenases (LPMOs), and swollenins also contribute to efficient biomass degradation. Due to their diverse catalytic properties, cellulases and related enzymes have numerous industrial applications in the textile, food, pulp and paper, pharmaceutical, and bioethanol sectors (Kuhad et al., 2011).

Almost invariably, cellulases are tightly regulated by induction and repression mechanisms since the synthesis and secretion of these enzymes is energy expensive for the organism (Amore et al., 2013). Thus, in the presence of easily metabolizable carbon sources like glucose in the growth medium, cellulase production is generally repressed. Most information on these aspects has been gained through studies on the hypercellulolytic fungus *Trichoderma reesei* (Ilmén et al., 1997). The induction signal from complex carbon sources like cellulose or hemicellulose is mediated by soluble products derived from their hydrolysis or transglycosylation by basal expression of cellulases. However, it remains unclear how the induction signals are delivered, although key regulators in different signal transduction pathways have been identified (Yan et al., 2021a). At the transcriptional level, this regulation is achieved through regulators such as the carbon catabolite repressor Cre1, Ace1, Ace2, Hap2/3/5 complex, Xyr1 etc. in *T.reesei* (Amore et al., 2013). Homologs of these regulators were also identified in other cellulase-producing fungi from genera *Aspergillus*, *Neurospora*, *Penicillium*, etc., but studies suggest divergence in modes of genome wide regulation in different species (Zhao et al., 2023). So the regulation must be studied for individual fungi for understanding/improving their cellulase production. The regulation by these factors is complex and involves cross talk or coordinated regulation and multiple target genes. Even though

these regulators and their role in cellulase regulation have been identified in other well studied cellulolytic fungi and their overexpression/knockout altered the cellulase production, knowledge on the regulatory pathway is still incomplete.

1.2 Rationale and Objectives of the study

The major cost involved in the production of bioethanol from lignocellulosic biomass is contributed by the hydrolysis step, due to the cost of cellulase enzymes (Saini et al., 2015). Additionally, cellulase from *Trichoderma reesei* (the industrial workhorse for cellulase production) is deficient in cellobiases (beta -glucosidases), causing accumulation of cellobiose in the hydrolysis medium, which results in end-product inhibition of the enzymes. So, supplementing with cellobiases from other sources is necessary for complete hydrolysis adding to the cost factor. Thus, the conversion of lignocellulosic biomass waste to bioethanol is not yet commercially feasible. Although significant enhancement in the enzyme production by these fungi has been achieved over the past years through random/rational strain improvement, it is limited by the fact that the complete mechanism of cellulase production and regulation remains to be understood. Understanding the regulation of enzymes that are involved in the hydrolysis of lignocellulosic biomass is an important area of research, since this opens possibilities to engineer enhanced cellulase production, and hence the cost of biomass hydrolysis and biorefinery operation.

Many species of *Penicillium* are reported to produce high titres of cellulase enzymes and has been used to hydrolyze lignocellulosic biomass. *Penicillium janthinellum* NCIM 1171 is a filamentous soil fungus known for its efficient cellulase production and hydrolysis efficiency (Adsul et al., 2004). Classical mutagenesis studies conducted at CSIR-NCL, Pune had yielded three mutants of *P. janthinellum* NCIM 1171 with enhanced cellulase production and they were named EMS-UV-8 (NCIM1366), EU-21, and EU2D-21(Adsul et al., 2007, 2009). The mutant strain *P. janthinellum* NCIM1366 produced cellulases efficient in the hydrolysis of natural biomass substrates and we have chosen it for the study. With further improvement at the genetic level, it has the potential to improve industrial cellulose production. However the cellulase enzyme system and mechanism of cellulase regulation are less explored and understood in this strain. This is mainly due to the unavailability of its genome sequence. The present study is intended to investigate the mechanism of production and regulation of cellulase enzymes from *Penicillium janthinellum* NCIM166. Whole genome sequencing and comparison of transcriptome profiles of induced and uninduced cultures of the organism were ex-

pected to reveal the regulators involved in the pathway. With the knowledge on the regulatory mechanism, it is possible to identify the target genes for strain engineering, for the future industrial applications.

The main objectives of the current study are

- Comparative study of the enzyme activity and hydrolysis efficiency of cellulases from *P. janthinellum* NCIM1366 and the industrial strain *T. reesei* RUT-C30 and elucidating the differences in their secretomes.
- Genome sequencing and analysis of *P. janthinellum* NCIM1366 for the identification of CAZyme genes and the orthologs for fungal regulators of cellulase expression.
- Comparative transcriptome analysis for identifying differentially expressed CAZymes under induced and uninduced conditions.
- Identification of the probable regulators/factors constituting cellulase regulon in *P. janthinellum* from the transcriptome data.

1.3 Review of literature

1.3.1 Structure of lignocellulosic biomass

Lignocellulosic biomass is the plant's dry matter and the most abundant raw material on the earth, with an annual production of 181.5 billion tonnes (Dahmen et al., 2019). Lignocellulosic biomass mainly comprises three polymers- polysaccharides cellulose and hemicellulose and aromatic polymer lignin. In addition, it also contains proteins, lipids, ash, pectin, and other low molecular-weight carbohydrates.

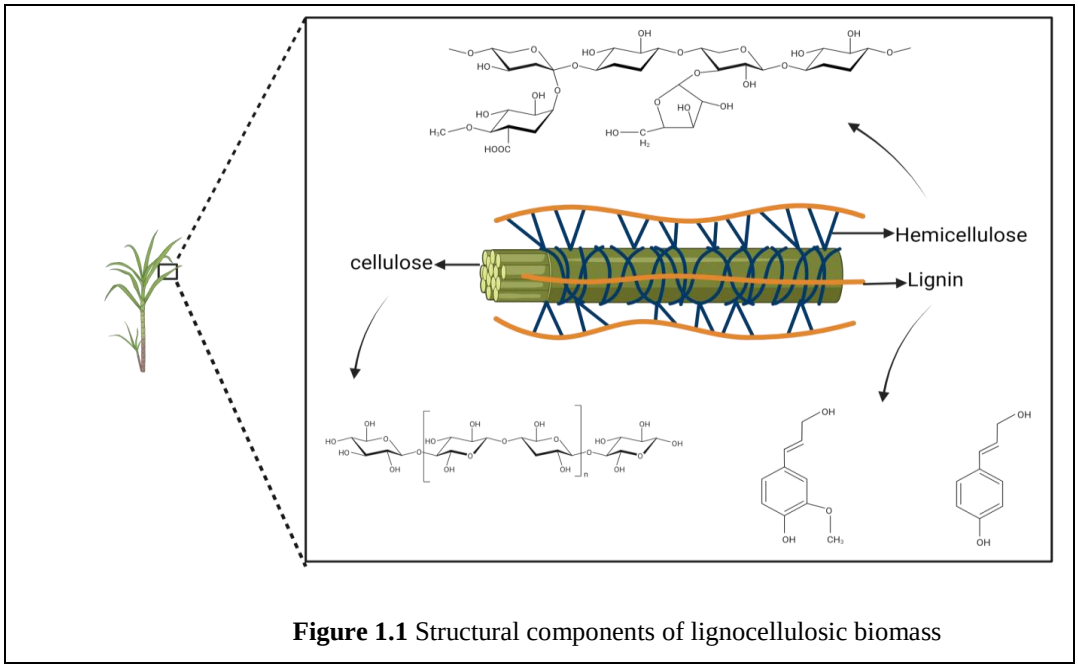


Figure 1.1 Structural components of lignocellulosic biomass

Cellulose is the most abundant polymer of lignocellulose, which amounts to 35-50 % in weight. In its chemical structure, β-D- glucopyranose units are linked together by β-(1, 4) glycosidic bonds with chemical formula $(C_6H_{10}O_5)_n$, where n represents the number of glucose groups in the linear chain, which ranges from several hundred to ten thousand. The cellulose molecules aggregate to form two distinct regions: crystalline and amorphous. In the crystalline region, the cellulose molecules are arranged orderly, and the hydroxyl groups in the crystalline region form intramolecular and intermolecular hydrogen bonds, forming elementary microfibrils. These microfibrils are then packed into fibers making cellulose molecules insoluble in most solvents. The amorphous region is arranged irregularly and has a relaxed structure because of the presence of free hydroxyl groups in this region (Chen, 2014).

Hemicelluloses are complex highly branched heterogenous polysaccharides with monosaccharides, including pentoses (xylose, rhamnose, and arabinose), hexoses (glucose, mannose, and galactose), and uronic acids (4-O-methyl glucuronic, D-glucuronic, and D-galacturonic acids). According to its chemical structure, there are three main types of hemicellulases- xylans, glucans, and mannans. The content and structure of hemicellulose in various plants differ, but it generally constitutes about 20-35 % dry weight. Hemicelluloses have a degree of polymerization, usually less than 200, and are generally non-crystalline(Betts et al., 1991).

Lignin is a highly irregular and insoluble polymer consisting of phenylpropanoid subunits, namely p-hydroxyphenyl (H-type), guaiacol (G-type), and syringyl (S-type) units(Malherbe & Cloete, 2002). Unlike other polymers, it has a random three-dimensional structure with no typical repeating unit and constitutes around 5-30 % dry weight. In addition, the inter-monomeric linkages are very stable, non-hydrolyzable carbon-carbon bonds or relatively inert ether bonds(Betts et al., 1991). It confers a rigid, impermeable resistance to microbial attack and oxidative stress to the plant. In lignocellulosic biomass, different polymers are associated with each other in a heteromatrix with a varying relative composition which depends on the type, species, and even source of the biomass(Bajpai, 2016). Cellulose is complexed with hemicellulose, lignin, and other components by covalent and non-covalent interactions constituting its tertiary structure. Cellulose microfibrils are surrounded by hemicellulose polysaccharides by covalent bonds and hydrogen bonds, and these covalent bonds are highly resistant to hydrolysis.

1.3.2 Enzymes for the degradation of biomass

By definition, cellulases are a group of enzymes involved in the cellulose pathway to glucose conversion and thus play an essential role in the global carbon cycle and several industries. Active research on these enzymes started in the early 1950s, owing to their potential to degrade lignocellulosic biomass to glucose and other soluble sugars(Reese and Mandels, 1984). They are produced by organisms from different taxonomic levels, which include bacteria, fungi, protozoans, plants, and animals, with fungi being the major industrial cellulase producers. The cellulase systems produced by these organisms are mainly two types-discrete non complexed cellulases and complexed cellulases. Filamentous fungi can penetrate cellulose through their hyphal extension and present their cellulase systems in confined cavities within cellulosic particles. These systems are therefore known as free or noncomplexed systems.

Some actinomycetes also produce such a kind of cellulase system. Complexed cellulase systems (cellulosomes) are generally produced by anaerobic bacteria which cannot effectively penetrate cellulose. They form protuberances on the cell wall, which are stable enzyme complexes that bind tightly to microcrystalline cellulose(Lynd et al., 2002).

According to Carbohydrate Active Enzyme database (CAZy), cellulases are included in Glycoside Hydrolase (GH) family. There are 115 glycoside hydrolase families and cellulases span over 13 families. In the general structure of cellulases, there are catalytic and non-catalytic modules(Jayasekara and Ratnayake, 2019). Cellulases are classified into different GH families based on the amino acid sequence and crystal structure of the catalytic module. Non-catalytic domain known generally helps in binding to cellulose during the catalysis and therefore known as Cellulose Binding domain (CBD). These domains are linked together by a linker domain.

1.3.3 Role of cellulases in bioethanol production

Second-generation (2G) bioethanol production which utilizes lignocellulosic biomass (agricultural residues) has emerged as a sustainable alternative to fossil fuels by addressing growing energy demands and environmental concerns. Key steps involved in this process are pretreatment of biomass to expose cellulose and hemicellulose, enzymatic hydrolysis to break down these polymers into fermentable sugars, fermentation to convert sugars into ethanol, and distillation to produce fuel-grade ethanol. Unlike first-generation bioethanol which uses food crops, 2G bioethanol uses abundant non-food biomass, helping address environmental concerns and energy demands(Gupta and Verma, 2015).

The efficiency of 2G bioethanol productions is largely depended on the activity and stability of these cellulase enzymes used in the hydrolysis step. This in turn can be influenced by external factors such as pH, temperature, and substrate concentration(Kumar et al., 2008). Researchers have also focused on engineering cellulases to improve their stability during simultaneous saccharification and fermentation (SSF), which enhances overall ethanol yield(Percival Zhang et al., 2006).

1.3.4 Cellulase producing organisms

Microorganisms, including bacteria, actinomycetes, fungi, and animals, are the two groups of organisms capable of degrading lignocellulose. The cellulose-degrading machinery and mechanisms differ in each group. Bacteria generally produce intracellular enzymes or cellulosomes with very few extracellular enzymes, and their levels are not so high. Actinomycetes

have low cellulase production, and more research needs to be done on their cellulases. Fungi are the major cellulase producers in the microbial community with high levels of extracellular enzyme production and are widely used in industries. The following sections will discuss cellulases from filamentous fungi(Chen, 2014).

1.3.5 Biomass degrading enzymes from filamentous fungi

In nature, when a filamentous fungus finds a lignocellulosic biomass or its constituent polymers in its vicinity, it secretes a group of enzymes and they depolymerize the biomass. The resulting monomeric products are absorbed back into the fungal mycelia. The enzyme cocktail secreted by fungi is composed of many enzyme entities which includes cellulases, hemicellulases and accessory enzymes. Among them, cellulases are the major group since the biomass is mainly formed by cellulose. Therefore the whole biomass hydrolyzing enzyme cocktails are sometimes referred as cellulases. *T. reesei* remains the primary organism used by industry for cellulase production because more than seventy years of expertise have established reliable technologies for its use and management (Bischof et al., 2016).

Cellulases

Cellulases belong to the enzyme group O-glycoside hydrolases (EC 3.2.1.) and they hydrolyze the glycoside bond present between two or more carbohydrates molecules or between non-carbohydrate and carbohydrate molecules. Three major groups of enzymes constitute cellulases based on their mode of action- endo-(1,4)-β-D-glucanase (EC 3.2.1.4) exo-(1,4)-β-D-glucanase (EC 3.2.1.91), and β-glucosidases (EC 3.2.1.21)(Schülein, 1988). Apart from these classical enzymes, lytic polysaccharide monooxygenases is a recent addition(Vaaje-Kolstad et al., 2010). Endoglucanases (EC 3.2.1.4) act by random cleavage of internal β-glycosidic bonds in the amorphous regions of the cellulose chain, thereby making chain ends accessible for further hydrolysis. According to Carbohydrate-Active Enzymes Database (CAZy), endoglucanases are spanned over 13 glycosyl hydrolase (GH) families (Yennamalli et al., 2013). The cellulase system from *Trichoderma reesei* contains at least five EGs (EGI/Cel7B, EGII/Cel5A, EGIII/Cel12A, EGIV/Cel61A and EGV/Cel45)(Saloheimo and Pakula, 2012). Cellobiohydrolases (EC 3.2.1.91) act mainly on the crystalline portion of the cellulose, catalyzing the release of glucose or cellobiose, from the ends of the cellulose fiber. They come under GH families 5–7 and 48. *Trichoderma reesei*'s cellulase system contains two CBHs (CBH1 and CBHII) (Lynd et al., 2002). Beta-glucosidases (E.C.3.2.1.21) hydrolyze cellobiose and short cello-oligosaccharides into glucose, which is the terminal reaction

of cellulose degradation process. This reaction has been identified as the rate-limiting step in the cellulose hydrolysis. Beta-glucosidases are mainly found in the GH families 1, 3, 4, 17, 30 and 116(Magwaza et al., 2024).

Hemicellulases

Hemicellulases encompass a variety of enzymes that hydrolyze the hemicellulosic fraction of lignocellulosic biomass. Important hemicellulases for biomass hydrolysis include xylanases (EC 3.2.1.8), β-xylosidases (EC 3.2.1.37), Endo-1, 4-β-mannanase (EC 3.2.1.78), β-mannosidase (EC 3.2.1.25) and α-arabinofuranosidases (EC 3.2.1.55), Acetyl xylan esterase (EC 3.1.1.72) and Ferulic acid esterase (EC 3.1.1.73) each acting on different components of hemicellulose. Hemicellulases are widely distributed across several glycoside hydrolase families, which include families 10, 11, 43, 51, 54, 62, 5, 26, 27, and 2. These enzymes often work synergistically with cellulases to enhance the breakdown of lignocellulosic biomass. Degradation of hemicellulose also exposes the cellulosic fraction and improves overall hydrolytic efficiency(Shallom and Shoham, 2003).

Lytic Polysaccharide Monooxygenases (LPMOs)

Lytic Polysaccharide Monooxygenases (LPMOs) are copper-dependent enzymes that enhance the degradation of polysaccharides cellulose, hemicellulose, chitin, and starch and pectin thereby improving biomass hydrolysis. Unlike classical hydrolases, LPMOs cleave glycosidic bonds via an oxidative mechanism, generating chain breaks in crystalline regions. They are classified under the auxiliary activity (AA) families in the CAZy database, mostly AA9 and AA10. In addition to polysaccharide oxidation, recent studies reveal LPMOs can enhance lignin degradation, suggesting a dual role. LPMOs significantly improved the efficiency of biomass saccharification and are now integral components of commercial enzyme cocktails for bioethanol production(Guo et al., 2022).

1.3.6 Mechanism of action

Biomass hydrolyzing enzymes act synergistically that involves the coordinated action of multiple enzymes to improve overall hydrolysis efficiency. This synergism is reflected in increased hydrolysis yields and reduced enzyme loading compared to individual enzymes acting alone, with the maximal effect often observed early in hydrolysis. The degradation process generally begins with hemicellulases and accessory enzymes that remove hemicellulose

and other side chains increasing accessibility to cellulose. Then, endoglucanases randomly cleave internal β-1, 4-glycosidic bonds in the amorphous region of cellulose, creating new chain ends. LPMOs assist by oxidatively cleaving crystalline regions, enhancing substrate accessibility. Cellobiohydrolases act on these ends to release cellobiose, which is subsequently hydrolyzed to glucose by β-glucosidases. This sequential and synergistic action prevents feedback inhibition (e.g., from cellobiose) and maximizes sugar release(Lynd et al., 2002).

1.3.7 Regulation of cellulases in filamentous fungi

Cellulase enzyme production is tightly regulated in filamentous fungi because their synthesis and secretion are energy-intensive processes. Therefore, fungi secrete these enzymes in the presence of cellulosic substrates in its environment given easily metabolizable sugars like glucose are absent. This ensures that the fungi do not waste valuable energy and resources by producing cellulases when easily metabolizable carbon sources are available(Amore et al., 2013). The mechanisms governing cellulase regulation been studied in numerous cellulolytic fungi mostly in *T.reesei* however the complete mechanism remains unresolved. In the following section, the known components of the regulatory mechanism will be discussed based on the current knowledge from the literature.

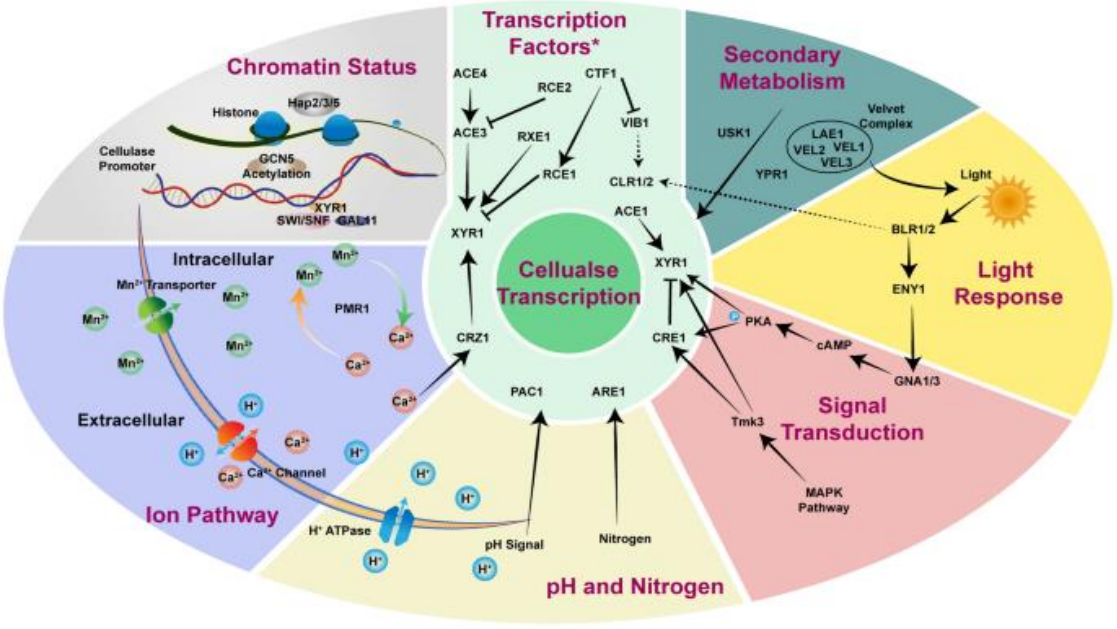


Figure 1.2 Representation of cellulase regulation mechanism in *T. reesei.* Adapted from (Yan et al., 2021b)

1.3.8 Role of transcription factors in cellulase regulation

Regulation of CAZymes in filamentous fungi has been studied largely at the level of transcription, leading to identification of numerous transcription factors involved in this process. It includes several pathway-specific transcription factors and additional factors associated with secondary metabolism, signal transduction pathways, and chromatin remodeling that has been shown to influence the transcriptional modulation of cellulase genes. A comprehensive list of transcription factors involved in the regulation of biomass utilization in filamentous fungi is shown in Table 1.1.

Table 1.1 List of transcription factors involved in the regulation of cellulase gene expression in filamentous fungi.

Transcription Factor	Key Organism(s)	Function/Key details in cellulase regulation	References
Xyr1/XlnR / XLR-1	*T. reesei* *A. niger* *N. crassa* *P. oxalicum*	Activator of cellulase and hemicellulase genes, Master regulator in *T. reesei*	(Klaubauf et al., 2014)
Clr-1/Clr-2 (ClrA/ClrB)	*N. crassa* *A. nidulans* *P. oxalicum*	Induction of cellulase genes and transporters Clr-1 activates Clr-2	(Coradetti et al., 2012)
Ace2	*T. reesei*	Specific activator of cellulases	(Aro et al., 2001)
Ace3	*T. reesei*	Activator of cellulases and hemi-cellulases Functions upstream of Xyr1	(Zhang et al., 2019)
Ace1	*T. reesei*	Repressor of cellulase/xylanase expression	(Aro et al., 2003)
Cre1 / CreA	*T. reesei* *A. niger* *N. Crassa* *P. oxalicum*	Mediates carbon catabolite repression (CCR) Repressor of cellulase genes in presence of glucose	(Campos Antoniêto et al., 2015)
Hap Complex (Hap2/3/5/ HapB/C/E)	*T. reesei* *A. nidulans* *N. crassa*	CCAAT-binding factor affects chromatin structure Modulates cellulase and xylanase gene expression	(Zeilinger et al., 2001)
PacC / Pac1	*A. nidulans* *T. reesei*	pH-responsive transcription factor	(Häkkinen et al., 2015)
Vel1/ VeA		Part of velvet complex Regulates development and is essential for cellulase expression	(Aghcheh et al., 2014)
VIB1	*T. reesei* *N. crassa*	Indirect activator Represses CCR	(Xiong et al., 2014)
Lae1/LaeA	*A. nidulans* *T. reesei*	Global methyltransferase regulator Controls secondary metabolism and cellulase gene expression	(Seiboth et al., 2012)
Xpp1	*T. reesei*	Repressor of xylanase-encoding genes	(Derntl et al., 2015)

Rxe1	*T. reesei*	Binds to the Xyr1 promoter Knockdown reduces cellulase gene expression.	(Wang et al., 2019)
BglR /Col-26	*T. reesei* *N. crassa*	Beta-glucosidase regulator	(Nitta et al., 2012)

XlnR is zinc binuclear cluster-type (Zn(II)2Cys6) transcription factor and it was first identified as the major activator for both cellulase and xylanase genes in *A. niger* (Van Peij et al., 1998). Subsequently, orthologs of this regulator were identified in various filamentous fungi, including Xyr1 in *T. reesei.* The deletion of Xyr1 has been shown to abolish the expression of the major cellulolytic and xylanolytic genes in *T.reesei* irrespective of the inducer used indicating its global effect on carbon utilization (Rauscher et al., 2006; Stricker et al., 2006; Mach-Aigner et al., 2008). However, Xyr1 is under the carbon catabolite repression mediated by Cre1 and also repressed by the specific transcription factor Ace1(Mach-Aigner et al., 2008). XlnR orthologs are present in most of the filamentous Ascomycoetes, but the set of genes controlled by this regulator is highly species specific and depends on the ecological niche and the lifestyle of the organism.

Carbon catabolite repression (CCR) is a conserved regulatory mechanism in fungi where the presence of a preferred carbon source like glucose suppresses the expression of genes involved in utilizing alternative carbon sources such as cellulose. In *Trichoderma reesei*, CCR is primarily mediated by the Cre1 transcription factor, a zinc finger protein that binds to specific DNA motifs (5′-SYGGRG-3′) upon phosphorylation in response to glucose. Cre1 negatively regulates key cellulase genes and by modifying or truncating Cre1 can alleviate repression and enhance cellulase expression even in the presence of glucose. Similar mechanisms are observed in other fungi, such as *Aspergillus* spp. (CreA) indicating a broadly conserved CCR strategy across filamentous fungi.

1.3.9 Role of sugar transporters in cellulase regulation

Sugar transporters play a vital and multifaceted role in regulating cellulase gene regulation in filamentous fungi by facilitating sugar uptake and participating in signaling pathways. Most of these tarnsporters come under Major facilitator superfamily (MFS) of transporters. In *Trichoderma reesei,* the cellobiose transporter CRT1 was identified as an importer of cellobiose, but later it was identified to be having dual function in transport and sugar sensing thus a “transceptor”(Havukainen et al., 2020). Other transporters including Stp1 in *T.*

reesei and CDT-1/CDT-2 in *Neurospora crassa*, contribute to cellodextrin uptake and the subsequent activation of cellulase genes(Zhang et al., 2013; Znameroski et al., 2013). These transporters regulate intracellular sugar concentrations, which are detected by transcriptional regulators such as XYR1. Additionally, sugar transporters influence CCR by modulating glucose sensing and the activity of repressors like CRE1, thus ensuring cellulase enzyme expression is appropriately controlled according to the type and availability of sugars. Collectively, sugar transporters function as critical modulators of cellulase production by integrating nutrient transport with signaling mechanisms(Mattam et al., 2022).

1.3.10 Role of signaling pathways in cellulase regulation

Signalling pathways play crucial role in cellulase regulation by coordinating environmental signals with transcriptional regulation. Recent advances have identified the involvement of multiple interconnected signaling pathways that sense environmental and intracellular cues and sunsequently modulate the cellulase gene expression.

G Protein-cAMP/PKA Pathway: In *T.reesei* GPCRs sense environmental signals such as carbon source and light and activate heterotrimeric G-protein subunits (GNA1, GNA3). This elevates cAMP through adenylate cyclase (ACY1). Protein Kinase A (PKA) responds to cAMP levels and phosphorylate of XYR1 leading to cellulase regulation(Schmoll and Hinterdobler, 2022).

MAPK (Mitogen-Activated Protein Kinase) Pathway: In *T.reesei*, three MAP kinases TMK1, TMK2, and TMK3 were shown to differentially influence cellulase regulation in a light-dependent manner(Schalamun et al., 2023).

Calcium (Ca^{2+}) Signaling Pathway: A cytosolic increase in Ca^{2+} level is shown to promote overexpression of *egl1* and *cbh1* significantly through the calcineurin-Crz1 **(**calcineurin-responsive transcription factor) pathway in *T.reesei*. (Liu et al., 2024).

Even though the component factors of various signaling pathways are known to affect cellulase expression, the precise molecular routes by which extracellular signals reach the nucleus and orchestrate nuclear responses remain partly unclear.

1.3.11 Influence of light on cellulase regulation in filamentous fungi

Cellulase genes regulated by environmental light conditions in in filamentous fungi such as *Trichoderma reesei* and *Neurospora crassa* through molecular mechanisms involving photoreceptors and signaling pathways. Photoreceptors BLR-1 and BLR-2 (Orthologs of WC-1

and WC-2 which forms White Collar Complex in *N.crassa*) binds to light-responsive elements in cellulase gene promoters, modulating their transcription based on light conditions in *T.reesei* (Froehlich et al., 2002; Schmoll, 2018). Also, the photoreceptor ENV1 (Ortholog of VVD in *N.crassa*) plays a role in photoadaptation, acting as a negative regulator to fine-tune cellulase production(Gyalai-Korpos et al., 2010). Moreover, light impacts cellulase regulation indirectly through crosstalk with heterotrimeric G-protein and cAMP signaling pathways(Schmoll, 2018). Not just the presence or absence of light but the difference in its intensity also plays a role in differential cellulase expression(Stappler et al., 2017). Such light-dependent regulation is conserved among filamentous fungi, offering an adaptive advantage by coordinating enzyme production with environmental light cues.

1.3.12 Strain improvemnt to improve cellulase production in filamentous fungi

Various strategies have been developed to enhance cellulase production in filamentous fungi. One effective approach is the genetic manipulation of transcription factors such as XYR1, CLR2, or ACE2, which regulate cellulase gene expression. Another widely used method involves alleviating carbon catabolite repression (CCR) by disrupting or modifying the CCR regulator Cre1 (or its homologs CreA/Mig1), allowing continued cellulase production even in the presence of glucose. Sugar transporter engineering, through overexpression of genes like *crt1* or *cdt-1*, can facilitate better uptake of inducers like cellobiose, which in turn enhances expression. Promoter engineering also improves production by replacing native regulatory sequences with stronger or inducible promoters. Additionally, strain improvement through random mutagenesis, adaptive evolution, or epigenetic regulation (e.g., via LAE1 manipulation) has yielded strains with superior enzyme secretion(Druzhinina and Kubicek, 2017). Protein and metabolic engineering further optimize cellulase activity and production by enhancing enzyme properties or redirecting cellular resources. Together, these strategies offer a comprehensive framework for optimizing cellulase yield in filamentous fungi.

1.3.13 *Penicillium janthinellum* NCIM1366

The development of *Penicillium janthinellum* 1366 strain was carried out at NCIM (National Collection of Industrial Microorganisms, CSIR-NCL Pune) through classical mutagenesis technique. The parental strain NCIM 1171 was exposed to ethyl methanesulfonate (EMS) followed by ultraviolet (UV) irradiation, resulting in strain EMS-UV-8 that showed notably

enhanced cellulase production, later named as NCIM 1366. These mutants demonstrated approximately double the filter paper (FPase) and carboxymethyl cellulase (CMCase) activities compared to the original strain, along with improved hydrolysis of Avicel in shake-flask experiments(Adsul et al., 2007). Given that the mutant strain generates cellulases effective in breaking down natural biomass substrates and that the cellulase system of this fungus remains relatively under-investigated, we have chosen to utilize it in our study.

Chapter 2

Cellulases from *P. janthinellum* and *T. reesei* – A comparative study

4.1 Introduction

Lignocellulosic biomass (LCB) primarily consists of three structural polymers—cellulose, hemicellulose, and lignin. Typically, cellulose is the major component, accounting for roughly 35–50 % of the plant's dry weight, followed by hemicellulose (~20–35 %) and lignin (~10–25 %) (Chen, 2014). A set of enzymes including cellulases, hemicellulases and ligninases, and other accessory enzymes generally carries out the degradation of lignocelluloses. Their relative proportions and quantities can determine the efficiency of biomass hydrolysis (Van Dyk and Pletschke, 2012). LCB is considered as one of the most promising renewable and sustainable feedstock for production of fuels and chemicals, but at the same time, its conversion is process intensive requiring a pretreatment to remove lignin /hemicellulose and hydrolysis using enzymes to recover fermentable sugars. The major cost involved in bioethanol production from lignocellulosic biomass is contributed by the hydrolysis step, owing to the cost of production of these enzymes (Johnson, 2016). Thus, reducing the cost of enzymes is crucial for making the process economical. Filamentous fungi are typically used for the production of these enzymes because of their ability to synthesize and secrete a wide array of plant cell wall-degrading enzymes (Su et al., 2012). The major sources of lignocellulose degrading enzymes include members of the fungal genera *Trichoderma*, *Aspergillus*, and *Penicillium* (Glass et al., 2013). With cellulose being the major component, cellulases play an important role in the degradation process. Cellulases are mainly classified into three major groups based on their mode of action on cellulose 1) endoglucanases, which randomly cleave internal β-1, 4 linkages in cellulose chain generating free ends 2) cellobiohydrolases, which act in a processive manner on either reducing or non-reducing ends of the cellulose chain, releasing cellobiose as the major product and 3) beta glucosidases that hydrolyze cellobiose into glucose (Zhang and Zhang, 2013).

Among the filamentous fungi, *Trichoderma reesei* is the most studied cellulase producer and is considered the model organism for cellulase research. It was originally isolated from the Solomon Islands during the Second World War, and the isolate was named *T. reesei* QM6a as part of the culture collection at the US Army Quarter Master Research and Development Center at Natick, Massachusetts. Attempts to improve the cellulase production from *T. reesei* QM6a by several generations of mutagenesis resulted in an enzyme hyper-secreting mutant named *T. reesei* RUT-C30 (Montenecourt and Eveleigh, 1979). Currently, *T. reesei* RUT-C30 is the predominantly used industrial cellulase producer, because of its ability to produce

a high titer of cellulases and a gene regulation mechanism that is highly adapted for cellulose utilization (Foreman et al., 2003), despite having a lower titer of beta glucosidases and much lesser number of CAZymes compared to certain other fungi. This is primarily because *T. reesei* produces the highest known titers of enzymes, the extracellular protein concentrations reaching as high as 100g/L (Cherry and Fidantsef, 2003) and there is a wealth of information accumulated on its genetics and gene regulation through works spanning several decades (Bischof et al., 2016). However, there are still efforts targeted at improving its enzyme production (Derntl et al., 2019; Chen et al., 2020) as the cost of cellulases cannot yet be considered economical for bio-refinery operations. Also, only a handful of companies are commercially manufacturing biomass hydrolysis enzymes; and the transportation and storage of the enzyme itself adds significantly to the cost (Ellilä et al., 2017). Availability of an alternate cellulase source, better than *T. reesei* RUT-C30 in terms of higher yields, a suitable ratio of different glucanases to effect improved biomass hydrolysis, and/or better production economics as determined by higher specific activities, shorter fermentation times etc., would be highly advantageous to the 2G-Ethanol industry, especially in countries where there is no commercial production of efficient biomass hydrolyzing cellulase. .

P. janthinellum NCIM 1171 is a filamentous soil fungus known for its efficient cellulase production and hydrolysis efficiency (Adsul et al., 2004, 2005). Classical mutagenesis studies conducted at CSIR-NCL, Pune had yielded three mutants of *P. janthinellum* NCIM 1171 with enhanced cellulase production and they were named EMS-UV-8 (NCIM1366), EU-21, and EU2D-21 (Adsul et al., 2007, 2009; Singhvi et al., 2011). The mutant EMS-UV-8, named as isolate NCIM 1366 was reported to have higher beta-glucosidase (BGL) activity compared to other mutants (Adsul et al., 2007). While its enzyme was not the best in terms of hydrolyzing pure cellulose (Adsul et al., 2007), it was efficient in the hydrolysis of pretreated rice straw (unpublished results). The strain was also used successfully for bioethanol production from pretreated wheat straw (Singhania et al., 2014). Since the mutant strain produced cellulases efficient in hydrolysis of natural biomass substrates and the cellulase system of the fungus is relatively unexplored, this chapter targeted to study the cellulases from *P. janthinellum* and compare it with the established cellulase hyper-producer *Trichoderma reesei* RUT-C30. Secretome analyses performed using Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) revealed a higher number of CAZymes in *P. janthinellum* compared to *T. reesei*, and a higher relative abundance of cellulases upon induction using cellulose, which may ex-

plain the higher activity and better biomass hydrolytic performance of enzyme preparation from the fungus.

4.2 Materials and methods

2.2.1 Microorganism and growth medium

Penicillium janthinellum NCIM1366 was kindly provided by the National Culture Collection of Industrial Microorganisms (NCIM), CSIR-National Chemical Laboratory, Pune, India and *Trichoderma reesei* RUT-C30 culture was a kind gift from Prof George Szakacs, Technical University of Budapest. The cultures were grown and maintained in Potato Dextrose Agar (PDA) slant. For enzyme production, spores were collected from 20-day-old PDA slants of *P. janthinellum* NCIM1366 and 5-day-old slants of *T. reesei* RUT-C30.

2.2.2 Enzyme Production

Cellulase enzyme production was carried out under submerged fermentation (SmF) using optimized media for both organisms. Both media were modified from the original Mandels and Weber medium (Mandels and Weber, 1969), and optimized for growth and cellulase production of the respective organisms. The enzyme production medium for *P. janthinellum* contained (in g/L): KH_2PO_4 (2.0), $CaCl_2.2H_2O$ (0.3), Urea (0.3), $MgSO_4.7H_2O$ (0.3), $(NH_4)_2SO_4$ (1.4), Peptone (0.75), Yeast extract (0.25), Tween-80 (0.5) and Trace elements: $FeSO_4.7H_2O$ (0.005), $MnSO_4.H_2O$ (0.0016), $ZnSO_4.7H_2O$ (0.0014), and $CoCl_2.6H_2O$ (0.002) with the pH of medium adjusted to 5.5. Cellulose (1% w/v) and wheat bran (2.5% w/v) were used as carbon sources and a spore suspension containing 1×10^5 spores/ml was used as inoculum at 1% (v/v) level. For *T. reesei*, the production medium contained (in g/L): KH_2PO_4 (2.0), $(NH_4)_2HPO_4$ (2.1), Yeast extract (2), NaCl (0.5), $CaCl_2.2H_2O$ (0.3), Urea (0.3), $MgSO_4.7H_2O$ (0.3), Tween 80 (0.5) and Trace elements: $FeSO_4.7H_2O$ (0.005), $MnSO_4.H_2O$ (0.0016), $ZnSO_4.7H_2O$ (0.0014) and $CoCl_2.6H_2O$ (0.002) with pH of the medium adjusted to 7.2. The carbon sources used were 0.1% Lactose, 2% cellulose, and 1.5% Wheat bran and the medium was inoculated at 1% (v/v) level with a 1×10^6 spores/ml spore suspension. Cultivation was carried out at 30 ± 2 °C and 200 rpm agitation. The culture fluid containing the secreted enzyme from both cultures was collected after 10 days of incubation by centrifugation (~13400 × g) and was assayed for total cellulase activity. Enzyme activity was expressed in Filter Paper Units (FPU) (Ghosh, 1987).

2.2.3 Comparison of total cellulase activity, endoglucanase activity, and beta-glucosidase activity

Both organisms were cultivated in the basic Mandels and Weber medium (Mandels and Weber, 1969) (Composition in g/L: KH_2PO_4 (2.0), $CaCl_2.2H_2O$ (0.3), urea (0.3), $MgSO_4.7H_2O$ (0.3), $(NH_4)_2SO_4$ (1.4), Peptone (0.75), yeast extract (0.25), Tween-80 (0.5) and Trace elements (g/L): $FeSO_4.7H_2O$ (0.005), $MnSO_4.H_2O$ (0.0016), $ZnSO_4.7H_2O$ (0.0014), and $CoCl_2.6H_2O$ (0.002) with pH of the medium adjusted to 5.0. Cellulose (1% w/v) was used as the sole carbon source. The inoculum size used was 1% v/v of a 1×10^5 spores/ml suspension for both cultures. Cultivation was carried out at 30 ± 2 °C and 200 rpm agitation. Samples were collected starting from the 2nd day of incubation, at 48-hour intervals. Total cellulase activity (Filter Paper Activity) and endoglucanase (CMCase) activity were determined by the IUPAC method (Ghose,1987) and beta-glucosidase activity was determined as described by Rajasree et al (Rajasree et al., 2013). One unit of beta-glucosidase activity was defined as the amount of enzyme required to liberate 1µg p-nitro phenol from pNPG (4-Nitrophenyl β-D- glucopyranoside) per milliliter.

2.2.4 Zymogram analysis for extracellular beta-glucosidases

Native Poly-Acrylamide Gel Electrophoresis (PAGE) of the extracellular enzymes was performed. After electrophoresis, the gels were washed with distilled water and incubated in a 10 mM solution of 4-Methylumbelliferyl β-D-glucopyranoside (MUG) in citrate buffer (0.05 M, pH 4.8) for 10 min at 50 ºC. The MUG-treated gels were visualized under UV light.

2.2.5 Comparison of hydrolytic efficiencies and Cellobiose accumulation

Different lignocellulosic biomass feedstock, pretreated similarly using either dilute acid or alkali, was used for the hydrolysis studies. For acid pretreatment, biomass (20% w/v) was mixed with 10% w/w of H_2SO_4 and was pretreated for 1h at 120 ± 2 °C. The biomass was cooled to room temperature and a slurry was made by adding 2× volume of water. The pH of the slurry was adjusted to 6.0 by adding 10N NaOH. Solid-liquid separation was performed using a nylon sieve and the biomass was washed twice with tap water. The biomass was used directly after correction of moisture, or air dried at room temperature and stored until used. For alkali pretreatment, 20% w/v rice straw and 10% w/w NaOH was mixed and pretreated at

120 ± 2 °C for 1h. Water (2× volume) was added and the pH of the pretreated slurry was adjusted to 6.0 by adding 10N H_2SO_4. The biomass was then processed as described above.

The hydrolysis reactions (20 ml reaction volume) were carried out in 100ml screw-capped glass conical flasks with the following conditions: 10% dry weight/w of pretreated biomass, 10 FPUs /g enzyme loading, and 0.05 % w/w surfactant (Tween 80). 0.5 % v/v of a commercial Penicillin /Streptomycin mixture (Himedia, India) was added to prevent any bacterial contamination. The hydrolysis reaction was carried out at 50 °C for 24 h and samples were collected at 0, 4, 8, 12, and 24 hours of hydrolysis. The amount of glucose, xylose, arabinose, mannose, and cellobiose in the samples was quantified by High-Perfomance Liquid Chromatography (HPLC) analysis using a Phenomenex Rezex® Pb-monosaccharide column. The analytical conditions were as follows: mobile phase-water; flow rate-0.6 mL/min; column temperature-80 ºC; detection-Refractive Index Detector (RID).

2.2.6 Comparison of extracellular protein production and secretome analysis

Both organisms were grown in basic Mandels and Weber medium (Mandels and Weber, 1969) with either glucose or cellulose at 1% (w/v) level as sole carbon sources, as described above. Samples were collected at 48-hour intervals, beginning on the 2nd day of incubation. The total secreted proteins present in the samples were estimated by the Bradford method (Bradford, 1976) with BSA as standard. SDS-PAGE was carried out as described by Laemmli (Laemmli, 1970), using 12% acrylamide gels. Proteins were visualized by staining with Coomassie Brilliant Blue R-250.

For secretome analyses, extracellular proteins were collected by centrifugation for 10 min at 4 °C and 13400×g after 10 days of incubation. The supernatants were further clarified by filtration through a 1µm Glass microfiber filter. The collected proteins were dialyzed against 50 mM ammonium bicarbonate buffer and concentrations were normalized. Samples were digested using trypsin following the standard protocol (Trauger et al., 2002). The proteomic profiling was performed in duplicates by Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) at the Mass Spectrometry & Proteomics Core facility of Rajiv Gandhi Centre for Biotechnology, Trivandrum, India. The protein samples were subjected to in-solution trypsin digestion using sequence-grade Trypsin (Sigma Aldrich, India). The LC/MS/MS analyses of the tryptic peptides were performed in a SYNAPT G2 High Definition Mass Spectrometer (Waters, UK), connected to a NanoACQUITY UPLC® chromatographic sys-

tem (Waters, UK) for the separation of the peptides. The LC-MS/MS acquired raw data was analyzed by Progenesis QI for Proteomics V3.0 (NonLinear Dynamics, Waters, UK) for protein identification using the protein database of *Trichoderma reesei* and *Penicillium* downloaded from the UniProt repository. Protein abundances were normalized across all the samples and Fold Changes (FC) were calculated as the raio of mean normalized abundances across the conditions. Prediction of the presence of secretion signal motifs was achieved using SignalP 5.0 (Almagro Armenteros et al., 2019).

2.2.7 Identification of CAZymes from secretome data

Annotation of the CAZy (Carbohydrate Active Enzymes) family of proteins as per the CAZy database was done through the dbCAN2 meta server (Yin et al., 2012).

2.3 Results and discussion

2.3.1 Comparison of extracellular protein production

Cellulose is a large polymer and utilization of it requires the secretion of enzymes by microorganisms to process it outside the cell, so that the simple sugars derived from its breakdown can be taken inside. The total secreted protein concentration in the presence of cellulose is indicative of the efficiency of the fungus in utilizing the polymer, as efficient cellulose digestion typically requires a milieu of different enzyme activities, in addition to cellulases. *T. reesei* showed a maximum protein secretion of 0.28 mg/ml on the 6th day of growth, whereas *P. janthinellum* secreted the maximum protein on the 10th day of growth, which was a ~1.8 time higher than *T. reesei* (Figure 2.1a). At all the time points tested, extracellular protein concentration was higher in *P. janthinellum*. SDS PAGE of the extracellular fractions from both fungi under un-induced (glucose grown) and induced (cellulose grown) conditions indicated a significant elevation in secreted proteins upon cellulose induction (Figure 2.1b). It was also observed that visibly, a greater number of extracellular proteins were secreted by *P. janthinellum*.

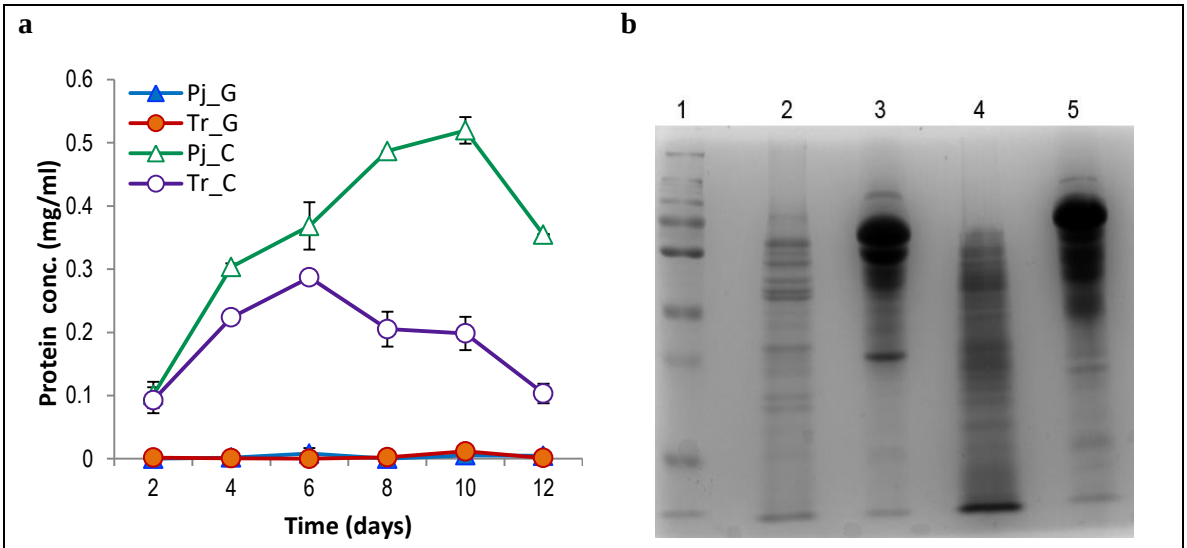


Figure 2.1 Extracellular protein productions by *T. reesei* and *P. janthinellum* in response to induction by cellulose.
a) Extracellular protein concentrations in *T. reesei* (Tr) and *P. janthinellum* (Pj) at different time points when cultivated in glucose and cellulose.
b) SDS-PAGE showing differences in protein production by the two fungi. Lane Information - 1) Page Ruler ® pre-stained protein ladder, 2) Tr grown in glucose 3) Tr grown in cellulose as sole C source 4) Pj grown in glucose 5) Pj grown in cellulose as sole C source.

2.3.2 Comparison of sugar release during hydrolysis

Experiments were designed to assess the efficiencies of both cellulase preparations for the hydrolysis of rice straw, pretreated using the most common methods of dilute acid or dilute alkali treatment at high temperatures. Both the dilute acid and dilute alkali pretreated rice straw were hydrolyzed better by *P. janthinellum* cellulases compared to enzymes from *T. reesei*, indicated by a significantly higher glucose release. At 24h, glucose release by *T. reesei* and *P. janthinellum* cellulases from acid-pretreated rice straw were 12.94 ± 0.8 g/L and 17.69 ± 0.47 g/L respectively, the latter showing a 37 % higher glucose release. Similar results were observed for alkali pretreated rice straw, where *P. janthinellum* enzyme released 27.24 ± 0.22 g/L of glucose which was 43% higher than the *T. reesei* cellulase. Also, the glucose release for both the acid and alkali-pretreated biomasses was higher with *P. janthinellum* cellulase at all the measured time points. Total sugar release from acid-pretreated rice straw, calculated as the sum of the concentrations of glucose, xylose, arabinose, and mannose in the hydrolysis mixture, was 17.88±0.5 g/L for *T. reesei* enzyme and 22.41 ±0.23 g/L for *P. janthinellum* enzyme. For alkali pretreated rice straw, the hydrolysate sugar contents were 29.49±0.57 g/L and 32.94±0.87 g/L respectively for *T. reesei* and *P. janthinellum* enzymes

(Figure 2.2a and Figure 2.2b). *P. janthinellum* enzyme released 25% higher glucose and 11% higher total sugars from alkali-pretreated rice straw sugars compared to *the T. reesei* enzyme.

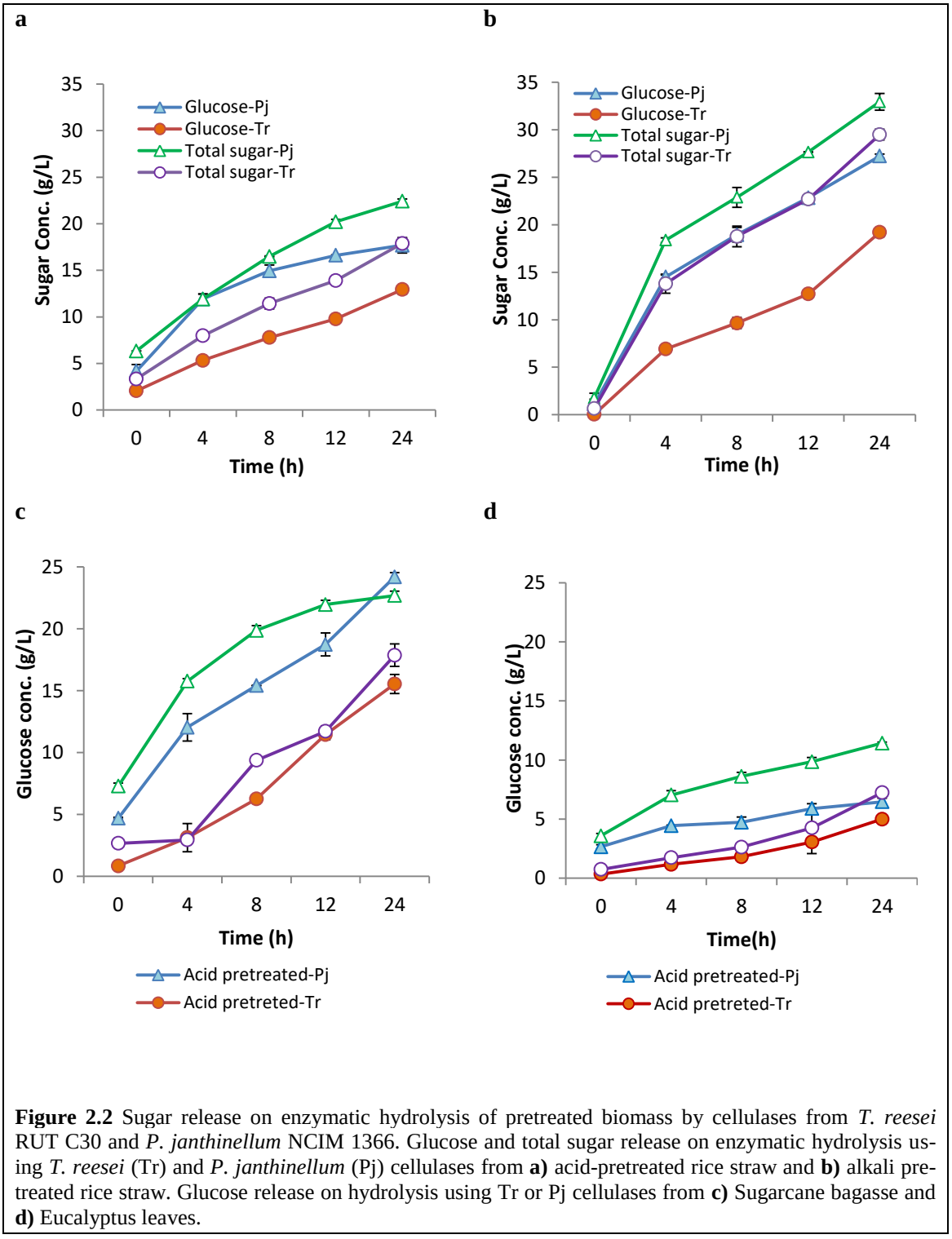


Figure 2.2 Sugar release on enzymatic hydrolysis of pretreated biomass by cellulases from *T. reesei* RUT C30 and *P. janthinellum* NCIM 1366. Glucose and total sugar release on enzymatic hydrolysis using *T. reesei* (Tr) and *P. janthinellum* (Pj) cellulases from **a)** acid-pretreated rice straw and **b)** alkali pretreated rice straw. Glucose release on hydrolysis using Tr or Pj cellulases from **c)** Sugarcane bagasse and **d)** Eucalyptus leaves.

The choice of substrate for the hydrolysis reaction often affects the efficiency of glucose release, and enzyme performance could be different on different biomass substrates. Two different biomasses with significantly different properties, viz. Sugar cane bagasse, and Eucalyptus leaves pretreated using acid or alkali, were used as substrates for testing the hydrolytic efficiency of enzymes from both fungi. Both the biomasses, regardless of the method of pretreatment were hydrolyzed better by *P. janthinellum* enzyme, indicated by higher glucose release. For acid-pretreated sugarcane bagasse, the glucose release at 24 h was 24.19±0.34 g/L for *P. janthinellum* enzyme and 15.54±0.76 g/L for *T. reesei* enzyme, while in the case of alkali-pretreated biomass it was 22.16±0.35 g/L and 17.87±0.91 g/L respectively for *P. janthinellum* and *T. reesei*. Similar results were obtained for eucalyptus leaves, for which the glucose release across treatments was less compared to other biomass types (Figure 2.2c and Figure 2.2d). The results are interesting for a "new" cellulase producer to outperform the established industrial producer irrespective of the substrate and the pretreatment method.

2.3.3 Comparison of cellobiose accumulation during hydrolysis

Cellobiose, the intermediate product of enzymatic cellulose hydrolysis, is produced through the action of exoglucanases and is the substrate for cellobiase/beta-glucosidase. Cellobiose accumulation can lead to product inhibition of upstream enzymes (endoglucanases and cellobiohydrolases) thus slowing down the whole hydrolytic process (Payne et al., 2015; Atreya et al., 2016). *T. reesei* is known to have limited cellobiase/beta-glucosidase (BGL) activity (Saloheimo et al., 2002); and while it may be advantageous for the organism in tight regulation of cellulose metabolism while growing on natural substrates, it is a disadvantage for the biomass hydrolyzing enzyme cocktails produced using the fungus. Often the *T. reesei* enzyme's lack of BGL activity is compensated by addition of BGL enzyme from other organisms. In this study, it was observed that the cellobiose accumulation in the hydrolysis mixtures was higher for the *T. reesei* enzyme compared to *P. janthinellum,* indicating an incomplete digestion due to the inherent low BGL activity of the former (Figure 2.3). For acid-pretreated rice straw, there was little or no cellobiose accumulation observed during the 24 h of hydrolysis, in the case of *P. janthinellum* enzyme, while about 5 g/L cellobiose was observed consistently from the 4th hour onwards in the case of *T. reesei*. In alkali-pretreated rice straw hydrolysis, the cellobiose concentration increased from 2.42 g/L in 4h to 7.45 g/L in 24 h for *T. reesei* while the maximum cellobiose accumulation in the case of *P. janthinellum* was only 0.76 g/L at 12 h after which it again decreased. Thus, unlike the *T.reesei* en-

zyme which has to be supplemented with external BGL for hydrolysis reaction (Berlin et al., 2007), *P. janthinellum* enzyme preparations may be used without the need for any blending or with minimal addition of synergistic BGL preparation (s).

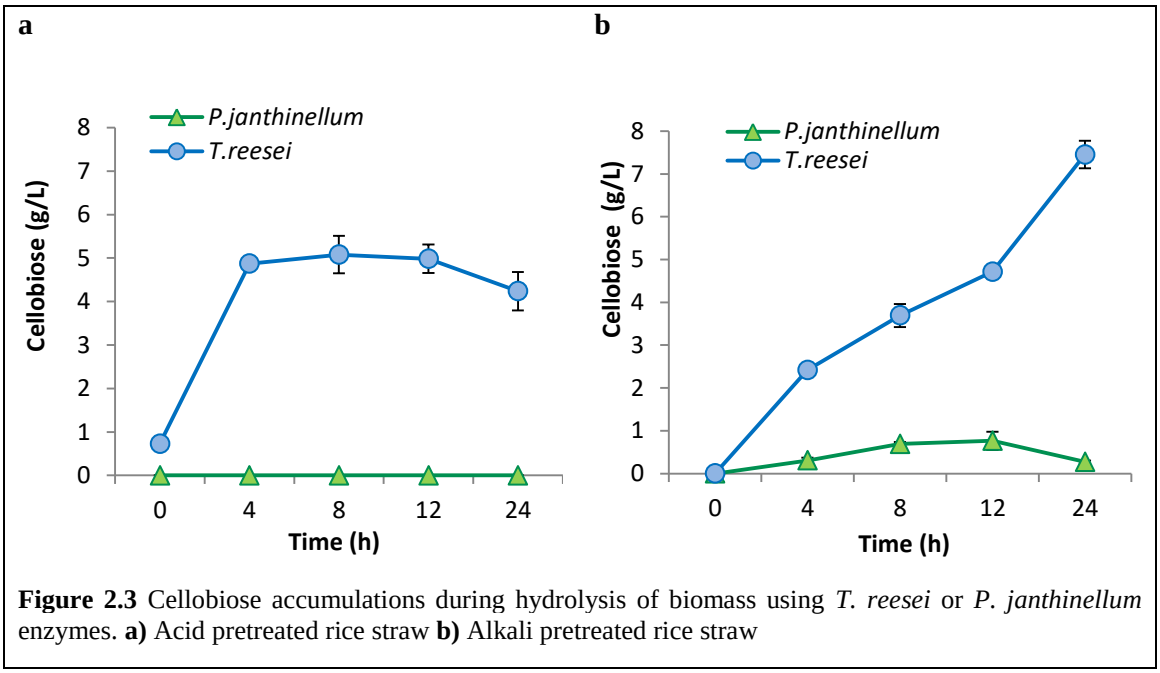


Figure 2.3 Cellobiose accumulations during hydrolysis of biomass using *T. reesei* or *P. janthinellum* enzymes. **a)** Acid pretreated rice straw **b)** Alkali pretreated rice straw

2.3.4 Comparison of cellulase enzyme activity

To know how each of the major components of the cellulolytic system contribute to the hydrolytic efficiency of *P. janthinellum* cellulase cocktail, standard cellulase assays were performed on the enzymes produced by the fungi. These included endoglucanases (EGs), cellobiohydrolases (CBHs), and β-glucosidases (BGLs), which act in synergism to hydrolyze cellulose (Zhang and Lynd, 2004). Since the parameters like media composition, pH, carbon source used, etc., can influence the quantity and variety of the cellulase components produced by fungi (Sukumaran et al., 2005), the basic mineral salts medium of Mandels and Weber (Mandels and Weber, 1969) with cellulose as the sole carbon source was used for both the organisms to obtain unbiased data. Extracellular enzyme production in this case was carried out using the same medium and under identical conditions of growth. Secreted enzymes from both fungi were analyzed for the total cellulase, endo-glucanase, and beta-glucosidase activities and the results indicated that all three major components in general exhibited increased activity with time, with peak levels observed on the 10th day of incubation, after which it decreased.

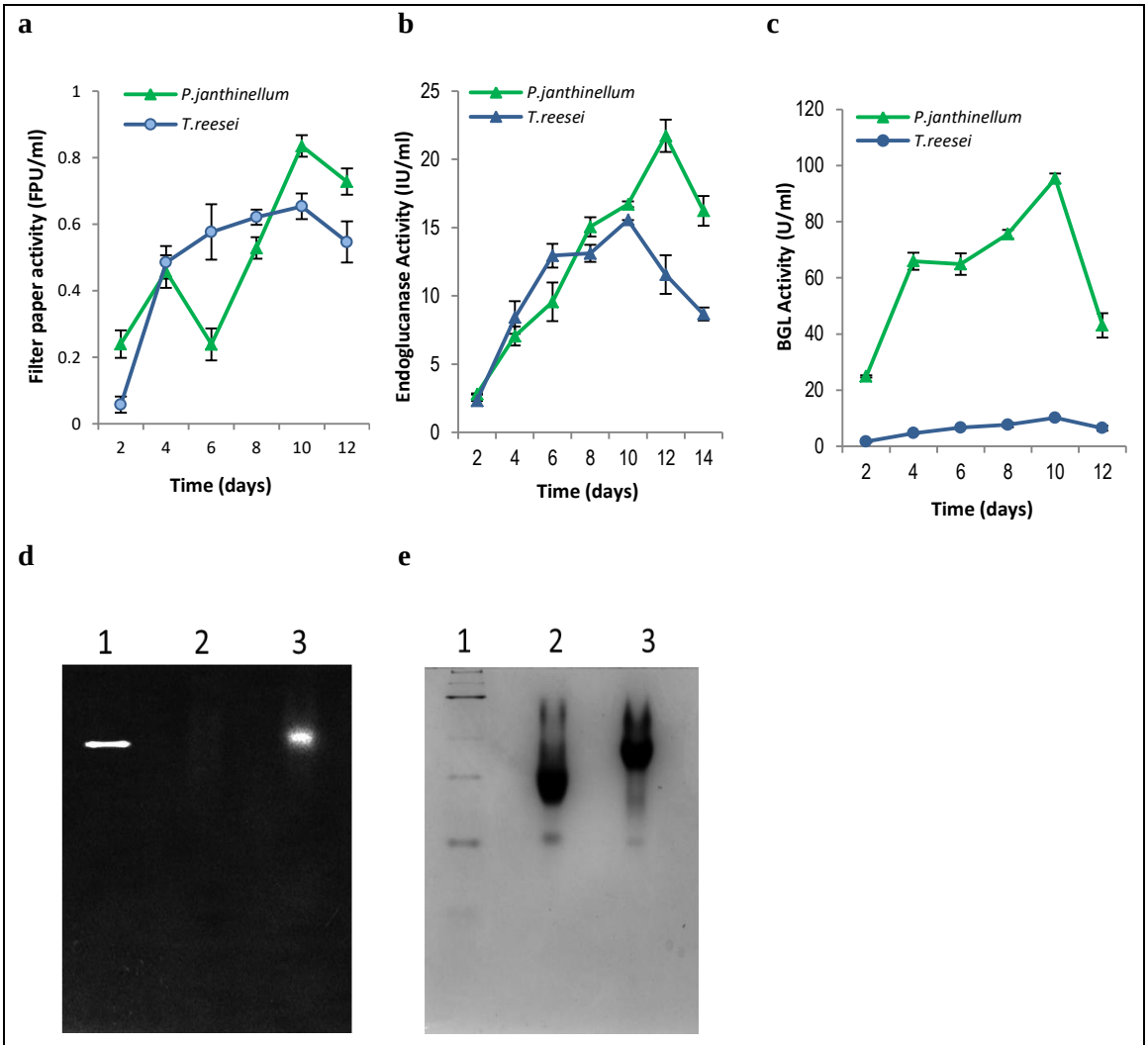


Figure 2.4 Extracellular cellulase production by *T. reesei* and *P. janthinellum* in Mandels and Weber medium with cellulose as sole carbon source. **a)** Total cellulase activity (Filter Paper Units), **b)** Endoglucanase activity (CMCase) **c)** Beta-glucosidase (BGL) activity **d)** Zymogram analysis showing BGL activity as MUG fluorescence **e)** Native PAGE showing protein profile of *T. reesei* and *P. janthinellum* secreted enzymes (Lane1-NativeMark™ unstained protein standard from Invitrogen, Lane2-*T. reesei* enzyme extract, Lane3-*P. janthinellum* enzyme extract for both **d** and **e**.

Both *T. reesei* and *P. janthinellum* showed maximum cellulase activity on the 10th day, but the FPAse activity of *P. janthinellum* (0.83 FPU/ml) was 28% higher than that of *T. reesei* (0.65 FPU/ml) (Figure 2.4a). Peak endoglucanase activity of 21.72 IU/ml was shown by *P. janthinellum* on the 12th day, whereas *T. reesei* showed maximum activity (15.55 IU/ml) on the 10th day (Figure 2.4b). *T. reesei* showed an endoglucanase activity of 12.93 IU/ml even on the 6th day, but the levels were raised only up to 15.55 IU/ml on the 10th day and were not sustained at further time points, probably indicating a feedback inhibition through glucose accumulation. *P. janthinellum* on the contrary, had a lower initial endo glucanase activity

(9.55 IU/ml) which steadily increased to 21.72 IU/ml on the 12th day and showed an ascending trend.

The largest difference in enzyme activity between the two fungi was observed in the case of beta-glucosidase (BGL) activity. The highest BGL activity in the case of *T. reesei* was 10.15 U/ml. *P. janthinellum* also showed the highest BGL activity on the 10th day but the activity was significantly higher at 95.42 U/ml compared to *T. reesei* (Figure 2.4c). Also, the fungus produced 24.88 U/ml activity on the 2nd day, whereas *T. reesei* could elaborate only 1.68 U/ml. These results are remarkable as the beta-glucosidase activity at peak levels by the two fungi is different by an almost 10-fold margin. Also, it becomes evident that the expression of BGL sets in early in the *P. janthinellum* which would allow it to hydrolyze cellulose faster and prevent cellobiose accumulation, which in turn may help to overcome an early setting in of feedback inhibition. The results were also confirmed by a Zymogram analysis which showed a prominent BGL activity band in *P. janthinellum*, whereas the *T. reesei* BGL band was barely visible (Figure 2.4d). The efficient removal of cellobiose from the reaction mixture by *P. janthinellum* can be attributed to the nearly 10-fold higher β-glucosidase activity exhibited by the fungus. Also, it may be speculated that the higher BGL activity may be one of the significant factors that can contribute to the higher hydrolytic efficiency shown by *P. janthinellum* cellulase. The results are indicative of an optimum enzyme cocktail from a single fungus that outperforms the conventional cellulase producer.

2.3.5 Comparative secretome analysis

Proteomic approaches have been widely used in filamentous fungi for the identification of both intracellular and extracellular proteins (Fernández and Novo, 2013). The genome of *T. reesei* QM6a, which is the parent strain of RUT-C30, was first sequenced in 2008 giving insight into its CAZyme system (Martinez et al., 2008). *T. reesei* is known to encode at least 10 cellulases, 16 hemicellulases, and a total of around 400 CAZymes in its genome. However, the composition of the secretome varies depending on the carbon source used, culture conditions, or experimental parameters. The first proteome analysis of *T. reesei* RUT-C30 identified a total of 22 proteins using lactose as a carbon source (Herpoël-Gimbert et al., 2008). Another study, using different carbon sources identified 230 extracellular proteins and 90 CAZymes (Adav et al., 2012).

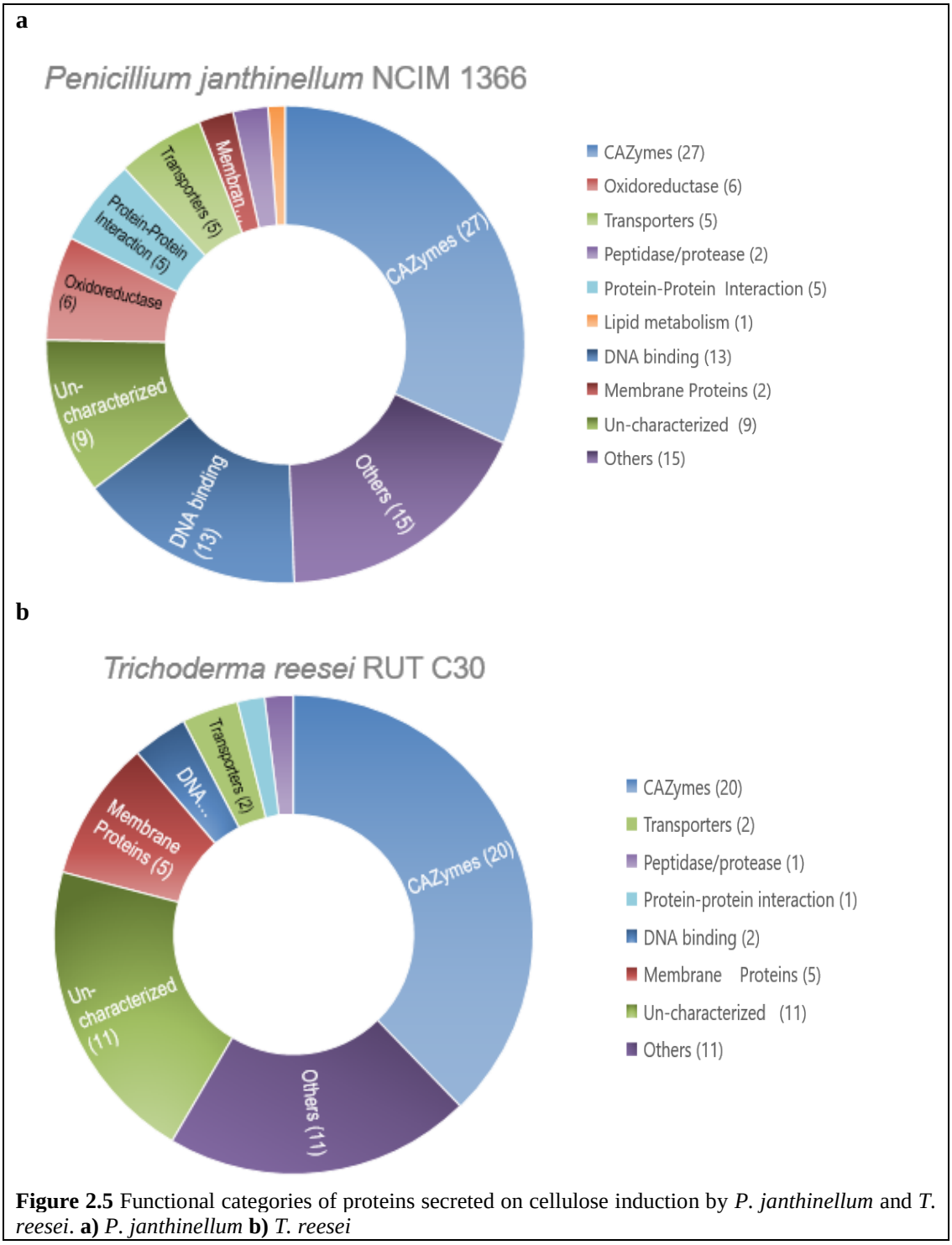


Figure 2.5 Functional categories of proteins secreted on cellulose induction by *P. janthinellum* and *T. reesei*. **a)** *P. janthinellum* **b)** *T. reesei*

In the present study, secreted proteins from both cultures (grown either with glucose as the carbon source or induced with cellulose) were identified and quantitatively analyzed using Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS). Samples were collected on the 10th day of growth, as both cultures exhibited peak filter paper activity on this day when incubated with cellulose, suggesting maximal secretion of the enzymatic repertoire at

this time point. Using a minimal mineral salt medium under identical conditions, a total of 53 proteins were identified in the *T. reesei* secretome, while *P. janthinellum* secreted 85 distinct proteins. Glucose was used as a non-inducing carbon source for comparative analysis. Annexure I and Annexure II lists all the proteins, their Uniprot accession number, molecular weights, number of unique peptides, normalized abundance in glucose and cellulose grown cultures, and fold change upon induction for *T. reesei* RUT-C30 and *P. janthinellum* NCIM1366 respectively. Our analysis detected a total of 53 proteins from *T. reesei* and 85 proteins from *P. janthinellum* in the 10th-day secretome. The distribution of proteins according to their biological function is shown in Figure 2.5. Among them, 27 proteins from *T. reesei* (51%) and 29 proteins from *P. janthinellum* (34%) were predicted to have an N terminal signal peptide using SignalP5.0 server. The identification of proteins without a signal peptide in the secretome could be indicative of the presence of cell lysis, cell death, or secretion through unconventional mechanisms (Miura and Ueda, 2018). Figure 2.6 shows the top 10 highly expressed proteins, as measured by the normalized abundance of their peptides in cellulose-induced cultures, compared to the control (grown in glucose). Most of the highly expressed proteins from both organisms were directly involved in the lignocellulose degradation.

Since it appeared from the forgoing studies that the reason for better overall cellulolytic activity and hydrolytic efficiency of *P. janthinellum* could be its secretion of a larger number of proteins, most of which are known to be involved in lignocellulose hydrolysis, it was speculated that the organism could elaborate more cellulolytic enzymes and/or accessory proteins compared to the industrial workhorse - *T. reesei*. An analysis of the CAZymes in the total secretomes was performed to understand their distribution in the extracellular proteins of both fungi. Among the 53 secreted proteins detected in *T. reesei*, 20 were identified as CAZymes (Figure 2.5a) and the number of CAZymes identified in the 85 identified secreted proteins of *P. janthinellum* was 27 (Figure 2.5b). As expected, most of the proteins identified from both the fungi were related to biomass degradation. Table 2.1 shows the list of CAZymes identified from the secretomes of *P. janthinellum* and *T. reesei*. The distribution of proteins among different CAZy families and the distribution of glycosyl hydrolase (GH) family proteins in both fungi are shown in Figure 2.7. CAZymes from *P. janthinellum* were mostly GH family proteins except one in the GT family. The CAZymes from *T. reesei* RUT-C30 were distributed to more CAZy families which included GH (Glycoside Hydrolases), CE (Carbohydrate Esterases), AA (Auxiliary Activities), and CBM (Carbohydrate Binding Module). In the case

of GH family proteins, *P. janthinellum* secretome had almost double the number of different Glycoside Hydrolases compared to *T. reesei*. GH family proteins from *P. janthinellum* spanned over 12 GH subfamilies while for *T. reesei* it was 11 GH subfamilies. GH subfamilies 5, 6, 7, and 11 were detected in both secretomes while GH subfamilies 3, 4, 16, 17, 30, and 72 were detected only in *T. reesei*, and GH subfamilies 2, 15, 27, 28, 36,43, 55 and 75 were detected only in *P. janthinellum*.

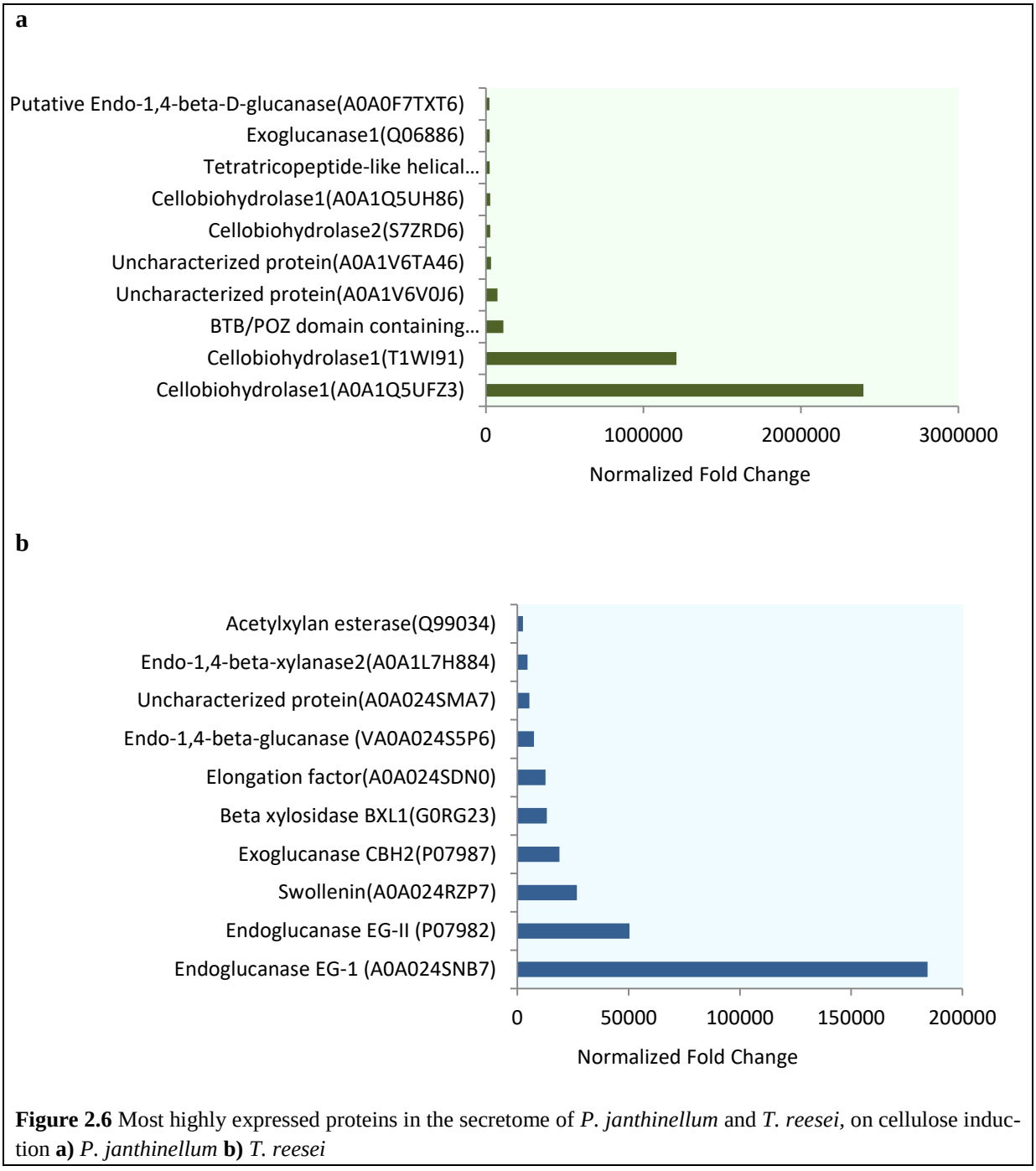
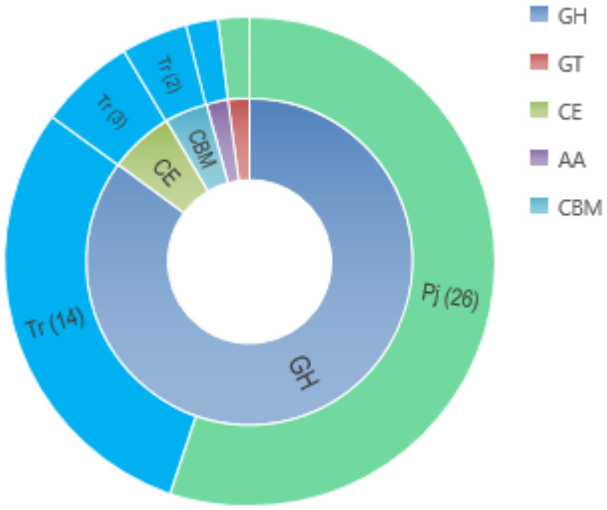


Figure 2.6 Most highly expressed proteins in the secretome of *P. janthinellum* and *T. reesei*, on cellulose induction **a)** *P. janthinellum* **b)** *T. reesei*

Among the CAZymes detected, a total of 17 enzymes which are directly involved in cellulose hydrolysis were detected, of which 3 were common to both fungi, which were cellobiohydrolase1 (CBH1) (Uniprot accession: P62694, A0A088DLG0), cellobiohydrolase2 (CBH2) (P07987, F1CHI2) and endoglucanase 1 (EG-1) (A0A024SNB7, A0A0F7TSC9). The two cellobiohydrolases showed higher abundance in *P. janthinellum* and the endoglucanase showed higher abundance in *T. reesei*. Apart from the EG-1, two other endoglucanases, EG-II (P07982) and EG-V (A0A024S5P6), are identified from *T. reesei*. In *P. janthinellum*, 10 cellobiohydrolases and 2 endoglucanases were identified from the common cellulases. There were 9 enzymes involved in hemicellulose degradation identified from the secretome of *T. reesei*, while 6 were identified in *P. janthinellum*, of which 2 were common with *T. reesei*. While *P. janthinellum* exhibited 10 times the BGL activity of *T. reesei*, no beta glucosidases were identified from both secretomes to support the extremely higher beta-glucosidase activity shown by *P. janthinellum*. Therefore, there was no way to confirm if the higher BGL activity obtained experimentally for *P. janthinellum* correlates to a higher amount of the corresponding protein in the secretome. It was previously observed that the BGL proteins had high specific activities and minute quantities can give high hydrolytic efficiencies, even though their proteins were undetectable by conventional means.

Chitin degrading enzymes were also identified from both secretomes but pectin degrading enzymes were identified only in *P. janthinellum*. Accessory activities known to aid cellulose hydrolysis in *T. reesei*, Swollenin (A0A024RZP7) and Lytic Polysaccharide Monooxygenase (A0A024SM10) were identified in the secretome of *T. reesei*, while these activities were not detected in *P. janthinellum*. In general, the relative abundance of most of the CAZymes was high in *P. janthinellum* compared to *T. reesei*, and one of the cellobiohydrolases (A0A1Q5UFZ3) from the GH7 family showed a very high relative abundance of 2.3 million. Though *P. janthinellum* did not show the accessory enzymes /activities in its secretome, it does not necessarily mean that the fungus lacks them, but they are not secreted under the tested conditions. However, it may be noted that multiple peptide tags may be matching the same *P. janthinellum* NCIM 1366 gene sequence in reality, which may not be captured on analyzing against the genome (s) of other *Penicillium* species' genomes as is the case here.

a



b

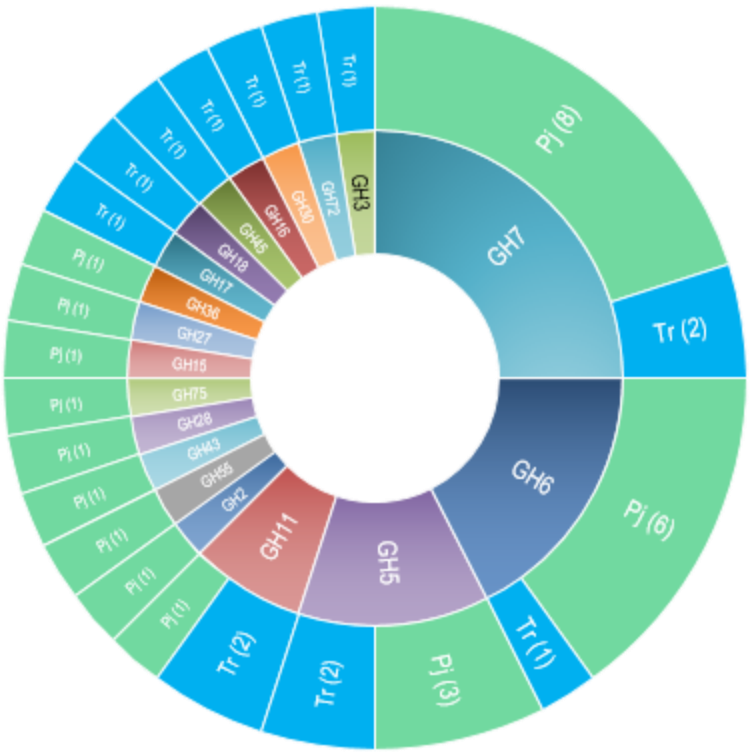


Figure 2.7 Distribution of secreted proteins of the fungi in different CAZyme families, and the glycosyl hydrolase family representation in the secretome. **a)** secreted protein distribution in CAZyme families **b)** distribution of glycosyl hydrolases in subfamilies. Tr – *T. reesei*, Pj – *P.janthinellum*

Table 2.1 CAZymes identified from the secretomes of *P. janthinellum* and *T. reesei.*

CAZy Family	Enzyme	UniProt Accession (s)	Tr-Glu	Tr- Cel	Tr-FC	Pj-Glu	Pj-Cel	Pj-FC	Presence
Cellulose hydrolysis/metabolism									
GH5	Endoglucanase EG-II	P07982	1.92	96905	50333	-	-	-	Tr
GH5	Endoglucanase B	A0A1Q5UIH7	-	-	-	14.5	1256	86	Pj
GH5	Putative Endo-1,4-beta-D-glucanase	A0A0F7TXT6	-	-	-	0	22950	22950	Pj
GH6	Exoglucanase 6A	A0A1Q5UI44	-	-	-	0	11744	11744	Pj
GH6	Exoglucanase CBH2	P07987, F1CHI2	13.6	256595	18844	0.24	4989	20734	Pj & Tr
GH6	Cellobiohydrolase	A0A1Q5UIA5	-	-	-	0.33	6916	20958	Pj
GH6	Cellobiohydrolase	A0A0F7TT91	-	-	-	-	-	-	Pj
GH6	Cellobiohydrolase	A0A1F5LUH2	-	-	-	-	-	-	Pj
GH7	Exoglucanase CBH1	P62694, A0A088DLG0	925	388093	419	0.13	2607	20054	Pj & Tr
GH7	Glucanase	S7ZRD6	-	-	-	0	28328	28328	Pj
GH7	Cellobiohydrolase1	A0A1Q5UFZ3	-	-	-	0.0036	8742	2396804	Pj
GH7	Cellobiohydrolase1	A0A1Q5UH86	-	-	-	0.36	10266	28516	Pj
GH7	Cellobiohydrolase1	T1WI91	-	-	-	0.00058	711	1225862	Pj
GH7	Exoglucanase1	Q06886	-	-	-	14.6	345207	23487	Pj
GH7	Cellobiohydrolase	A0A0F7TQ87	-	-	-	-	-	-	Pj
GH7	Endoglucanase EG-1	A0A024SNB7, A0A0F7TSC9	0.42	77136	183567	29.5	12605	426	Pj & Tr
GH45	Endo-1,4-beta-glucanase V	A0A024S5P6	3	22586	7407	-	-	-	Tr]
Hemicellulose hydrolysis /metabolism									
GH2	Beta mannosidase	A0A2H3IHX4	-	-	-	11.8	240	20	Pj
GH3	Beta xylosidase BXL1	G0RG23	2.8	37715	13278	-	-	-	Tr
GH5	Mannan endo-1_4-beta-mannosidase	Q99036, A0A1Q5SW99	12.8	1387	107	24.4	1891	77	Pj & Tr
GH11	Endo-1,4-beta-xylanase2	A0A1L7H884, A0A088S933	0	4530	4530	20	833	40	Pj & Tr
GH11	Endo-1,4-beta-xylanase1	A0A2H3A3Q5	110	32909	299	-	-	-	Tr
GH17	Beta-1,3-endoglucanase	A0A024SAF4	385	264	1.4	-	-	-	Tr
GH27	Alpha-galactosidase	A0A0F7KIJ2	-	-	-	0	5173	5173	Pj
GH30	Endo beta 1,4-xylanase4	A0A024RV01	34.5	3052	88	-	-	-	Tr
GH43	Putative exo-beta-1,3-galactanase	S7ZCW4	-	-	-	78.2	1900	24	Pj
GH55	Exo-beta-1,3-glucanase	S8BDR6	-	-	-	117	13.9	8.4	Pj
CE05	Acetyl xylan esterase	Q99034	19.5	49211	2522	-	-	-	Tr
CE16	Acetyl esterase	A0A024SFG8	10.8	24711	2273	-	-	-	Tr
CE16	Acetyl esterase	A0A2H3A8G3	18.6	1573	84	-	-	-	Tr
Chitin hydrolysis/metabolism									
GH18	Chitinase2	A0A024S0K1	11.2	597	53	-	-	-	Tr
GH75	Endo-chitosanase	A0A1Q5UKX2	-	-	-	422	1340	3.1	Pj

CAZy Family	Enzyme	UniProt Accession (s)	Tr-Glu	Tr- Cel	Tr-FC	Pj-Glu	Pj-Cel	Pj-FC	Presence
GH84	Putative bifunctional alpha-glucuronidase/ N-acetyl beta-glucosaminidase	S7ZNU5	-	-	-	534	1572	2.9	Pj
Starch hydrolysis/metabolism									
GH15	Glucoamylase	A0A093V3D6	-	-	-	223	0	0.0044	Pj
Pectin									
GH28	Endopolygalacturonase	A0A1Q5UPK8	-	-	-	56	3631	64	Pj
Accessory Activities/Other Enzymes									
AA9	Polysaccharide monooxygenase CEL61A	A0A024SM10	2.85	2817	987	-	-	-	Tr
CBM1	Swollenin	A0A024RZP7	0.75	20249	26748	-	-	-	Tr
CBM1	CBM1 domain-containing protein	A0A024SJK4	104	3661	35	-	-	-	Tr
GH16	Glucanosyl transferase	A0A024S7E6	787	2014	2	-	-	-	Tr
GH36	Putative galactinol sucrose galactosyl transferase	A0A1V6YI05	-	-	-	0	2804	2804	Pj
GH72	1_3-beta-glucanosyltransferase	A0A024RXP9	7.9	3225	405	-	-	-	Tr
GT31	Glycosyl transferase	B8MG06	-	-	-	20	825	40	Pj

Tr, *Trichoderma reesei* RUT C30, Pj, *Penicillium janthinellum* NCIM 1366,
Tr-Cel, Pj-Cel, *T. reesei,* or *P. janthinellum* grown with cellulose as carbon source (induced),
Tr-Glu, Pj-Glu, *T. reesei,* or *P. janthinellum* grown with glucose as a carbon source (uninduced)
Tr-FC, Pj-FC, Fold change in normalized protein abundance between induced and uninduced in *T.reesei* or *P. janthinellum*

The proteins identified from the secretome may not be a complete representation of all the CAZymes secreted by the organism, as the study used only a single time point and pure cellulose as the sole carbon source. The highest differentially expressed protein from *T. reesei* was the GH7 family endoglucanase EG-1, which showed an 183567-fold increase in expression upon cellulose induction. However, CBH1 is known to be the major secreted protein of *T. reesei* on cellulose induction (Peterson and Nevalainen, 2012). The difference in this study might be a result of the culture conditions and/or the time point of analysis. It could also result from the processing of samples where the insoluble cellulose fraction, which could bind the enzyme, was removed to obtain the supernatant used for analyses. The normalized fold difference shown by the most highly expressed protein from *P. janthinellum* was almost 2 million, and this was a cellobiohydrolase from the GH7 family. While *T. reesei* secreted a wider variety of enzymes involved in lignocellulose hydrolysis, it was the *P. janthinellum*

that secreted more glycosyl hydrolases and especially very high levels of exoglucanases. The study provides preliminary information on the presence of all major cellulolytic and hemicellulolytic activities in the fungus and a very high induction in the presence of cellulose, which could account for its enhanced hydrolytic performance.

2.4 Conclusions

Here, we provide the first-ever secretome analysis of *Penicillium janthinellum* NCIM1366 and its comparison with the established cellulase hyper-producing industrial strain-*Trichoderma reesei* RUT-C30. The analyses have highlighted the better hydrolytic efficiency, enzyme activity, protein production, and secretion efficiency of *P. janthinellum*, which indicates its potential as a future industrial cellulase producer. Further exploration and a deeper understanding of the reasons for its better cellulase production warranted genome and transcriptome level studies on the fungus, which was undertaken in the upcoming chapters. Further, targeted genetic modifications are expected to improve its performance even more, providing a worthy alternative for *T. reesei*, or complement it in the cellulase applications for biomass conversion.

Chapter 3

Whole-genome sequencing of *Penicillium janthinellum* NCIM1366 and identification of genes involved in cellulose metabolism

3.1 Introduction

Advancement of high-throughput and cost-effective sequencing technologies has enabled the generation of an extensive number of genome sequences across all domains of life. Unlike in the past, when genomic information on filamentous fungi was scarce, the last two decades have seen a remarkable surge in genome data from these organisms, largely driven by various fungal genome initiatives (Grigoriev et al., 2014). *Neurospora crassa* was the first filamentous fungus to have its entire genome sequenced, marking a significant milestone in the field of fungal genomics. (Galagan et al., 2003). The availability of genome sequences has facilitated in-depth studies of industrially important filamentous fungi and has supported the development of genetic engineering techniques tailored for these organisms.

The genome sequence of the cellulolytic model organism *Trichoderma reesei* QM6a, made available in 2008, provided valuable insights into the molecular basis of its efficient biomass hydrolysis (Martinez et al., 2008). Unexpectedly, despite the industrial relevance and effectiveness of the carbohydrate-active enzymes of *T. reesei*, its genome encodes fewer cellulases and hemicellulases than those of other biomass-degrading fungi. However, many *T. reesei* genes encoding carbohydrate-active enzymes are organized non-randomly into clusters situated between regions of synteny with other Sordariomycetes. This genomic arrangement enables coordinated regulation of these enzymes, contributing to the fungus's high hydrolytic efficiency. Genome comparison of the mutant *T. reesei* RUT-C30 with the wild-type QM6a revealed a truncated *cre1* gene, which accounts for its enhanced cellulase production by partially relieving carbon catabolite repression (CCR) in the presence of glucose (Seidl et al., 2008). The availability of the genome sequence facilitated a shift from traditional random mutagenesis to targeted genome engineering of key regulators such as *cre1* and *xyr1*, enabling precise and efficient manipulation of the organism's cellulase production pathway (Kubicek et al., 2009).

As demonstrated by the studies described in Chapter 1, *P. janthinellum* NCIM 1366 exhibits significantly higher enzyme activities and superior biomass conversion under the tested conditions, compared to *T. reesei* RUT-C30. Sequencing the genome of *P. janthinellum* NCIM 1366 is essential for analyzing the repertoire of its biomass-degrading enzymes and elucidating the mechanisms underlying cellulase regulation, thereby uncovering the basis for its enhanced activity. Furthermore, this genomic information will facilitate the development of a

genetic toolbox for *P. janthinellum*, enabling future genetic engineering strategies to improve enzyme production and strengthen its potential as an industrial cellulase producer.

Here, we report the whole-genome sequence of *Penicillium janthinellum* NCIM 1366. The genome reveals a richer repertoire of carbohydrate-active enzymes (CAZymes) compared to other well-studied cellulase producers within the *Penicillium* and *Aspergillus* genera, as well as *Trichoderma reesei*. In addition, we have identified key lignocellulose-degrading enzymes and homologs of known regulators of cellulase expression from other filamentous fungi, which may serve as promising targets for genome engineering in this organism. These findings expand our understanding of the genomic landscape of *P. janthinellum*, offering valuable insights into the mechanisms of cellulase production and supporting future efforts to enhance this strain for industrial applications.

3.2 Materials and methods

3.2.1 Organism and culture conditions

P. janthinellum culture was maintained in Mandels and Weber medium (Mandels and Weber, 1969) supplemented with 1% cellulose as the sole carbon source to maintain its cellulase activity. For genomic DNA isolation, the fungus was cultured in PDA (Potato Dextrose Agar) slants for 15 days and was used for spore collection.

3.2.2 Genomic DNA isolation

Spores were collected from 15-day old PDA slants of *P. janthinellum* and inoculated into 100 ml PDB medium. Mycelia were collected after 3 days of growth and were used for Genomic DNA isolation using GeneJET® Plant Genomic DNA Purification Kit (Thermo Scientific™). The concentration and quality of the DNA were determined using a Nanodrop® spectrophotometer and 1% agarose gel, respectively.

3.2.3 Whole genome sequencing and assembly

Genome sequencing was carried out using the Next Generation Sequencing Technology (Illumina HiSeq 2500 platform). Cutadapt v1.8 was used to remove the adapter sequences from the Illumina reads. All low-quality data (Q<30) were filtered out using Sickle v1.33. And the reads that are of length less than 30 bases were discarded. Further, duplicate reads were removed using FastUniq.

The cleaned reads were subjected to Kmergenie (Chikhi and Medvedev, 2014) to predict the optimal k-value and assembly size. De novo assembly was performed using MaSuRCA(Zimin et al., 2013), SPades (Bankevich et al., 2012), Velvet (Zerbino and Birney, 2008), and SOAPdenovo2 (Luo et al., 2012). The default k-mer sizes were used for MaSuRCA and SPades assembly, and the k-mer value predicted by kmergenie was used for Velvet and SOAPdenovo2 assembly. After the assembly generation, we have checked the presence of conserved genes in the assembled contigs using BUSCO v2 (Simão et al., 2015). The MaSuRCA assembly was used for all further downstream analysis.

3.2.4 Gene prediction and annotation

Genes were predicted from the MaSuRCA assembled contigs using Augustus 3.1(Stanke et al., 2008) and the resulting coding sequences were translated into predicted protein sequences. The predicted proteins were annotated by comparison with the UniProt Fungi database using the BLASTP program, and matches with E-value <= 10^{-5} and similarity score >= 40% were retained for further annotation. From the top BLASTP hit of each gene, the organism name was extracted for organism annotation. The predicted proteins from Augustus were annotated against UniProt, Pathway, and other databases. The gene ontology (GO) terms for predicted proteins were extracted wherever possible. The tRNA genes were predicted from the contigs using tRNAscan-SE (Lowe and Eddy, 1997).

3.2.5 Identification of CAZymes and key cellulase genes

Annotation of CAZy (Carbohydrate Active Enzymes) family of proteins was done from predicted amino acid sequences as per the CAZy database (Lombard et al., 2014) using the dbCAN2 meta server (http://bcb.unl.edu/dbCAN2) (Yin et al., 2012), which also predicted the secretory signals. Cellulase genes and other regulatory genes were identified from Uniprot annotation data. CAZyme data for *T. reesei*, *A. niger*, *A. nidulans*, *P. chrysogenum*, *P. brevicompactum*, and *P. oxalicum* were retrieved from the CAZy database.

3.3 Results and discussion

3.3.1 Genome sequencing, assembly, and annotation

The whole-genome shotgun sequence of *Penicillium janthinellum* NCIM1366 was generated through the Illumina HiSeq NGS platform and the complete sequence information has been deposited at GenBank under the accession number NPFE00000000.1. The important features of the genome sequence are shown in Table 3.1. The 37.6 Mbp genome of *P. janthinellum* was assembled into 1773 contigs with an N50 value of 71.6 Kb. From the assembled contigs, 12003 genes were predicted. Among them, 11532 (96.08%) proteins showed a significant BLASTP match (E-value <=1e-5 and Similarity score >=40%). The top BLASTP hit of each gene was studied, and the organism name was extracted. The 15 organisms that most frequently appeared among the source organisms of the homologs are shown in Figure 3.1. The majority of the top BLASTP hits were associated with *Penicillium brasilianum* (approximately 8,000 proteins), suggesting a close evolutionary relationship between these species.Functional annotation of the predicted protein-coding genes revealed that approximately 7579 genes (63% of the total) were assigned Gene Ontology (GO) terms. The total number of different GO (Gene Ontology) terms identified in molecular function, biological process, and cellular component categories is provided in Table 3.2.

Table 3.1 Features of the *Penicillium janthinellum* NCIM 1366 Whole-Genome Sequence

Genomic feature	Value
Genome size	37.6Mbp
Genome coverage	100x
GC content	50.72 %
Total number of contigs	1773
Largest contig	276987 bp
N50	71.6 Kb
Number of predicted Proteins	12003
Number of genes with a significant BLASTP hit and UniProt information	11532 (96.08%)
Number of tRNA genes	206

The top 10 terms in each category are shown in Figure 3.2. Among the 1992 GO terms identified within the Cellular Component category, the majority of genes were associated with the 'integral component of membrane,' reflecting a high prevalence of membrane-bound trans-

porters. This may indicate an adaptation to facilitate the efficient uptake and processing of biomass-derived substrates.

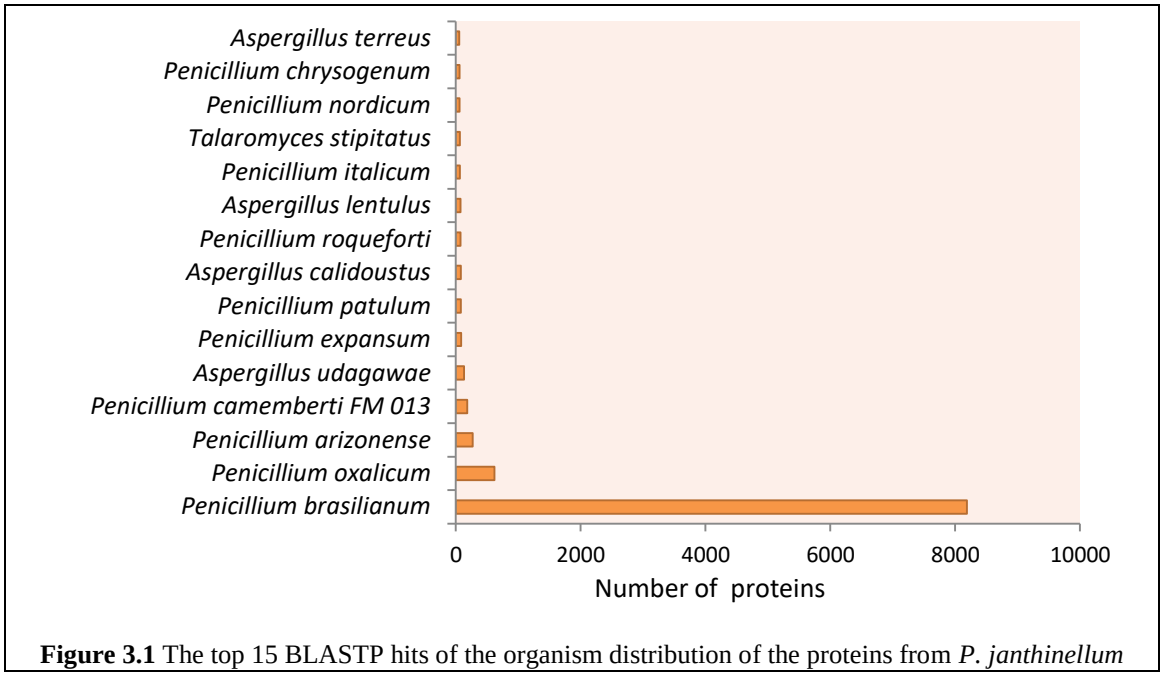


Figure 3.1 The top 15 BLASTP hits of the organism distribution of the proteins from *P. janthinellum*

The second most abundant category corresponded to transcription-related components, suggesting a substantial representation of transcription factors in the genome (Figure 3.2a). A similar pattern was observed for the Biological processes category, which encompassed a total of 2866 terms. Here, the most prominent category was 'transmembrane transport', followed by 'zinc ion binding' and the third being 'carbohydrate metabolic processes' (Figure 3.2b), highlighting its cellulolytic lifestyle. Interestingly, 'small GTPase-mediated signal transduction' was identified as a major category within the Biological Process annotations, which may be indicative of their involvement in cellulose sensing and signal transduction pathways, analogous to their established roles in cellulase regulation observed in other filamentous fungi such as *T. reesei* (Li et al., 2023). In the Molecular functions category also, the terms 'DNA binding' and 'Zn ion binding' were prominent, further underscoring the significant representation of transcription factors and regulatory proteins (Figure 3.2c).

Table 3.2 The total number of Gene Ontology different terms identified from the genome of *P. janthinellum*

Category	Number of terms
Cellular Components	1992
Biological Processes	2866
Molecular Functions	2721

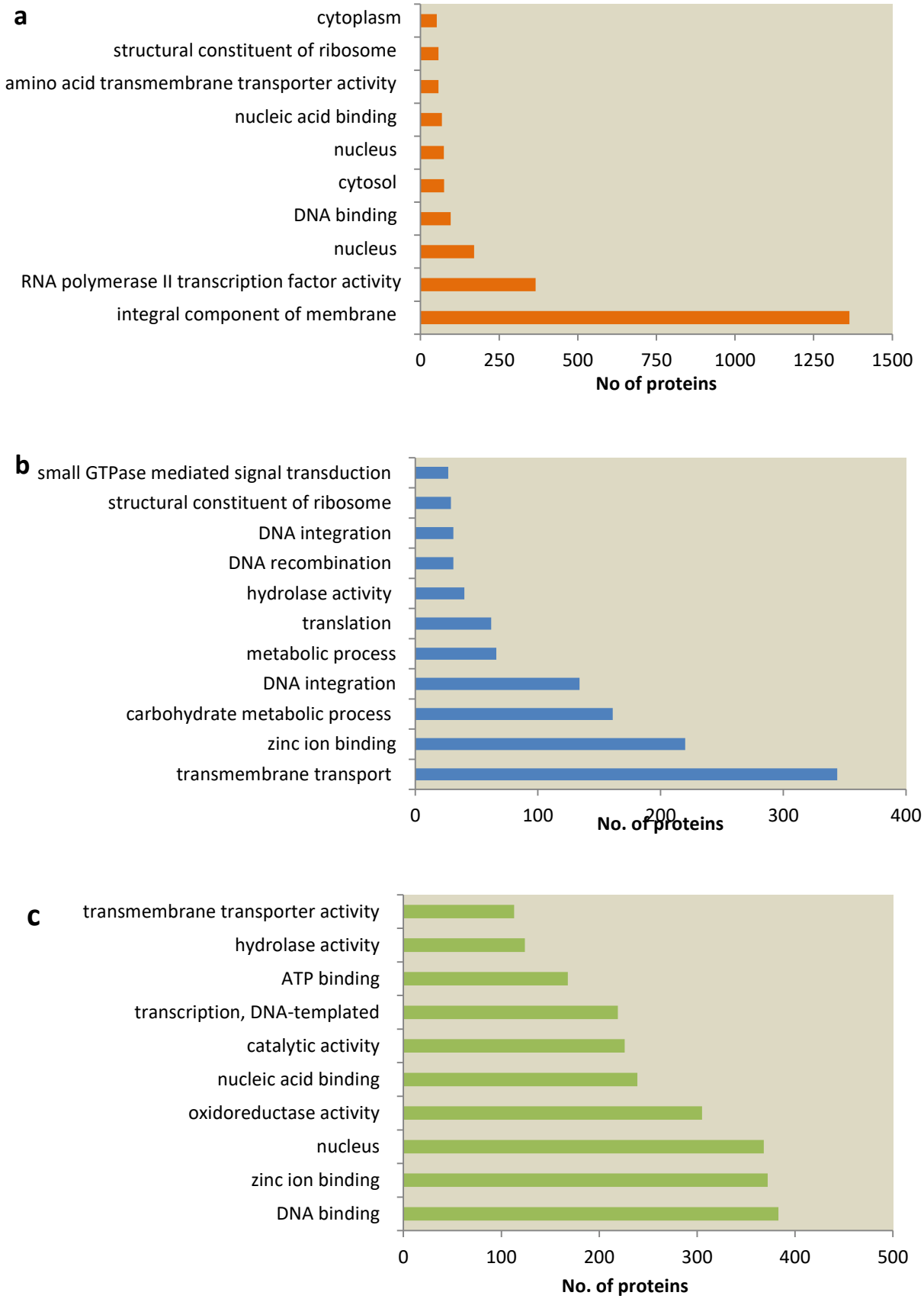


Figure 3.2 Top 10 terms in GO annotation category of *Penicillium janthinellum* NCIM1366: **a)** cellular components**, b)** biological processes, **c)** molecular functions

3.3.2 Identification of CAZymes

Carbohydrate Active Enzymes (CAZymes) were identified from the genome according to the CAZy database using the dbCAN tool with the criteria 'number of tools ≥ 2'. The genome showed enrichment in enzymes that act on carbohydrates, with a total of 657 CAZymes, which comprise 5.4% of all predicted genes. This percentage exceeds the typical CAZyme content observed in other cellulolytic fungi, which generally ranges from 2–4% (Ferreira Filho et al., 2017; Li et al., 2017), suggesting an evolutionary enrichment of CAZyme genes in its genome that may be critical for supporting its specific lifestyle or ecological niche. SignalP analysis showed the presence of Signal Peptide in 269 proteins, indicating their extracellular location. The distribution of CAZymes by family is presented in Figure 3.3. This comprised 391 Glycoside Hydrolases (GHs), 86 Auxiliary Activity enzymes (AAs), 8 Carbohydrate-Binding Modules (CBMs), 44 Carbohydrate Esterases (CEs), 114 Glycosyltransferases (GTs), and 14 Polysaccharide Lyases (PLs). Figure 3.4 illustrates the distribution of CAZymes across various subfamilies within each enzyme family.

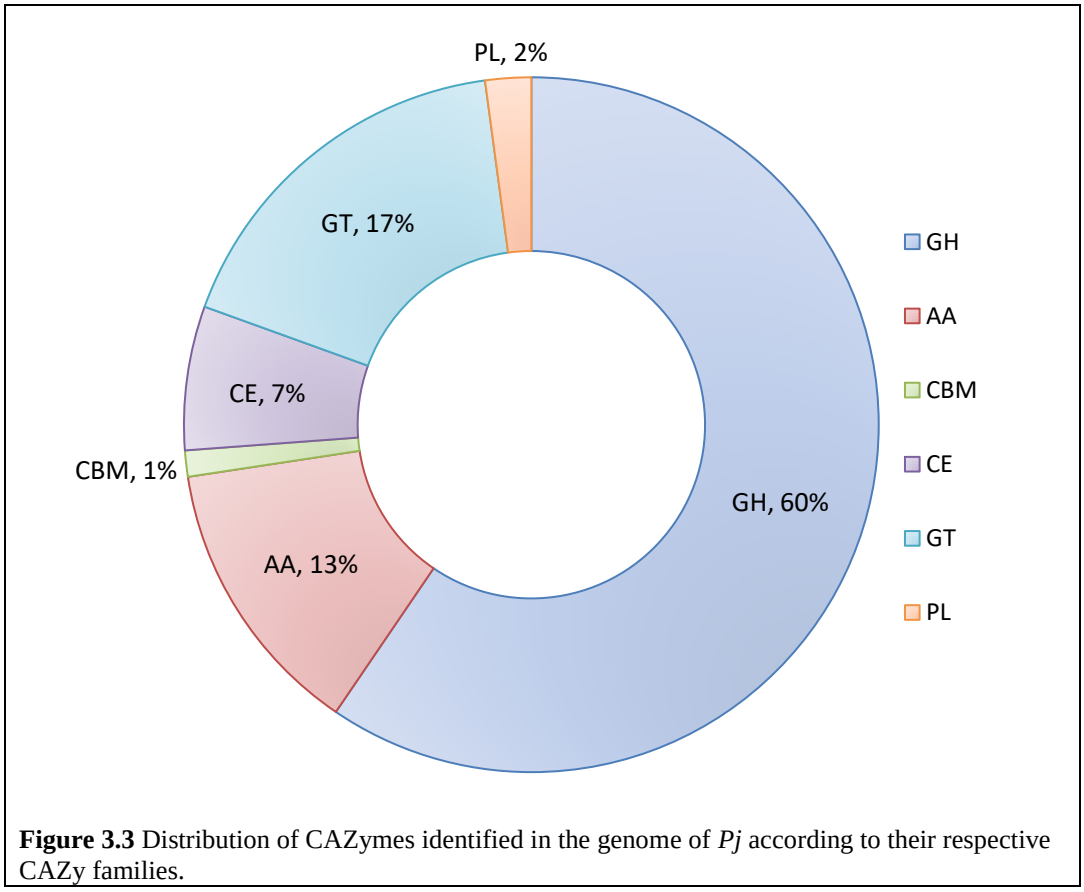
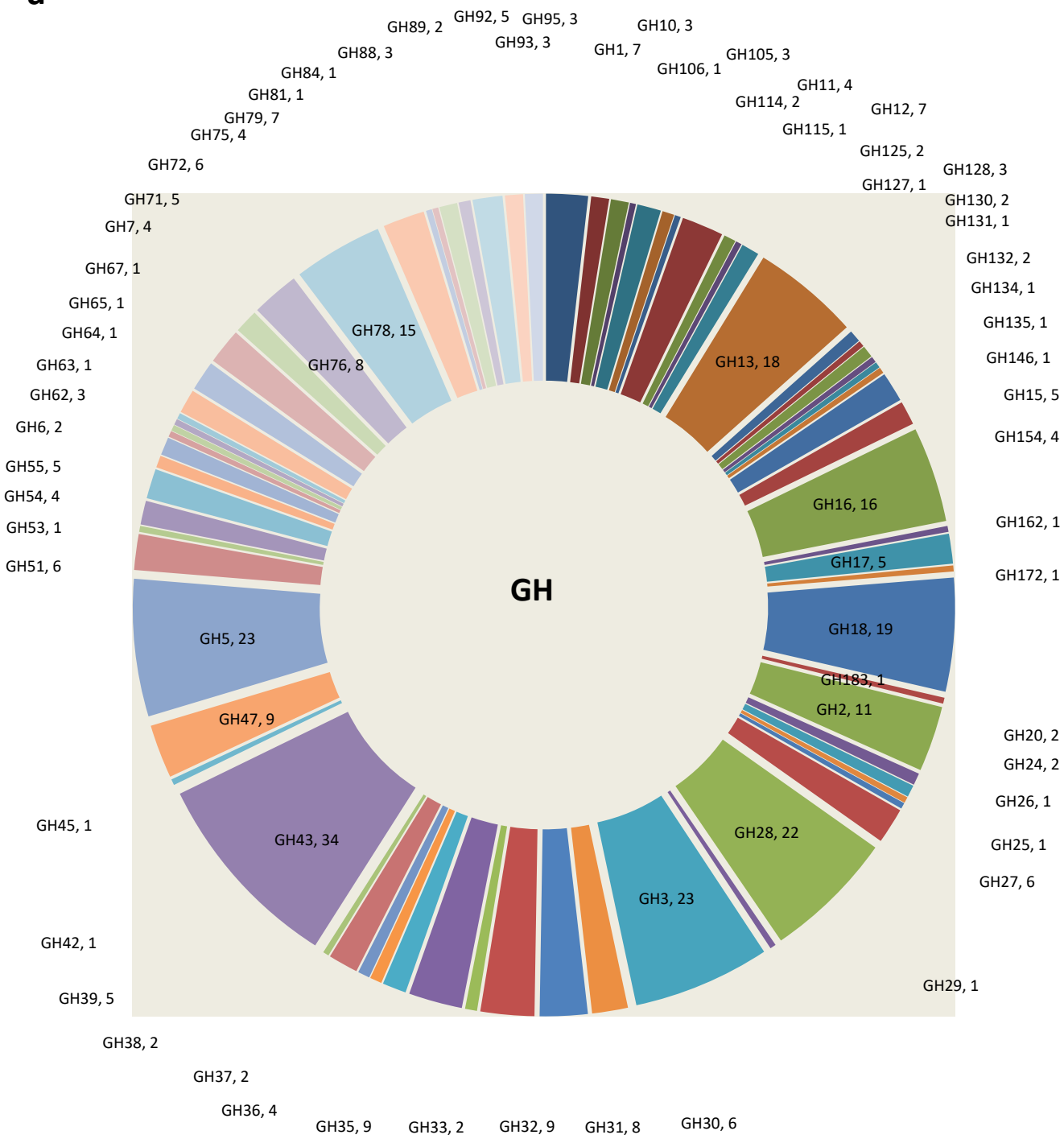
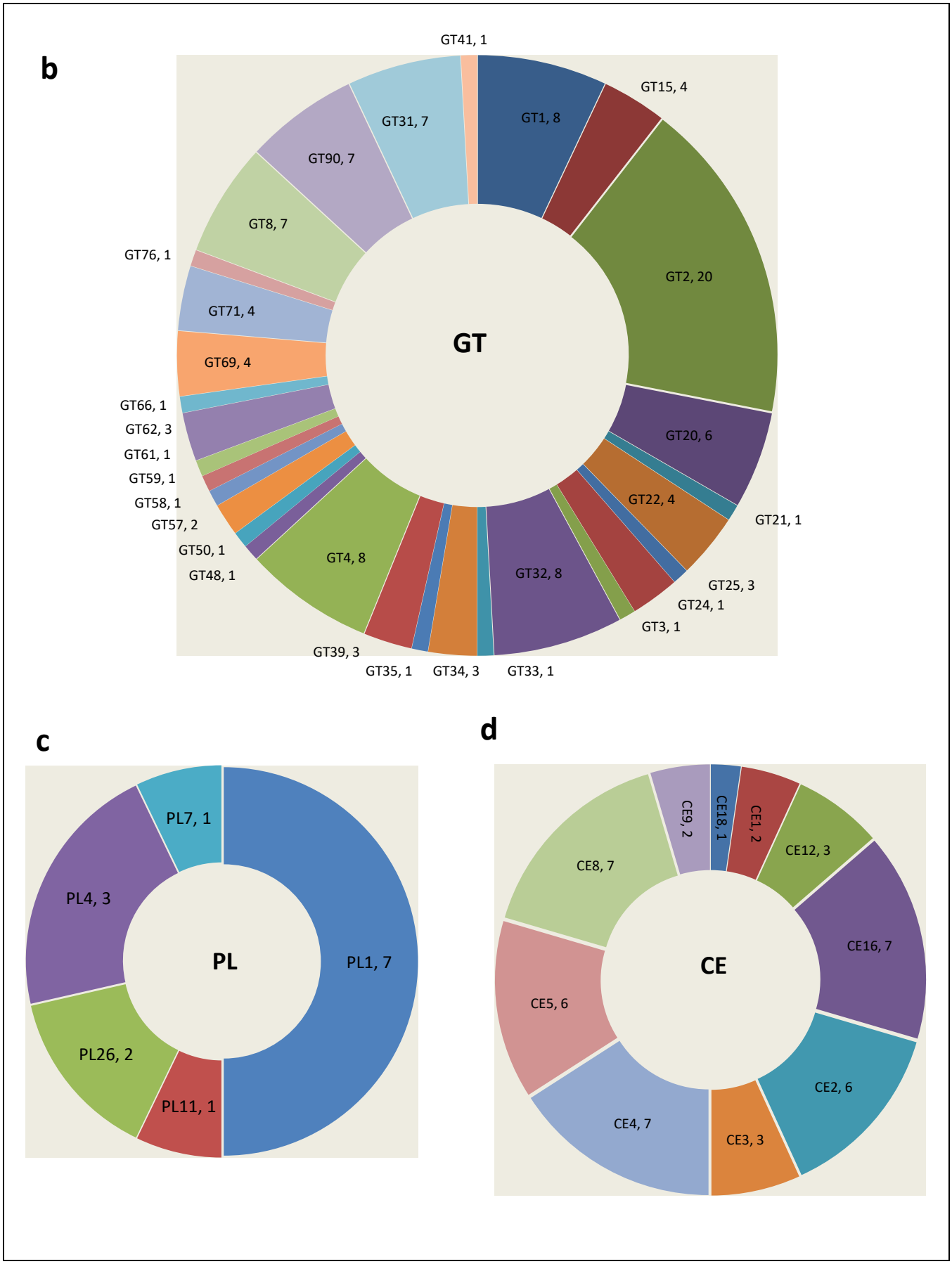


Figure 3.3 Distribution of CAZymes identified in the genome of *Pj* according to their respective CAZy families.

a


GH
GH89, 2
GH92, 5
GH95, 3
GH10, 3
GH93, 3
GH1, 7
GH88, 3
GH105, 3
GH106, 1
GH84, 1
GH11, 4
GH81, 1
GH114, 2
GH12, 7
GH79, 7
GH115, 1
GH75, 4
GH125, 2
GH72, 6
GH128, 3
GH127, 1
GH71, 5
GH130, 2
GH7, 4
GH131, 1
GH67, 1
GH132, 2
GH65, 1
GH134, 1
GH64, 1
GH78, 15
GH135, 1
GH63, 1
GH76, 8
GH13, 18
GH146, 1
GH62, 3
GH15, 5
GH6, 2
GH154, 4
GH55, 5
GH16, 16
GH54, 4
GH162, 1
GH53, 1
GH17, 5
GH51, 6
GH172, 1
GH5, 23
GH18, 19
GH183, 1
GH47, 9
GH2, 11
GH20, 2
GH24, 2
GH45, 1
GH26, 1
GH43, 34
GH28, 22
GH25, 1
GH3, 23
GH27, 6
GH42, 1
GH29, 1
GH39, 5
GH38, 2
GH37, 2
GH36, 4
GH35, 9
GH33, 2
GH32, 9
GH31, 8
GH30, 6


b
GT
GT41, 1
GT15, 4
GT1, 8
GT31, 7
GT90, 7
GT8, 7
GT76, 1
GT71, 4
GT69, 4
GT66, 1
GT62, 3
GT61, 1
GT59, 1
GT58, 1
GT57, 2
GT50, 1
GT48, 1
GT2, 20
GT20, 6
GT21, 1
GT22, 4
GT25, 3
GT24, 1
GT3, 1
GT32, 8
GT4, 8
GT39, 3
GT35, 1
GT34, 3
GT33, 1
c
PL
PL7, 1
PL4, 3
PL1, 7
PL26, 2
PL11, 1
d
CE
CE9, 2
CE18, 1
CE1, 2
CE12, 3
CE8, 7
CE16, 7
CE5, 6
CE2, 6
CE4, 7
CE3, 3

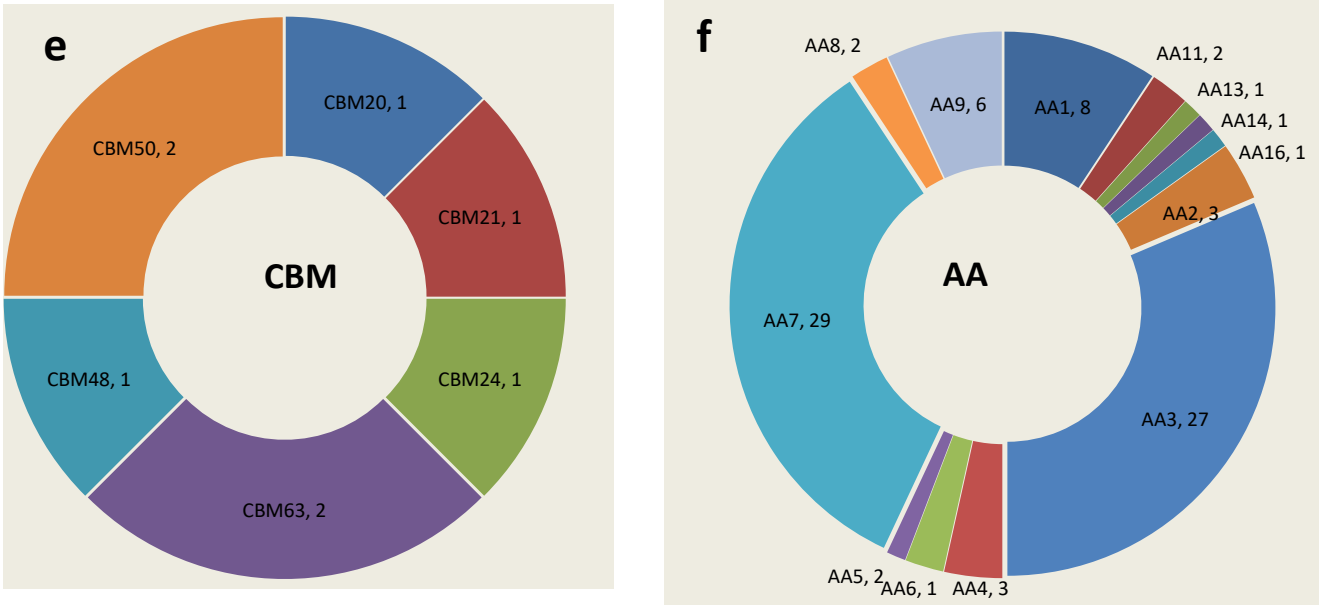


Figure 3.4 Distribution of CAZymes from *P.janthinellum* across various subfamilies within each enzyme family **a)** Glycoside Hydrolases, **b)** Glycosyltransferases **c)** Polysaccharide Lyases **d)** Carbohydrate Esterases **e)** Carbohydrate Binding Module **f)** Auxiliary Acivity

Glycoside hydrolases (GHs) (EC 3.2.1.-) play a central role in biomass degradation by catalyzing the hydrolysis of glycosidic bonds in complex plant polysaccharides such as cellulose, hemicellulose, and pectin. The Glycoside Hydrolases represent the major category of CAZymes in *P. janthinellum*, accounting for 60% of the total CAZymes, with 391 GH genes classified into 76 subfamilies. A detailed analysis of the GH subfamilies reveals that the most abundant families are GH43 (34 genes), followed by GH3 (23), GH5 (23), GH28 (22), GH18 (19), GH13 (18), GH16 (16), GH78 (15), and GH2 (11) (Figure 3.4a). These predominant enzyme families encode enzymes such as xylosidases and arabinofuranosidases (GH43), comprises beta-glucosidases (GH3) , endoglucanases and mannanases (GH5), polygalacturonases (GH28), chitinases (GH18), α-amylases (GH13), β-1,3-glucanases (GH16), α-L-rhamnosidases (GH78), and β-galactosidases (GH2), which collectively contribute to the extensive carbohydrate degradation capability of *P. janthinellum*. The second most abundant enzyme family is the Glycosyltransferases (GTs), comprising 114 genes, which accounted for 17% of the total CAZymes. They catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds (Lairson et al., 2008). The Glycosyltransferases were included in 30 subfamilies and majorly represented subfamilies within the GT group are GT2 (20 genes), followed by GT1 (8), GT4 (8), and GT32 (8) (Figure 3.4b). The third largest group of CAZyme families was the Auxiliary Activity (AA)

enzymes, comprising a total of 86 genes (13%), distributed across 13 subfamilies. The abundance of enzymes in Auxiliary Activity family indicates a robust oxidative enzyme system that compliments hydrolytic cellulases and hemicellulases in breaking down lignocellulosic biomass (Levasseur et al., 2013). Among these, the largest representation was from AA7, followed by AA3, AA9, and AA1. AA9 (formerly known as GH61) are Lytic polysaccharide monooxygenases (LPMOs), which oxidatively cleave cellulose, enhancing cellulose depolymerization and synergizing with classical cellulases (Horn et al., 2012). AA1 represents laccases, involved in lignin modification and detoxification (Janusz et al., 2020). Carbohydrate Esterases (CEs) accounted for 7% of the total CAZymes, with 44 genes classified into 10 subfamilies. The major subfamilies included CE8, CE4, CE16, CE2, and CE5, indicating a strong capacity to remove side-chain modifications such as acetyl and feruloyl groups from plant cell wall polysaccharides, thus facilitating further hydrolysis by other enzymes (Armendáriz-Ruiz et al., 2018) . Polysaccharide Lyases (PLs) and Carbohydrate-Binding Modules (CBMs) were the least represented families, constituting 2% and 1% of the total CAZymes, respectively.

3.3.3 Comparison with CAZymes from other fungi

The biomass degradation capacity of *P. janthinellum* is likely influenced by multiple factors. As discussed in the previous section, its genome encodes a large and diverse array of CAZymes, classified into various CAZy families. Comparing the distribution of these CAZymes within its genome to those of other known cellulase producers can help elucidate the possible reasons behind its enhanced enzymatic performance. To this end, we compared the number and distribution of CAZymes across different CAZy families in *P. janthinellum* with those in other cellulase-producing filamentous fungi from the genera *Penicillium* and *Aspergillus*, as well as with the model cellulase producer *Trichoderma reesei*. The results of this comparison are shown in Figure 3.5.

The total number of CAZymes in *P. janthinellum* surpasses that of *Trichoderma reesei* as well as two cellulolytic filamentous fungi from the genus *Aspergillus* (*A. niger* and *A. nidulans*) and three from the genus *Penicillium* (*P. chrysogenum*, *P. brevicompactum*, and *P. oxalicum*). These fungi are among the most studied and widely used industrial producers of biomass-hydrolyzing enzymes. Notably, *P. janthinellum* stands out with a substantial difference in the total number of CAZymes. Among the other fungi, *T. reesei* exhibited the lowest number of CAZymes (413), while *A. niger* showed the highest (557). However, *P. janthinellum* had a remarkable total of 657 CAZymes, over 100 more than *A. niger*. The largest differ-

ence comes from the Glycoside Hydrolase (GH) family, which includes most of the cellulolytic enzymes directly involved in cellulose hydrolysis. While the other fungi had a comparable number of GH genes, ranging from 207 in *T. reesei* to 267 in *A. nidulans*, *P. janthinellum* possessed 382 GH genes, a uniquely high count regardless of genus. Interestingly, other *Penicillium* species showed GH numbers similar to *Aspergillus* species and *T. reesei*, indicating that the elevated CAZyme and GH content is a species-specific feature of *P. janthinellum*.

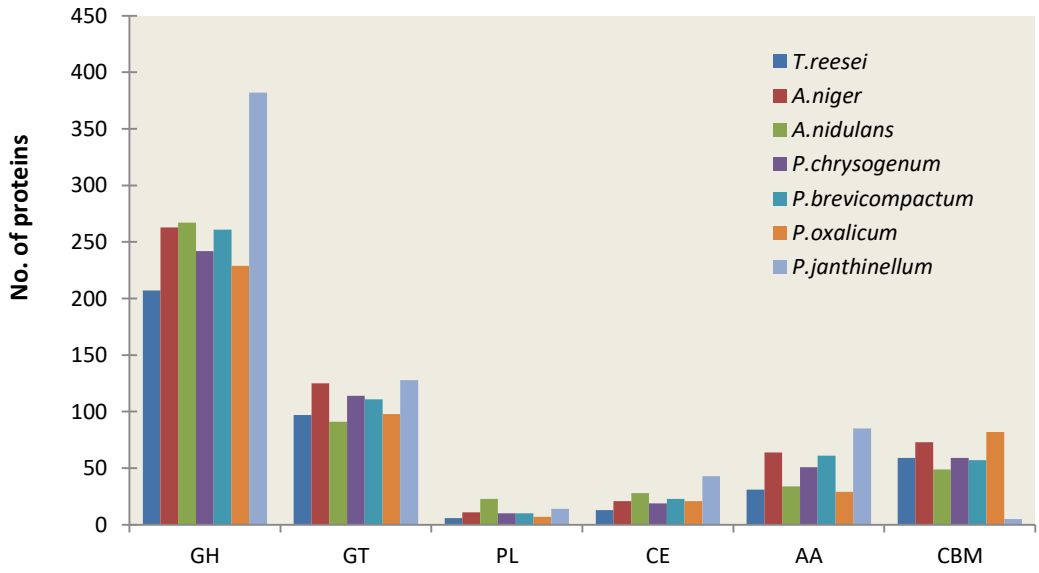


Figure 3.5 Comparison of the number of CAZymes in *P. janthinellum* with other cellulase-producing filamentous fungi.

The extensive array of cellulases present in the genome of *Penicillium janthinellum* might be one of the reasons for its enhanced cellulase production. However, it is important to note that the mere presence of numerous genes does not necessarily translate to higher overall enzyme production; rather, it provides greater versatility in substrate utilization and adaptability to diverse ecological niches (Mäkelä et al., 2014). In *T. reesei*, although the total number of CAZymes is lower compared to other cellulolytic fungi, these genes are often organized non-randomly into clusters. Such clustering and co-expression may contribute to its efficient cellulase production (Martinez et al., 2008). Unfortunately, similar analyses are not currently feasible for *P. janthinellum* due to the lack of a chromosome-level genome assembly. The genome assembly is at contig level, which limits the ability to investigate gene clustering and synteny in this species.

3.3.4 Identification of key cellulase genes

The cellulase genes that are critical for lignocellulosic biomass degradation were identified by analysis of the CAZyme annotation and the Uniprot data. The major cellulase genes identified from this analysis are summarized in Table 3.3. In the *P. janthinellum* genome, we identified 4 cellobiohydrolases (CBH), 9 endoglucanases (EG), and 14 β-glucosidases (BGL). The endoglucanases are distributed across GH5, GH7, GH12, and GH45 families, while only GH5 and GH12 EG are identified from *T.reesei* (Lynd et al., 2002), reflecting a broad enzymatic repertoire for internal cleavage of cellulose chains. Notably, one GH5 and one GH7 endoglucanase lack a signal peptide, suggesting possible intracellular roles.

Two cellobiohydrolases belong to GH6 family and the other two belong to GH7 family, all of which possess signal peptides, indicating their extracellular secretion. This arrangement reflects a robust system for the synergistic hydrolysis of cellulose where GH6 family CBHs generally act on the non-reducing ends of cellulose chains and GH7 family CBHs on the other hand, usually attacks the reducing ends of cellulose chains (Payne et al., 2015). Thus, compared to *T. reesei*, which relies heavily on a single GH6 and GH7 CBH pair, *P. janthinellum* harbors twice the number of GH6 and GH7 CBHs with secretion signals, suggesting possible functional redundancy or specialization that might contribute to its observed superior biomass conversion under certain conditions.

In *Penicillium janthinellum*, β-glucosidases (BGLs) are distributed across the GH1 and GH3 families, with 7 out of 14 predicted to possess signal peptides. This relatively large and diverse set of BGLs indicates a robust capacity to hydrolyze cellobiose and soluble cello-oligosaccharides into glucose, thereby alleviating product inhibition of upstream cellulases (Lynd et al., 2002). By contrast, *T. reesei* is well known to have a notoriously low endogenous β-glucosidase activity, with most of them from GH3 family. Only a few GH3 enzymes are produced, and these are secreted at relatively low levels (Peterson and Nevalainen, 2012). Consequently, in industrial applications, *T. reesei* cellulase preparations often require supplementation with external β-glucosidases to achieve efficient cellulose saccharification and avoid cellobiose accumulation (Berlin et al., 2007). This result clearly explains the lower accumulation of cellobiose observed in the hydrolysis medium for *P. janthinellum* compared to *T. reesei* RUT-C30, as documented in Chapter 1, highlighting the functional significance of the broader and more abundant β-glucosidase system in *P. janthinellum*.

Table 3.3 Key cellulase genes identified from the genome of *P. janthinellum*, CAZy family and presence of signal peptide

Gene	*Annotation*	*GH family*	*Signal peptide*
Endoglucanases			
g263_t1	*Endoglucanase 3*	*GH45*	*Yes*
g1099_t1	*Putative Endo-1,4-beta-D-glucanase*	*GH12*	*Yes*
g1614_t1	*Glucanase (EC 3.2.1.-)*	*GH7*	*No*
g2758_t1	*Putative Endoglucanase Cel5C*	*GH5*	*Yes*
g3362_t1	*Endoglucanase 1 (EC 3.2.1.4)*	*GH5*	*Yes*
g3817_t1	*Glucanase (EC 3.2.1.-)*	*GH7*	*Yes*
g4652_t1	*Putative Endoglucanase*	*GH5*	*No*
g4919_t1	*Endoglucanase 2 (EC 3.2.1.4)*	*GH5*	*Yes*
g9888_t1	*Putative Endo-1,4-beta-D-glucanase*	*GH5*	*Yes*
Cellobiohydrolases			
g243_t1	*Glucanase (EC 3.2.1.-) (Fragment)*	*GH7*	*Yes*
g1612_t1	*Glucanase (EC 3.2.1.-)*	*GH6*	*Yes*
g3309_t1	*Glucanase (EC 3.2.1.-)*	*GH6*	*Yes*
g7476_t1	*Glucanase (EC 3.2.1.-)*	*GH7*	*Yes*
Beta-Glucosidases			
g1642_t1	*Beta-glucosidase (EC 3.2.1.21)*	*GH3*	*Yes*
g1737_t1	*Putative Beta-glucosidase*	*GH3*	*Yes*
g1783_t1	*Periplasmic beta-glucosidase*	*GH3*	*No*
g4752_t1	*Beta-glucosidase (EC 3.2.1.21)*	*GH3*	*Yes*
g4875_t1	*Beta-glucosidase (EC 3.2.1.21)*	*GH3*	*No*
g5107_t1	*Beta-glucosidase (EC 3.2.1.21)*	*GH3*	*Yes*
g6682_t1	*Beta-glucosidase (EC 3.2.1.21)*	*GH3*	*No*
g8031_t1	*Putative beta-glucosidase*	*GH3*	*Yes*
g8604_t1	*Putative Beta-glucosidase 1B*	*GH1*	*No*
g9196_t1	*Beta-glucosidase (EC 3.2.1.21)*	*GH3*	*No*
g10145_t1	*Beta-glucosidase (EC 3.2.1.21)*	*GH3*	*No*
g11004_t1	*Probable beta-glucosidase G*	*GH3*	*No*
g2463_t1	*Putative Glycoside hydrolase family 1*	*GH1*	*Yes*
g7617_t1	*Uncharacterized protein*	*GH3*	*Yes*
LPMO			
g10717_t1	*Putative Cellulose monooxygenase Cel61A*	*AA9*	*Yes*
g9705_t1	*Putative Endo-1,4-beta-glucanase*	*AA9*	*No*
g4854_t1	*Putative Endoglucanase IV*	*AA9*	*Yes*
g8173_t1	*Putative Endoglucanase*	*AA9*	*Yes*
g185_t1	*Uncharacterized protein*	*AA9*	*Yes*
g6918_t1	*Uncharacterized protein*	*AA9*	*Yes*

Additionally, 6 lytic polysaccharide monooxygenase (LPMO) genes belonging to the AA9 family were identified. LPMOs are a relatively recent addition to the cellulase system (Vaaje-Kolstad et al., 2010). Unlike traditional hydrolytic cellulases, they employ an oxidative

mechanism to cleave cellulose, acting directly on the crystalline regions and thereby enhancing overall biomass depolymerization.

We have identified 69 hemicellulase genes in this study (Annexure III). These hemicellulases belong to both the Glycoside Hydrolase (GH) and Carbohydrate Esterase (CE) families. Among the GH families, GH43, GH35, and GH51 were predominantly represented, which included enzymes such as arabinofuranosidases, xylosidases, and galactosidases. Furthermore, two acetyl xylan esterases were identified from the CE family, underscoring the organism's potential to remove acetyl groups from xylan, thereby facilitating subsequent hydrolysis by xylanases (Biely, 2012). In summary, the identification of a large and diverse repertoire of critical enzymes involved in biomass hydrolysis underscores the organism's considerable potential for lignocellulosic biomass degradation.

3.3.5 Identification of transcription factors

Transcription factors represent a major group of genes that play a crucial role in the regulation of cellulase production in filamentous fungi. They are among the most extensively studied genetic elements involved in cellulase regulation. Numerous transcriptional regulators have been identified in various filamentous fungi, and their roles in modulating cellulase gene expression have been well established (Mattam et al., 2022). Most of these transcription factors have been identified and extensively studied in *T. reesei* (Bischof et al., 2016). Studying the regulators in *P. janthinellum* is important, as it may reveal the underlying reasons for its enhanced cellulase production. Moreover, these regulators could serve as valuable targets for genome engineering to develop strains with an optimized enzyme cocktail, suitable for various applications in the cellulase industry.

Initially, all the transcription factors annotated in the genome were extracted and analyzed using UniProt data. In total, 805 transcription factors were identified , among which 628 were belonging to GAL4-like Zn_2Cys_6 binuclear cluster DNA-binding domain (Figure 3.6). The GAL4-like Zn_2Cys_6 binuclear cluster DNA-binding domain is exclusively found in fungi and these transcription factors typically play a regulatory role in numerous secondary metabolic biosynthesis (Todd et al., 2014). Several well-characterized cellulase regulators, such as XYR1 in *Trichoderma reesei* belong to this family (Stricker et al., 2008).

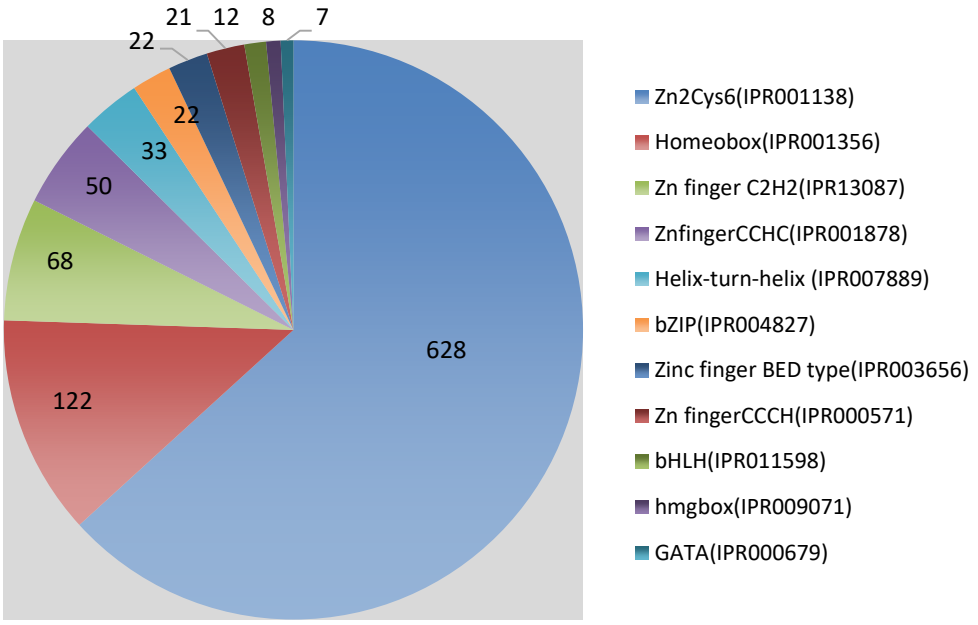


Figure 3.6 Domain comparison of transcription factors identified from the genome of *P. janthinellum*

There is a wealth of information on transcriptional regulators in filamentous fungi. From our analysis, we identified 12 transcription factors in *P. janthinellum* the homologs of which are well known to influence cellulase and hemicellulase expression in other filamentous fungi (Table 3.4). Among them, we have identified two copies of XlnR from the genome sequence (g9593_t1 and g8827_t1), which might contribute to its enhanced cellulase production. XlnrR was first identified in *Aspergillus* which affected the levels of cellulase and hemicellulases (Van Peij et al., 1998). Later it was identified and often considered as the central regulator of cellulase expression in *Trichoderma reesei* (Mach-Aigner et al., 2008). Among them, g9593 showed more similarity to the identified XlnR/Xyr1 gene sequences. The divergence in sequence between the two XlnR genes strongly suggests a gene duplication event, followed by functional divergence. This kind of evolution often leads to expanded regulatory capacity, likely contributing to the high lignocellulolytic efficiency of *P. janthinellum*.

In addition to XlnR/Xyr1, several other important transcription factors were identified from the genome that are known to regulate cellulase and hemicellulase expression. Notably, the CreA/Cre1 gene was detected, which generally acts as a negative regulator of cellulase expression through carbon catabolite repression (CCR), limiting enzyme production in the presence of easily metabolizable sugars like glucose (Ilmén et al., 1996). Other key transcription factors identified include ClrB, a Zn_2Cys_6 a factor shown to be essential for cellulase gene

induction in *Aspergillus* species (Coradetti et al., 2012) and ortholog to Clr-2 from *Neurospora crassa* (Coradetti et al., 2013). Additional regulators detected were PacG/Vib1, which is linked to cellulase induction under low pH and stress conditions (Xiong et al., 2014). We have also identified the Hap complex subunits, HapB and HapE (Hap2/3/5 complex in *T.reesei*) which participate in CCAAT-binding transcriptional complexes influencing many metabolic genes (Brakhage et al., 1999). Other transcription factors which integrate environmental signals (light, pH) with secondary metabolism and cellulase expression which include WC1 (a light-responsive regulator) and VeA, LaeA, and PacC (global regulators) (Tisch and Schmoll, 2010) were also identified from the genome. Furthermore, we identified Ace1 and Ace3, specific activators of cellulase genes in *Trichoderma* reesei (Aro et al., 2003).

Table 3.4 Orthologs of known transcription factors identified from the genome

Transcription Factor	Gene in the *Pj* genome
ClrB	g4518_t1
Xyr1/XlnR	g9593_t1, g8827_t1
CreA/Cre1	g1055_t1
AraR	g1951_t1
PacG/Vib1	g6695_t1
HapB	g8646_t1
Wc1(blr1)	g7400_t1
VeA	g3764_t1
Ace3	g8603_t1
PacC	g11812_t1
HapE	g10331_t1
Ace1	g1656_t1
LaeA	g2476_t1
clbr	g2815_t1

The presence of these regulators suggests that they may play a role in cellulase regulation in *P. janthinellum*. However, the mechanisms by which they function could differ in this fungus, as the roles of similar regulators are known to vary even between species of the same genus. Therefore, drawing definitive conclusions would require further functional studies on each of these transcription factors.

3.4 Conclusions

In this study, we successfully sequenced and analyzed the whole genome of *Penicillium janthinellum* NCIM 1366. The genome, at 37.6 Mbp, encodes an extensive array of carbohydrate-active enzymes (CAZymes), comprising approximately 5.4% of all predicted genes-notably higher than the CAZyme content reported in other industrially relevant fungi, including *Trichoderma reesei*. This remarkable enrichment, particularly in glycoside hydrolases, underscores the organism's evolutionary adaptation toward efficient plant biomass utilization. We identified a large repertoire of cellulolytic genes, including multiple endoglucanases, cellobiohydrolases, β-glucosidases, lytic polysaccharide monooxygenases (LPMOs), and hemicellulases collectively highlighting the genetic basis for the superior biomass hydrolysis observed in *P. janthinellum*. The genome also revealed a substantial number of transcription factors, with a striking dominance of GAL4-like Zn_2Cys_6 binuclear cluster DNA-binding domain proteins, alongside key orthologs of well-characterized cellulase regulators such as XlnR/Xyr1, CreA/Cre1, ClrB, Ace1, Ace3, and others. These regulators likely form a complex network orchestrating cellulase and hemicellulase gene expression, offering promising targets for future genetic engineering to enhance enzyme production. Overall, this comprehensive genomic analysis provides valuable insights into the molecular machinery underpinning lignocellulose degradation in *P. janthinellum.* These findings position *P. janthinellum* as a highly promising candidate for industrial applications in biomass conversion and biofuel production.

Chapter 4

Transcriptome analysis of the cellulose-induced expression of CAZymes in *Penicillium janthinellum* NCIM 1366

4.1 Introduction

Transcriptome sequencing utilizes high throughput platforms to capture and sequence the entire mRNA pool of specific tissues or cells at a given time point. It is also referred to as RNA-Seq or NGS-based mRNA sequencing. It has emerged as an indispensable tool in molecular biology, offering insights into how environmental and physiological parameters influence gene expression. In contrast to traditional methods that focus on individual genes, transcriptome data enabled the study of global expression patterns and thereby facilitated the identification of regulatory networks and the effect of genetic mutations (Wang et al., 2009).

Carbohydrate-active enzymes (CAZymes) play a pivotal role in the degradation of biomass and require the synergistic action of diverse extracellular enzymes. However, the biosynthesis and secretion of these enzymes are metabolically expensive and are tightly regulated in filamentous fungi in response to many environmental factors. The type and availability of carbon sources rank among the most crucial factors influencing the process (Kubicek et al., 1993). When readily utilizable sugars such as glucose are present, fungi suppress the production of cellulases and other carbohydrate-active enzymes (CAZymes) via a regulatory mechanism called carbon catabolite repression(Campos Antoniêto et al., 2015). The specific repertoire of CAZymes secreted is therefore influenced by the carbon source and may change over time. Many transcriptome-based studies have been useful in understanding the changes in their expression in response to the available carbon source in the growth medium (Bischof et al., 2013; Ries et al., 2013; Chen et al., 2014; Borin et al., 2017; Li et al., 2023).

We have identified the CAZyme genes in *P. janthinellum* through whole genome sequencing and analysis. However, transcriptome profiling-based studies on the pattern of expression of CAZymes, particularly cellulases, will help to identify the genes that contribute to its biomass degradation potential at specific growth conditions. The primary objective of the transcriptome analysis is to study the CAZyme expression pattern and to identify the co-expressed genes that may regulate cellulase expression in *P. janthinellum*. An induction study was conducted to determine the most effective carbon source for rapidly inducing cellulase expression. Based on these findings, the optimal time point at which key cellulase genes begin to show significant expression was identified and subsequently RNA-Seq analysis was performed at that stage. In this chapter, a comparison between the cellulose-induced transcriptome and the non-induced (glucose) transcriptome is presented, focusing on changes in the expression patterns of CAZymes. This study can serve as a foundation for understanding the

regulatory mechanisms governing cellulase expression in *P. janthinellum* and may facilitate the identification of candidate genes for further functional studies.

4.2 Materials and methods

4.2.1 Optimization of growth conditions for inducer addition

Mandels and Weber medium (Mandels and Weber, 1969) (100 ml) supplemented with 1% w/v (10 mg/ml) glucose as the sole carbon source was inoculated with 1×10^7 spores from *P. janthinellum* and incubated at 30 °C with agitation at 200 rpm. Samples were collected at different time points and the concentration of remaining glucose in the medium was determined using HPLC analysis. For growth calculation at same conditions, biomass was collected by filtration through pre-weighed Whatman No.1 filter paper. The filter paper along with the fungal mycelium was dried at 50°C for 48 hours in an oven and re-weighed to obtain mycelial dry-weight.

4.2.2 Selection of inducer for gene expression study

P. janthinellum spores (1×10^7 spores) were inoculated in 100 ml of Mandels and Weber medium supplemented with 1% w/v glucose as the sole carbon source and incubated at 30 °C with 200 rpm agitation. After 72 hours of growth (when glucose is depleted), the biomass is collected and washed with distilled water to remove residual glucose. Then the mycelia were transferred to fresh Mandels and Weber medium supplemented with 1% (w/v) of the various carbon sources: cellulose, cellobiose, cellobiose octa-acetate, Avicel®, carboxymethyl cellulose (CMC), or glucose (control). Culture supernatant samples were collected at 2h, 6h, 9h, 12h and 24h for a qualitative Congo Red assay to assess cellulase production.

For the assay, sterile discs were placed on CMC-Agar plates and each disc was loaded with 10 µl of the culture supernatant and incubated at 50 °C for 1 hour. Then the plates were flooded with 0.1% Congo red solution and stained for 15 minutes. Subsequently, 1M NaCl was added for de-staining and then plates were observed for clearance zone around the discs.

4.2.3 Expression analysis of cellulase genes and regulators under induced and uninduced conditions

P. janthinellum spores (1×10^7 spores) were inoculated in 100 ml of Mandels and Weber medium supplemented with 1% w/v glucose as the sole carbon source and incubated at 30 °C with 200 rpm agitation. After 72 hours of growth, the biomass was collected and washed with distilled water to remove residual glucose. Then the mycelia were transferred to fresh Mandels and Weber medium supplemented with 1% (w/v) of either cellulose or glucose (control) as carbon source/inducer. Mycelial samples (100 mg) were collected at 2h and 4h and 6h post-induction and washed with distilled water. Mycelia were flash frozen in liquid nitrogen and ground to a fine powder using Mortar and Pestle.

RNA isolation from the powdered mycelia was done using PureLink™ RNA Mini kit (Thermo scientific). Isolated RNA (1μg) was treated with DNaseI (Thermo Scientific) to remove the contaminating DNA, and was then used for cDNA synthesis. cDNA synthesis was carried out using RevertAid™ First Strand cDNA Synthesis Kit (Thermo Scientific), with oligo-dT primers as per manufacturer's protocol. GAPD (internal control) and the selected cellulase genes were PCR amplified from the cDNA following standard PCR protocols.

4.2.4 Transcriptome sequencing and analysis of differentially expressed genes

Isolation of RNA was done as described above. The sequencing, *de-novo* assembly and analysis of the transcriptome were performed at OmicsGen LifeSciences Pvt Ltd, Kochi. The quality checking of the RNA was done by Agilent bioanalyzer 2100. rRNA was removed from the sample using Ribo ZeroTM rRNA removal kit. Library construction was done using NEBNext Ultradirectional RNA Library Prep kit. Library purification, detection and quantification were done using AxyPrep Mag PCR clean-up kit, Agilent Bioanalyzer 2100, Quibit respectively. It was followed by CBOT auto-clustering and sequencing by Illumina HiSeq system.

The sequence data was analyzed using bioinformatics tools. Bcl2fastq (v2.17.1.14) was used to processed the original image data for base calling and preliminary quality analysis. Quality assessment of the sequencing data was evaluated using FastQC (v0.10.1). Raw sequencing data (Pass Filter Data) was processed to remove adapters and reads of low quality using Cutadapt (version 1.9.1) (Martin, 2011). The resulting clean data was used for subsequent bioin-

formatics analysis. The genome *de-novo* assembly was performed using Trinity (Grabherr et al., 2011). Repeated contigs were removed by CD-HIT (Fu et al., 2012) to obtain the unigene file. ORF detection was performed using TransDecoder. Unigenes were annotated using Nr database, SWISS-PROT, gene ontology (GO) database, The Kyoto Encyclopedia of Genes and Genomes (KEGG) Database and COG (Cluster of Orthologous Groups of proteins) database. The level of gene expression is measured by read density, the higher read density corresponding to a higher the level of the gene expression. Gene expression was calculated using RSEM (V 0.6.1) (Li and Dewey, 2011), which measures gene expression in FPKM (Fragment per Kilo bases per Million reads). EdgeR was used for differential expression analysis (Robinson et al., 2010). Fold Change is calculated as ratio of FPKM values between induced and uninduced samples. Differential expression is reported as logarithnmic value of Fold Change value (log_2FC) and genes with $|log_2FC| \geq 2$ (i.e., $\geq$ 4-fold change) are considered significantly differentially expressed.

4.2.5 Identification of CAZymes and key cellulase genes

Annotation of CAZy (Carbohydrate Active Enzymes) family of genes was performed as per the CAZy database (Lombard et al., 2014) using the dbCAN2 meta server (http://bcb.unl.edu/dbCAN2) (Yin et al., 2012), Cellulase genes and other regulatory genes were identified from the Annotation data.

4.3 Results and discussions

4.3.1 Optimization of growth conditions for inducer addition

CAZyme gene expression in filamentous fungi is modulated by different inducing carbon sources. However, before inducer addition, it is critical to establish sufficient fungal biomass, which will facilitate the identification of early gene expression responses and minimize the influence of ongoing growth. Previous studies have shown that fungal biomass accumulation is slower when grown on cellulase-inducing carbon sources, such as cellulose or lactose (Messner and Kubicek, 1991). To overcome this limitation, the fungus was initially cultured in Mandels and Weber medium containing 1% glucose as the sole carbon source. The point of glucose depletion was determined by HPLC analysis, and the data is shown in Table 4.1. By 72 hours, only 1.5% of the initial glucose was remaining in the medium, and the fungus also exhibited maximum growth, as determined by mycelial dry weight measurements (Fig-

ure 4.1). Based on these results, inducer addition was carried out at 72 hours of growth in glucose-containing medium.

Table 4.1 Glucose consumption profile of *P. janthinellum* during growth

Time (h)	Glucose conc. (g/L)	Percentage of Glucose remaining
0	10.00	100.0
48	3.20	32.0
60	2.03	20.3
72	0.15	1.5
84	Not detected	-

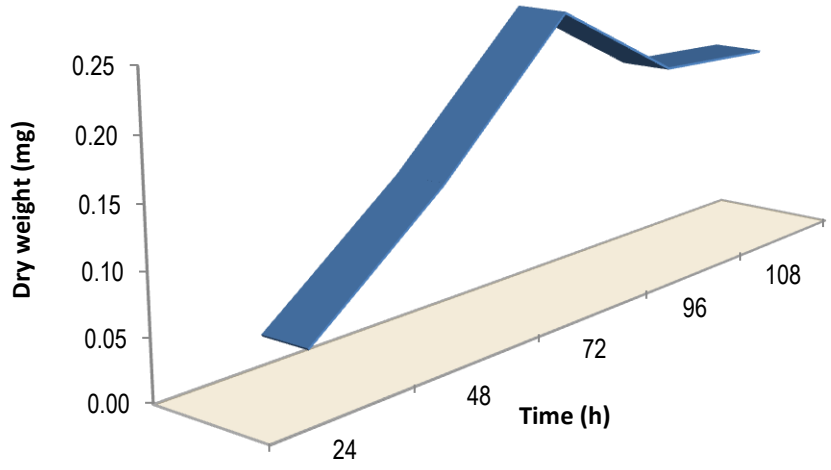


Figure 4.1 Time-course analysis of changes in mycelial dry weight of *P. janthinellum* during growth in Mandels and Weber medium with glucose as the sole carbon source

4.3.2 Selection of inducer for transcriptome studies

Earlier reports suggest that the presence of cellulose, xylan, or mixtures of plant polymers in the culture medium induces the production of cellulolytic and xylanolytic activities by *T. reesei* (Mach and Zeilinger, 2003). Apart from these, pure mono or oligosaccharides and their transglycosylation products, such as sophorose, β-cellobiono-1, 5-lactone, D-xylose, xylobiose, galactose, and lactose, have been identified as inducers that stimulate the production of cellulase and hemicellulase enzymes (Amore et al., 2013). Six carbon sources were selected for the study, which included lactose, cellulose, cellobiose, cellobiose octa-acetate, Avicel, and glucose (negative control). The selection of these carbon sources was based on earlier

studies that showed their role in inducing cellulolytic enzymes (Ryu and Mandels, 1980; Kamagata et al., 1991; Znameroski et al., 2012; Ivanova et al., 2013). Carbon sources, which are whole biomass components such as wheat bran or rice straw, were avoided in this study to prevent RNA contamination, which could interfere with transcriptome data quality.

Fungal mycelia were harvested and thoroughly washed with distilled water to remove residual glucose after 72 hours of growth. Mycelia were then transferred to fresh Mandels and Weber medium supplemented with 1% (w/v) of the test carbon sources, which served as inducers for cellulase production. Culture filtrates were collected at different time points after induction, to assess the cellulase activity using a qualitative Congo Red assay (Figure 4.2).

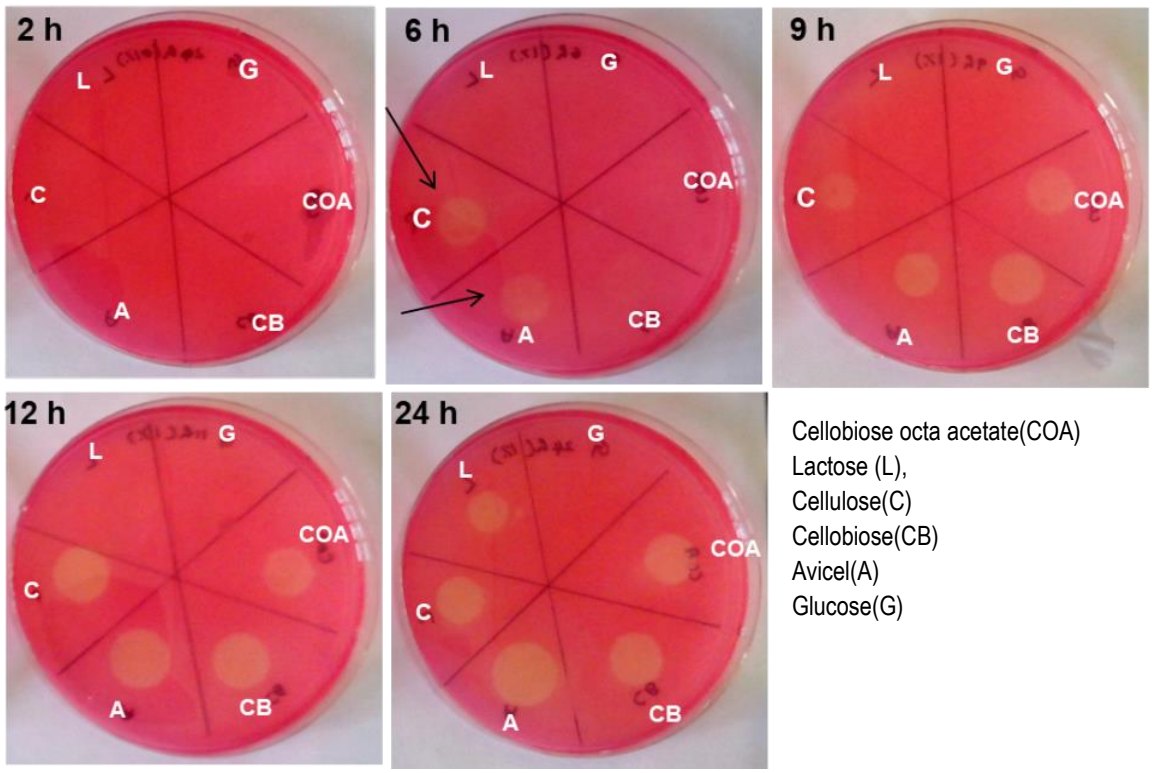


Figure 4.2 Time-course analysis of cellulase activity in *P. janthinellum* using Congo red assay during growth on different carbon sources as inducers

All the tested carbon sources acted as inducers of cellulase production, indicated by the clearing zone. However, cellulose and Avicel induced cellulase production as early as 2 hours after induction, suggesting their efficiency as fast-acting inducers. While both are cellulose-based carbon sources, Avicel is characterized by a high degree of crystallinity (Hall et al., 2010). Since the lower solubility and heterogenous nature of Avicel may complicate downstream sample processing and RNA extraction, cellulose was selected as the inducer for transcriptome studies. However, it is important to note that the most effective inducer for cellulase production can only be definitively determined through quantitative enzyme assays.

4.3.3 Selection of time point for transcriptome analysis

The objective of this experiment was to determine the earliest time point at which major cellulase gene expression begins. This is expected to help in identifying the co-expressed regulators through transcriptome analysis. Two of the key cellulase genes -g4919_t1 (endoglucanase) and g7476_t1 (cellobiohydrolase) were selected as markers. The GAPD gene (g2773_t1) was used as the internal control. The culture was induced with cellulose and mycelial samples were collected at 0h, 2h and 4h. Samples at the same time points from a culture with glucose as carbon source was used as control (non-inducing). Total RNA was isolated from the collected mycelial samples and cDNA was synthesized with oligodT primer. The cDNA at different time points were used for templates for amplification of the selected cellulase genes using real time primers. The results are shown in Figure 4.3. GAPD expression remained the same across all the conditions confirming normalization. Transcript accumulation was not observed for glucose induced cultures for both the marker genes at all the time points. The transcript accumulation of cellobiohydrolase gene started as early as 2h of induction indicated by a product at 189 bp (Figure 4.3c). However, the expression of the endoglucanase gene started only at 4h of induction indicated by 144 bp product (Figure 4.3b). The 4h time-point was selected for the transcriptome analysis, since it was the earliest at which both the cellulase markers were expressed.

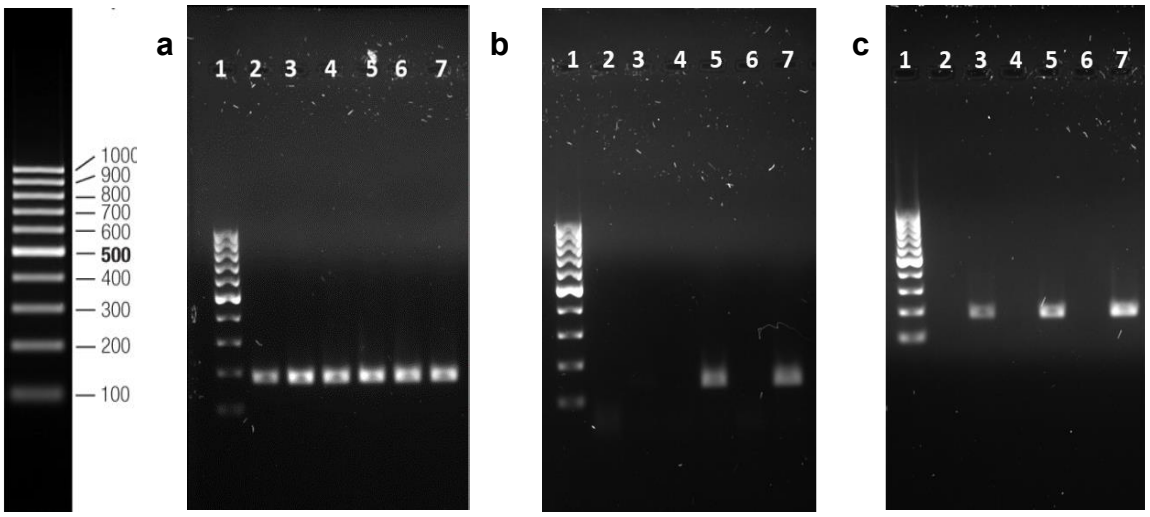


Figure 4.3 Agarose gel images showing the RT-PCR products of genes using real-time primers on cellulose-induced and non-induced (glucose) cultures of *P. janthinellum* at different time points a) *gapd* (g2773_t1) (control) b) *eg* (g4919_t1) c) *cbh* (g243_t1). Lane information for all three gels:1-DNA ladder, 2-2nd hour glucose, 3-2nd hour cellulose ,4-4th hour glucose, 5-4th hour cellulose, 6-6th hour glucose, 7-6th hour cellulose. DNA ladder image is shown on the left and is same for all the gels.

4.3.4 Overview of the transcriptome data

The transcriptome sequencing was performed using NGS for the 4h post cellulose induction samples in comparison with the glucose-induced culture at the same time point as a control. The sequences were *de-novo* assembled into a total of 22023 transcripts (unigenes). Among them, 10523 transcripts were down-regulated (log2 (FC) <0) and 11480 transcripts were up-regulated (log2 (FC) >0). Among these, 2062 transcripts were solely expressed in glucose, 3221 transcripts were solely expressed in cellulose, and 16739 transcripts were expressed in both glucose and cellulose (Figure 4.4). Among them, 1167 genes with significant differential expression were selected based on the fold change data filtered by the criteria $|\log2\ (FC)| \geq 2$, which equates to a minimum fourfold increase or decrease in the gene expression. This stringent criterion is expected to select the genes having biologically relevant fold change with high confidence and low background noise (Love et al., 2014). Among the differentially expressed genes, 815 were upregulated and 352 were downregulated, as shown in Figure 4.5. All the further analysis on the transcriptome was performed on the genes with significant differential expression.

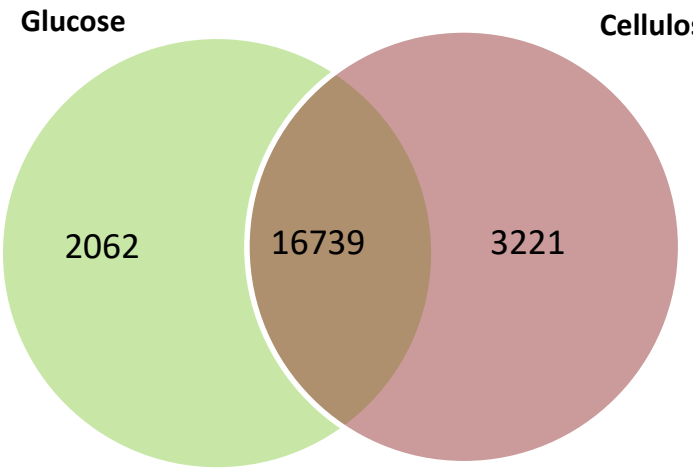


Figure 4.4 Venn diagram showing the distribution of unigenes at different conditions in the transcriptome analysis

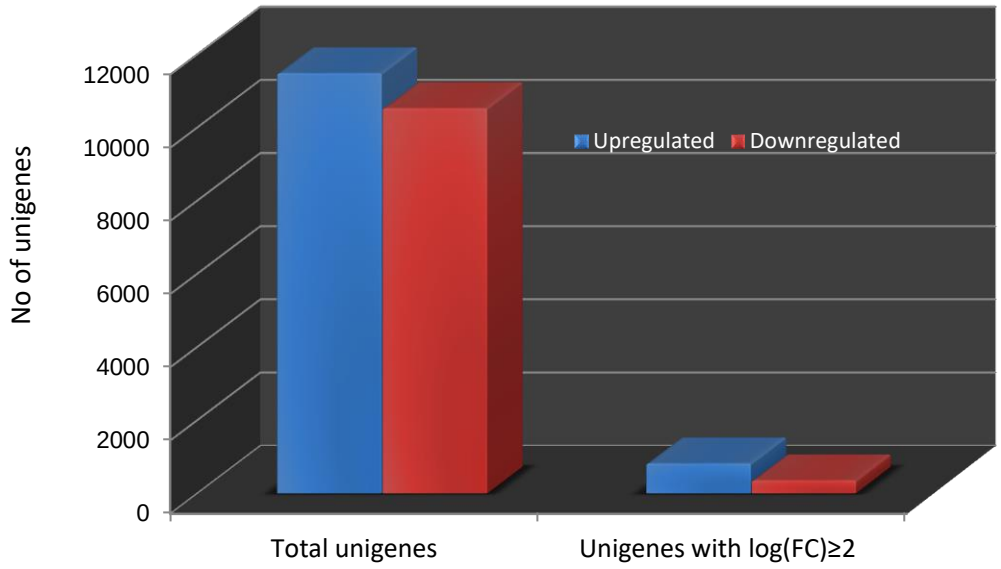


Figure 4.5 Expression data of Unigenes identified from the transcriptome of *P. janthinellum*

The functional annotation of the differentially expressed genes was performed using the Clusters of Orthologous Groups (COG) database and the distribution is shown in Figure 4.6. Among the annotated functional categories, carbohydrate metabolism and transport was the predominant category, which is expected for a cellulose induced transcriptome. It was followed by lipid metabolism and amino acid metabolism, which may be representative of broader metabolic changes in association with the cellulose induction.

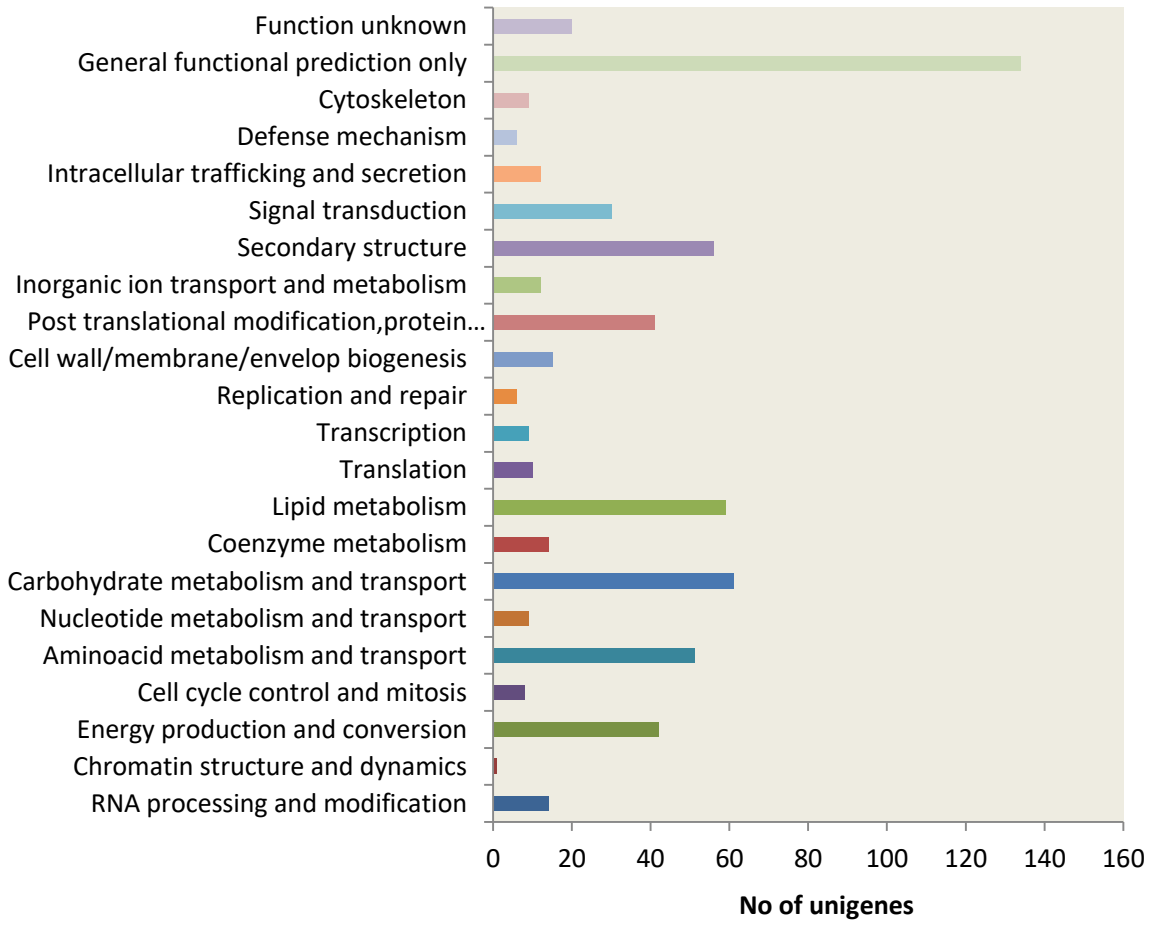


Figure 4.6 COG functional annotation categories identified from the transcriptome analysis of *P. janthinellum*

4.3.5 CAZymes identified from the transcriptome data

We have identified CAZymes from the differentially expressed genes using dbCAN (Yin et al., 2012). A total of 226 genes coding for CAZymes were identified, which accounted for 19.4% of all the differentially expressed genes. The list of CAZymes is shown in Annexure IV. Their classification according to the CAZy database is shown in Figure 4.7. Among the CAZymes, 137 genes (64%) belonged to the Glycoside Hydrolase (GH) family, representing the most abundant category. This was followed by 36 genes (16%) in Glycosyltransferase (GT), 28 genes (12%) in Auxiliary Activity (AA), 18 genes (8%) in Carboxyl Esterase (CE), 6 genes (3%) in Carbohydrate-Binding Modules (CBM) and a single gene in Polysaccharide Lyases (PL) families. Fold change in expression of the identified CAZymes for these CA-

Zymes is represented in Figure 4.8. Among them, majority of the CAZyme genes were up-regulated, only a small subset showed downregulation.

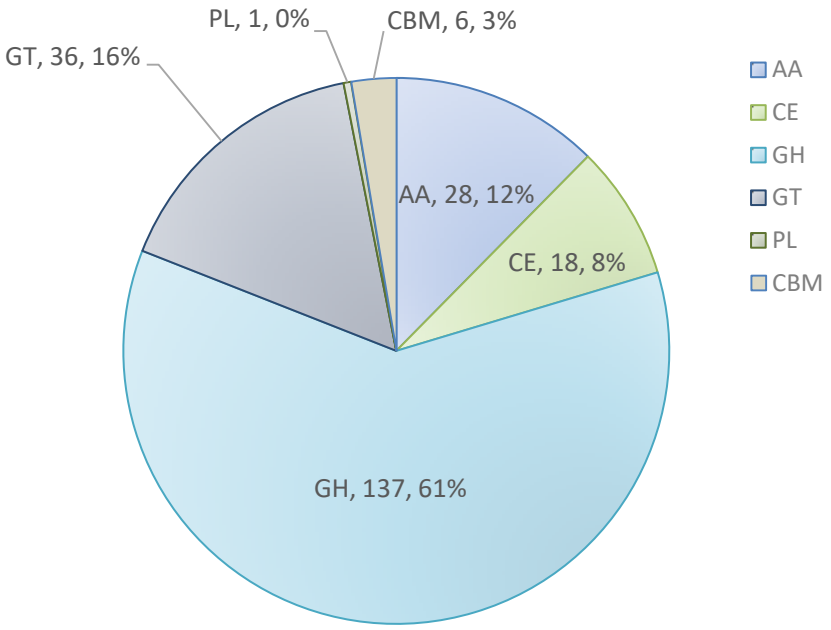


Figure 4.7 Distribution of CAZymes that showed significant differential expression in the transcriptome of *P. janthinellum* across different CAZy families

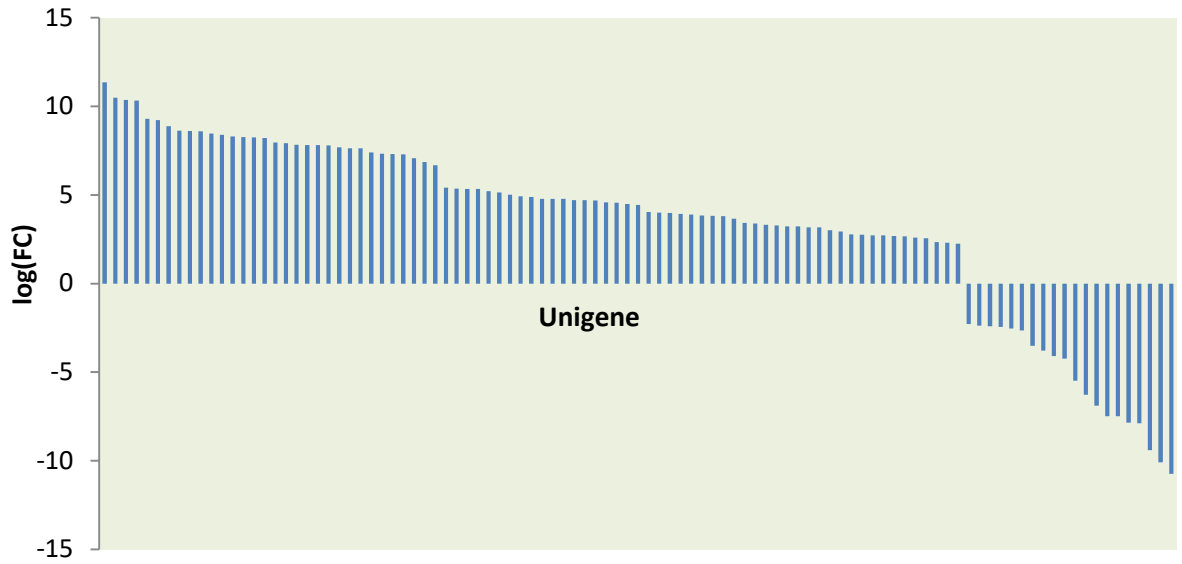


Figure 4.8 Fold change in expression of CAZymes that showed significant differential expression in the transcriptome of *P. janthinellum*

A detailed comparison between the genome and transcriptome data identified cellulase and hemicellulase genes that showed differential expression, and the data is shown in Table 4.2. It included 14 genes representing cellulases (4 CBH, 6 EG, 4 BGL), 16 hemicellulases and 2 LPMOs. These enzymes represent the core components of fungal cellulases and play a vital role in biomass degradation by fungi (Gao et al., 2011). From the 9 endoglucanases (EG) annotated in the genome data, 6 were detected in the transcriptome data and all of them showed a very high upregulation with log2 (FC) values ranging from 5.33 to 8.61. The highest log2 (FC) value of 8.61 was observed for g2758_t1 (*Cel5C*) which represents a fold change of 390 in comparison with glucose culture. It was followed by the marker gene g4919 (Endoglucanase 2) that was selected for optimizing the time point for transcriptome analysis; with a log2 (FC) value of 8.46. All the four cellobiohydrolases identified from the genome were detected in the transcriptome with high upregulation, with log2 (FC) ranging from 7.82 to 8.88. This corresponds to a fold change in expression of around 225-470 fold. Among the 14 beta-glucosidases annotated in the genome, only four were detected from the transcriptome, all showing significant upregulation with log2 (FC) value ranging from 3.3 to 7.9. In contrast, all the cellobiohydrolases (CBH) and majority of the endoglucanases (EG) were detected in the transcriptome with significant upregulation. The absence of other beta glucosidases might be because the onset of their expression starts at a later time point after sufficient accumulation of cellobiose in the medium. This aligns with a study in *T. reesei* which showed the secretion of CBH and EG started earlier than that of BGL (Li et al., 2019). Among the upregulated BGL genes, three of them had a signal peptide, possibly indicating their extracellular location, and one was without signal peptide and may be assumed to function intracellularly. It is speculated that these beta glucosidases might be contributing to the production of soluble inducer from cellulose in *P. janthinellum* consistent with its role in inducing cellulases in response to insoluble cellulose (Zhou et al., 2012). Two out of the 6 LPMO genes could also be identified and these showed a strong upregulation (log2 FC =8.59 and 7.68) indicating active oxidative enhancement of cellulose hydrolysis. Among the 69 hemicellulases in the genome, only 15 were detected and they showed log2 FC values ranging from 3.65- 11.35. These included GH10 Xylanase A (log_2FC 10.32), GH2 β-mannosidase (log_2FC 11.35), GH43 β-xylosidases and exo-β-galactanases, GH35, GH36, GH54, GH67, GH30, CE1 and CE5 carbohydrate esterases, and GH27 α-galactosidase.

Table 4.2 Cellulases and hemicellulases differentially expressed on cellulose induction

Unigene	*Gene*	*log2 (Fold Change)*	*GH family*	*Signal peptide*	*Annotation*
EG					
TRINITY_DN16044_c0_g1_i1	*g263_t1*	*6.85*	*GH45*	*Yes*	*Endoglucanase 3*
TRINITY_DN10691_c0_g1_i1	*g1099_t1*	*6.67*	*GH12*	*Yes*	*Putative Endo-1,4-beta-D-glucanase*
TRINITY_DN15972_c1_g1_i1	*g2758_t1*	*8.61*	*GH5*	*Yes*	*Putative Endoglucanase Cel5C*
TRINITY_DN13469_c0_g1_i1	*g3817_t1*	*5.33*	*GH7*	*Yes*	*Glucanase (EC 3.2.1.-)*
TRINITY_DN10938_c0_g1_i1	*g4919_t1*	*8.46*	*GH5*	*Yes*	*Endoglucanase 2 (EC 3.2.1.4)*
TRINITY_DN2819_c0_g1_i1	*g9888_t1*	*7.80*	*GH5*	*Yes*	*Putative Endo-1,4-beta-D-glucanase*
CBH					
TRINITY_DN7078_c0_g2_i1	*g243_t1*	*8.31*	*GH7*	Yes	*Glucanase (EC 3.2.1.-) (Fragment)*
TRINITY_DN10072_c0_g1_i1	*g1612_t1*	*7.82*	*GH6*	Yes	*Glucanase (EC 3.2.1.-)*
TRINITY_DN12902_c0_g1_i1	*g3309_t1*	*8.62*	*GH6*	Yes	*Glucanase (EC 3.2.1.-)*
TRINITY_DN7078_c0_g1_i1	*g7476_t1*	*8.88*	*GH7*	Yes	*Glucanase (EC 3.2.1.-)*
BGL					
TRINITY_DN8487_c0_g2_i1	*g4752_t1*	*4.58*	*GH3*	*Yes*	*Beta-glucosidase (EC 3.2.1.21)*
TRINITY_DN19387_c0_g1_i1	*g10145_t1*	*4.78*	*GH3*	*No*	*Beta-glucosidase (EC 3.2.1.21)*
TRINITY_DN24416_c0_g1_i1	*g2463_t1*	*7.90*	*GH1*	*Yes*	*Putative Glycoside hydrolase family 1*
TRINITY_DN5845_c0_g1_i1	*g7617_t1*	*3.30*	*GH3*	*Yes*	*Uncharacterized protein*
LPMO					
TRINITY_DN18851_c0_g1_i1	*g10717_t1*	*8.59*	*AA9*	*Yes*	*Putative Cellulose monooxygenase Cel61A*
TRINITY_DN3397_c0_g1_i2	*g4854_t1*	*7.68*	*AA9*	*Yes*	*Putative Endoglucanase IV*
Hemicellulases					
TRINITY_DN2919_c0_g3_i1	*g549_t1*	*7.63*	*GH26*	*Yes*	*Putative beta-1,4-mannanase*
TRINITY_DN21704_c0_g1_i1	*g748_t1*	*7.81*	*GH10*	*Yes*	*Beta-xylanase (EC 3.2.1.8)*
TRINITY_DN5374_c0_g1_i1	*g2256_t1*	*11.35*	*GH2*	*No*	*Putative beta-mannosidase*
TRINITY_DN3800_c0_g1_i1	*g2952_t1*	*5.48*	*GH43*	*Yes*	*Putative exo-beta-1,3-galactanase*
TRINITY_DN9077_c0_g2_i8	*g4267_t1*	*7.10*	*GH43*	*No*	*Putative Beta-xylosidase*
TRINITY_DN22093_c0_g1_i1	*g4996_t1*	*4.71*	*GH43*	*No*	*Beta-D-xylosidase (EC 3.2.1.37)*
TRINITY_DN6904_c0_g2_i1	*g6737_t1*	*5.21*	*GH54*	*No*	*Probable alpha-N-arabinofuranosidase B*
TRINITY_DN2456_c0_g1_i1	*g8569_t1*	*10.32*	*GH10*	*Yes*	*Endo-1,4-beta-xylanase A (Xylanase A)*
TRINITY_DN981_c0_g1_i1	*g8709_t1*	*4.88*	*GH36*	*No*	*Probable alpha-galactosidase G*
TRINITY_DN25274_c0_g1_i1	*g9001_t1*	*5.34*	*GH35*	*Yes*	*Beta-galactosidase (EC 3.2.1.23)*
TRINITY_DN10150_c0_g1_i1	*g9309_t1*	*4.57*	*GH67*	*No*	*Alpha-glucuronidase*
TRINITY_DN19635_c0_g1_i1	*g9310_t1*	*8.26*	*GH30*	*Yes*	*Putative endo-beta-1,4-xylanase*
TRINITY_DN16894_c0_g1_i1	*g3484_t1*	*4.68*	*CE5*	*No*	*Acetylxylan esterase 2 (EC*

					3.1.1.72) (AXE II)
TRINITY_DN17336_c0_g1_i1	*g7103_t1*	*3.98*	*CE1*	*Yes*	*Acetylxylan esterase A (EC 3.1.1.72) (AXE I)*
TRINITY_DN13454_c0_g1_i1	*g966_t1*	*3.65*	*GH27*	*Yes*	*Alpha-galactosidase (EC 3.2.1.22)*
TRINITY_DN8904_c0_g1_i4	*g7438_t1*	*7.39*	*GH35*	*No*	*beta-galactosidase, putative*

4.4 Conclusions

Cellulose and Avicel were identified as inducers with quickest response in *P. janthinellum*. Transcriptome of *P. janthinellum* 4h post-induction with cellulose showed CAZymes with significant differential expression and this accounted for 19.4 % of the total differentially expressed genes. These Unigenes included 14 cellulases (4 CBH, 6 EG, 4 BGL), 15 hemicellulases and 2 LPMO's. The early-expressed CAZymes identified from this study reflect only the initial response to cellulose. However, the CAZyme composition is expected to change at later time points or with different carbon sources. This early snapshot will serve as a foundation for studying cellulase regulation mechanisms in *P. janthinellum*.

Chapter 5

Identification of the probable regulators of CAZymes in *Penicillium janthinellum* NCIM1366

5.1 Introduction

In filamentous fungi, biomass hydrolyzing enzymes are under multi-stage stringent regulation because their synthesis and secretion is energy-intensive and resource-demanding. While the precise regulation mechanism of cellulases remain elusive, studies on different filamentous fungi-especially *T. reesei*, has identified some of the critical regulatory elements and some of these have become targets of strain engineering. Even though some of homologous regulators are conserved, their functions and the regulation mechanism differ even between species of the same genera. So the regulation mechanism cannot be interpolated and has to be studied at individual species level. Transcriptome profiling is an effective approach to study multi-layered cellulase regulation encompassing signaling, transport, and transcriptional control.

The most critical/studied level of cellulase gene regulation is at the transcriptional stage. Several transcription factors have been identified as either activators or repressors of gene expression. In *T. reesei* as well as in closely related species, the network is built around XYR1 (XlnR homolog) - the master activator of cellulase and hemicellulase genes, ACE proteins (ACE1, ACE2, ACE3 and ACE4), CRE1/CREA- repressor mediating carbon catabolite repression (CCR), and other specific transcription factors (BglR, VIB1, AZF1, CLP1, and RXE1) that fine-tune expression. Apart from these specific transcription factors, the secondary metabolism regulator LaeA and the chromatin status regulator complex HAP2/3/5 are also involved in cellulase regulation (Bischof et al., 2016; Yan et al., 2021). Environmental cues such as light (through the photoreceptor ENVOY and the BLR complex) and ambient pH (through PAC1) also affect cellulase production (Tisch and Schmoll, 2010; Häkkinen et al., 2015). In *Neurospora* and *Aspergilli*, the regulation is centered on positive transcription factors CLR-1/ClrA and CLR-2/ClrB with role of XlnR mostly limited to hemicellulase regulation (Coradetti et al., 2012; Raulo et al., 2016).

Sugar transporters are part of Major Facilitator Superfamily (MFS) and play pivotal role in cellulase regulation of filamentous fungi. They are involved in either the uptake of inducers such as glucose, xylose, and cellobiose into the cell or, which can act as sugar sensors. In *T. reesei*, STP1 transports cellobiose and glucose and CRT1 function as cellobiose transporter and transceptor that help activate XYR1(Zhang et al., 2013; Wang et al., 2022). In *Neurospora crassa*, CDT-1 and CDT-2 function as key cellodextrin transporters, with CDT-2 being essential for cellulase induction via the activation of transcription factors such as CLR-1 and CLR-2 (Cai et al., 2014). The signaling pathways are the intermediate between the sugar transporters and transcription factors. Certain pathways involved include heterotrimeric G-

protein-coupled receptors (GPCRs), the cyclic AMP-protein kinase A (cAMP-PKA) signaling cascade, the Mitogen-Activated Protein Kinase (MAPK) pathway, and the calcium–calcineurin signaling system.(Mattam et al., 2022a).

Cellulose is an inducing carbon source for cellulase expression and it is insoluble, thus unable to reach inside the cell. Cellulose degradation starts with its partial hydrolysis and release of soluble oligosaccharides. These molecules act as inducers, sensed by membrane-bound receptors or transported inside by specific sugar transporters, and trigger downstream signaling pathways. In this chapter, the transcriptome data under cellulose induction conditions was analyzed focusing on the expression dynamics of Transcription Factors, Sugar transporters and Signaling pathway genes. Through this integrative analysis, it was aimed to elucidate the probable regulatory mechanism underlying cellulase induction, including identification of novel transcription factors, transporters acting as inducers/transceptors, and signaling pathways modulating the transcriptional network. This will lay the foundation for predicting and engineering enhanced cellulase expression in *P. janthinellum*.

5.2 Materials and methods

5.2.1 Transcriptome data analysis and annotation

Transcriptome profiling and expression analysis data from Chapter 4 was used in this study. Annotation and expression data from Chapter 4 were used with the same fold change cut of $|\log_2 FC| \geq 2$ was used for selecting all the genes in this study.

5.2.2 Motif scanning for binding site analysis

Promoter regions (1 kb upstream of the transcription start site) of cellulase genes were extracted from the genome data. These promoter regions were scanned for the presence of the known binding motifs using the FIMO (Find Individual Motif Occurrences) tool from the MEME Suite with a significance threshold of $p < 0.001$(Grant et al., 2011).

5.2.3 Domain Analysis Using the Conserved Domain Database (CDD)

To functionally characterize the protein sequences of interest, domain analysis was performed using the Conserved Domain Database (CDD) available at the NCBI database (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml).

5.3 Results and discussion

5.3.1 Identification of differentially expressed transcription factors from the transcriptome data

Transcription factors (TFs) are the major determinants of cellulase regulation in filamentous fungi. Most of the studies on the cellulase regulation involve the identification of transcription factors (Mattam et al., 2022a). To identify the probable transcription factors in *P. janthinellum* that are involved in the regulation of cellulolytic machinery we have utilized the cellulose-induced transcriptome data. The expression analysis of TFs in *P. janthinellum* under cellulose-induction reveals significant transcriptional changes. We could identify a total of 36 transcription factors which are differentially expressed with $|\log_2 FC| \geq 2$, expected to be biologically relevent. It included 26 transcription factors which were markedly upregulated, with $\log_2$ fold change ($\log_2$ FC) values between +2.26 and +10.37, while 10 TFs were downregulated, showing $\log_2$ FC values from -2.21 to -10.91. The list of all the differentially expressed transcription factors and their $\log_2$ FC values are shown in Table 5.1. The highest upregulation was observed for a hypothetical gene (g3042_t1) with a $\log_2$ FC of +10.3 and the highest downregulation was shown by another hypothetical gene (g2815_t1) with a log2 (FC) value of -10.9. The high differential expression values for these genes suggest that they may play important roles in cellulase regulation and also indicate a possibility as novel cellulase regulators. From the list, orthologs of some of the important cellulase regulators from other filamentous fungi were also identified. Cellulolytic activator ClrB (*g4518_t1*, $\log_2$ FC = +2.26) and Xylanolytic activator XlnR (*g8827_t1*, $\log_2$ FC = +3.84) showed moderate but significant upregulation, confirming their conserved function in regulating cellulases. LaeA (*g7355_t1*, $\log_2$ FC = +8.28), a global regulator of secondary metabolism, showed significant upregulation which reinforces its role in cellulase regulation similar to other fungi. Involvement of other pathways in enhancing biomass degradation was identified by the upregulation of additional TFs involved in nitrogen metabolism (Nit4, g11678_t1, $\log_2$ FC = +3.26).

Downregulated transcription factors involved in glucose repression, stress response, and alternative metabolic pathways were identified from the transcriptome data. It included glucose-responsive repressor Rgt1 (*g4218_t1*, $\log_2$ FC = -8.97), the UPR pathway bZIP TF HacA (*g7295_t1*, $\log_2$ FC = -9.17), and the calcium-responsive regulator Prz1 (*g4413_t1*, $\log_2$ FC = -9.88). These results are indicative of the activation of the cellulolytic system upon depletion of glucose (Portnoy et al., 2011). AraR (TRINITY_DN17015), which is a regulator

of arabinan degradation showed upregulation with log_2 FC value of +7.51. Transcription factors with log_2 FC values above ±8, which include *LaeA*, *Rgt1*, and *HacA* are highly likely to be involved in cellulose utilization making these TFs promising targets for strain engineering in *P. janthinellum*.

Table 5.1 List of differentially expressed transcription factors identified from the transcriptome data of *P. janthinellum* under cellulose induction

geneID	logFC	Annotation
TRINITY_DN6843_c0_g1_i1	10.36	hypothetical protein PMG11_09846
TRINITY_DN9813_c0_g1_i2	10.23	Zn(2)-C6 fungal-type DNA-binding domain
TRINITY_DN8779_c0_g5_i1	9.99	hypothetical protein PMG11_10186
TRINITY_DN8719_c0_g1_i2	8.28	Secondary metabolism regulator LAE1
TRINITY_DN6700_c0_g1_i2	8.25	hypothetical protein PDE_07001
TRINITY_DN5665_c0_g3_i1	8.13	Putative Zn cluster transcription factor
TRINITY_DN4063_c0_g2_i1	7.78	hypothetical protein PMG11_08776
TRINITY_DN9286_c0_g1_i1	7.73	Putative Transcription factor cys6
TRINITY_DN17015_c0_g1_i1	7.51	Arabinanolytic transcriptional activator araR
TRINITY_DN9609_c0_g1_i2	7.45	hypothetical protein PMG11_10205
TRINITY_DN1979_c0_g1_i1	3.84	Xylanolytic transcriptional activator xlnR
TRINITY_DN9609_c0_g1_i1	3.52	Asperfuranone cluster transcription factor afoA
TRINITY_DN8779_c0_g1_i1	3.37	hypothetical protein PMG11_10186
TRINITY_DN5719_c0_g1_i1	3.25	Nitrogen assimilation transcription factor nit-4
TRINITY_DN4768_c0_g1_i1	3.22	Putative C6 transcription factor
TRINITY_DN7417_c0_g1_i2	3.09	Putative C6 transcription factor
TRINITY_DN9767_c0_g1_i23	2.91	hypothetical protein PMG11_09729
TRINITY_DN9386_c0_g1_i5	2.88	Transcriptional regulatory protein moc3
TRINITY_DN2044_c0_g1_i1	2.78	Putative transcription factor kapC
TRINITY_DN8347_c0_g1_i2	2.64	hypothetical protein
TRINITY_DN4045_c0_g2_i1	2.60	specific transcription factor domain protein
TRINITY_DN9767_c0_g1_i8	2.55	hypothetical protein PMG11_09729
TRINITY_DN5507_c0_g1_i1	2.55	hypothetical protein PMG11_05290
TRINITY_DN1590_c0_g1_i1	2.46	hypothetical protein PMG11_09057
TRINITY_DN4818_c0_g1_i1	2.41	Transcription factor vrtR2
TRINITY_DN7417_c0_g1_i1	2.38	Cutinase transcription factor 1 beta
TRINITY_DN16478_c0_g1_i1	2.26	Putative Cellulolytic transcriptional activator ClrB
TRINITY_DN8940_c0_g1_i2	-2.21	Putative Fungal Zn binuclear cluster domain containing protein
TRINITY_DN8880_c0_g1_i2	-3.12	hypothetical protein PMG11_02836
TRINITY_DN9767_c0_g1_i2	-7.41	hypothetical protein PMG11_09729
TRINITY_DN5665_c0_g3_i2	-7.94	Putative Zn cluster transcription factor
TRINITY_DN9767_c0_g1_i14	-8.63	hypothetical protein PMG11_09729 [Penicillium brasilianum]
TRINITY_DN9693_c0_g4_i1	-8.96	Glucose transport transcription regulator RGT1
TRINITY_DN8422_c0_g1_i3	-9.16	TPA: bZIP transcription factor HacA
TRINITY_DN8076_c0_g1_i8	-9.87	Putative Transcriptional regulator prz1
TRINITY_DN9767_c0_g1_i27	-10.91	hypothetical protein PMG11_09729

5.3.2 A dual-domain protein with both transporter and transcription factor domains could be a novel regulator

Domain prediction of the proteins corresponding to the highly differentially expressed transcription factors were performed using the NCBI Conserved Domain Database (CDD) with default parameters. Interestingly, one of the transcription factors was identified to contain a transporter domain suggesting dual function. The gene was *g8937* (TRINITY_DN9813_c0_g1_i2) with $\log_2$ FC value of 10.23 and the domain analysis data is shown in Figure 5.1. The protein contains a Sugar_tr (sugar transporter) domain belonging to the Major Facilitator Superfamily (MFS) at the N-terminal region. In the C-terminal region, two conserved domains which are characteristic of fungal transcription factors-GAL4-like (GAL) domain and a fungal transcription factor middle homology region (fungal_TF_MHR)-were identified. So, we hypothesize that *g8937* may act as a sugar-responsive regulator, which connects the environmental cellulose signaling to transcriptional regulation in *P.janthinellum*. Further functional validation would be necessary to confirm its regulatory role in cellulase production.

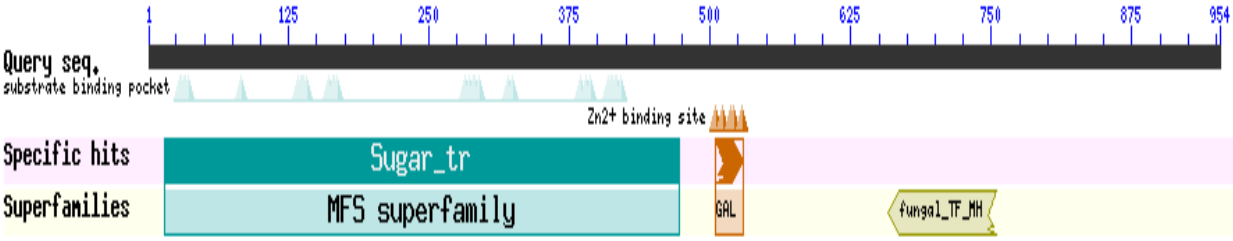


Figure 5.1 Conserved domain data for the dual-domain protein (g8937) identified from transcriptome

5.3.3 Promoter Analysis of Differentially Expressed Cellulases for Predicted Binding Sites

Correlating the gene expression levels from the transcriptome data with the transcription factor binding sites is a strategy for the identification of key transcription factors (Huang et al., 2024). In this study, we have selected differentially expressed transcription factors XlnR and ClrB whose binding sites are already known. In an *in-silico* study, 5'-GGC(A/T)(3)-3' motifs are identified as XYR1/XlnR binding site and it was widespread in the 5'-upstream region of all the XYR1-regulated genes (Furukawa et al., 2009). The consensus binding site of ClrB in *A. nidulans* was identified as 5′-CGGN$_8$CCG-3′(Li et al., 2016). In the current study, sequences 1 kb upstream of the transcription start site (TSS) of cellulase and hemicellulase

genes were extracted from the genome as promoter. Earlier studies have identified that promoter activity of filamentous fungi are generally contained within this length (Punt et al., 1990). Binding site prediction was performed using FIMO from the MEME Suite with a significance threshold of $p < 0.001$(Bailey et al., 2009; Grant et al., 2011).

Surprisingly, we couldn't identify any XlnR binding sequences in the promoters of cellulase or hemicellulose genes. This may be due to the fact that the XlnR binding site differs among different genera or even between species in filamentous fungi, and the binding site is not yet identified in *Penicillium* genus (Van Peij et al., 1998; Klaubauf et al., 2014). However, ClrB binding motifs were detected in key cellulase and hemicellulose genes as represented in Figure 5.2.

The presence ClrB binding sites in the promoter regions of key cellulase genes is shown in Figure 5.2a. Among the four upregulated BGL genes, three genes contained ClrB binding sites, supporting ClrB's role as a positive regulator of β-glucosidase expression. All the four upregulated CBH genes showed the presence of ClrB binding sites. In case of Endoglucanase (EG) also, ClrB binding sites were found in all the upregulated genes. The distribution of these binding sites both near and distal to the TSS may reflect different regulatory strengths or involvement of co-regulators. So, it is hypothesized that the key cellulase genes are under direct ClrB-mediated activation. As shown in Figure 5.2b, ClrB motifs were also detected upstream of some of hemicellulase genes. However, only six of the upregulated genes contained a putative ClrB binding site. The presence of binding sites across the upregulated cellulase and hemicellulase genes suggests the role of ClrB as a master transcriptional activator of cellulolytic enzymes similar to its ortholog Clr2 in *Neurospora crassa* (Coradetti et al., 2013; Craig et al., 2015).

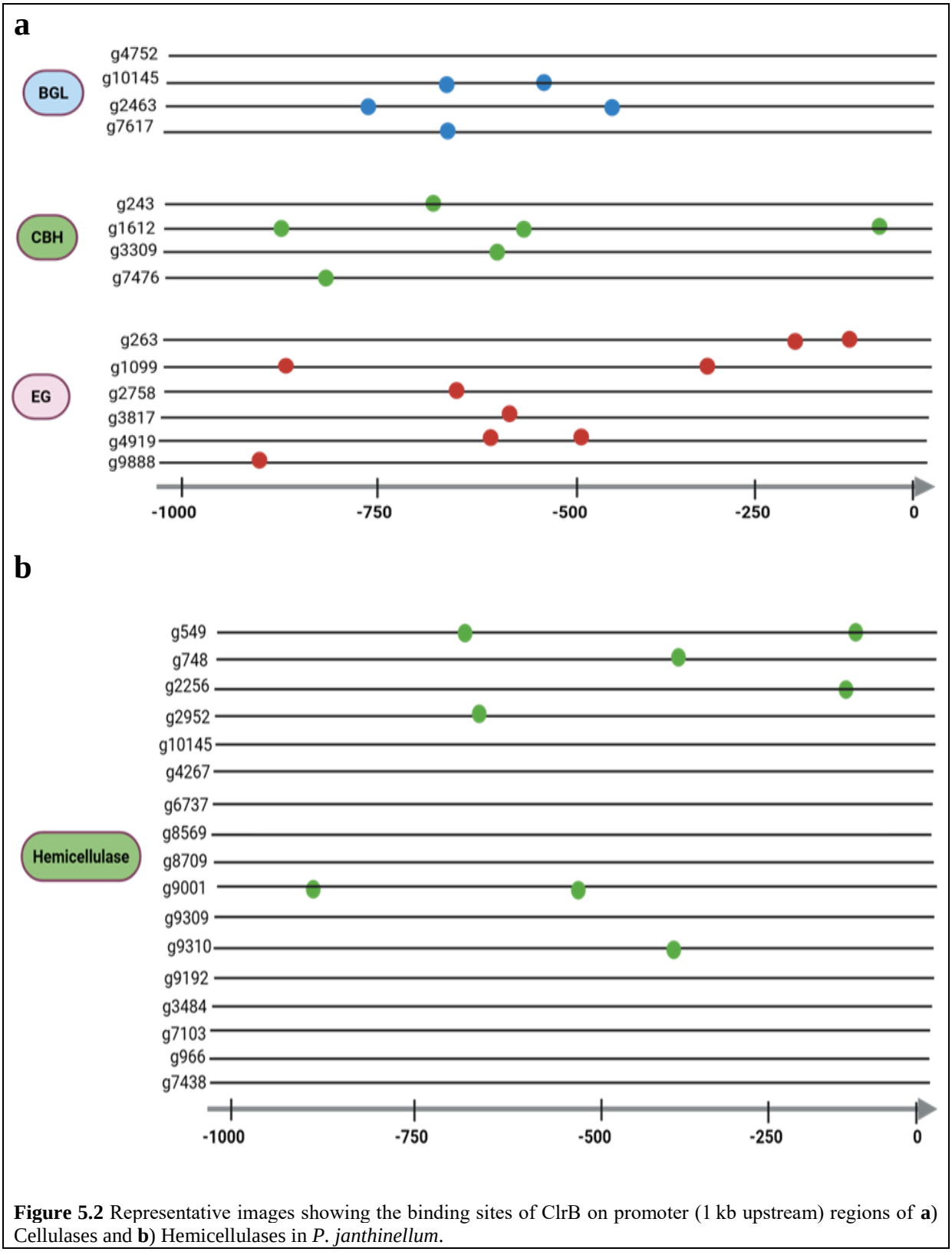


Figure 5.2 Representative images showing the binding sites of ClrB on promoter (1 kb upstream) regions of **a**) Cellulases and **b**) Hemicellulases in *P. janthinellum*.

5.3.4 Promoter Analysis of Differentially Expressed Transcription Factors for Predicted Binding Sites

Cross-regulation between transcription factors is identified as a mechanism that fine-tunes the expression of lignocellulolytic genes in filamentous fungi (Mattam et al., 2022b). So the promoters of differentially expressed transcription factors were investigated for the binding sites. ClrB binding sites were detected in the promoter of XlnR represented in Figure 5.3a.

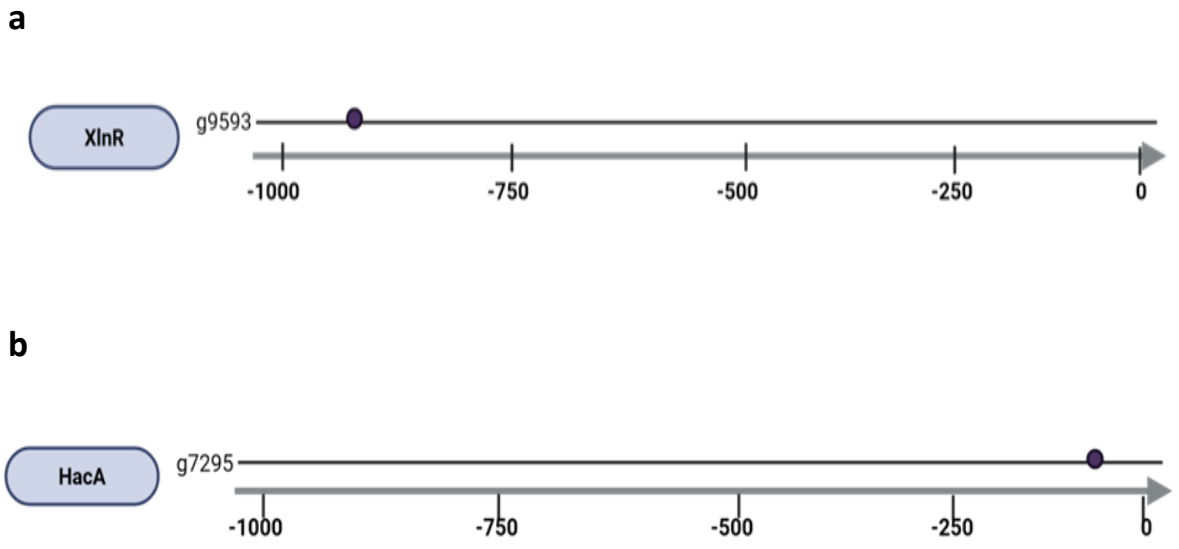


Figure 5.3 Representative image showing the binding sites of ClrB on promoter (1 kb upstream) **a)** XlnR and **b**) HacA in *P. janthinellum*.

This suggests that ClrB may directly regulate XlnR expression, providing a hierarchical layer of control where ClrB acting upstream of XlnR during cellulase regulation. A ClrB binding motif was also found very close to the TSS or within the 5' UTR of HacA (Figure 5.3b). HacA is a central regulator of the unfolded protein response (UPR) and is activated under conditions of high secretory stress, such as massive enzyme production. The presence of a ClrB motif here suggests that ClrB might modulate HacA expression to ensure cellular stress responses are coordinated with enzyme induction, especially during active cellulase secretion. This is particularly relevant, as cellulase overproduction is known to induce endoplasmic reticulum stress, and the UPR is essential for maintaining protein folding and secretion efficiency (Saloheimo and Pakula, 2012). The potential regulation of *hacA* by ClrB links cellulase gene induction with cellular homeostasis and stress management.

Together, these observations highlight that ClrB is not only a direct activator of structural cellulase genes but also modulates a broader transcriptional network by regulating other major regulators such as XlnR and HacA. This reflects a multi-tiered regulatory architecture**,** wherein a master regulator like ClrB amplifies the cellulolytic response both directly (on enzyme genes) and indirectly (through other TFs)**.** This insight is valuable for metabolic engineering strategies aimed at enhancing cellulase production by targeting both direct enzyme promoters and upstream regulatory nodes.

5.3.5 Identification of differentially expressed sugar transporters from the transcriptome data

Sugar transporters (STs) are generally involved in the uptake the monomeric sugars generated from complex polysaccharides by the action of cellulases into the cell. Apart from this, they also regulate cellulase production in filamentous fungi through a combination of environmental sensing of sugars and intracellular signaling (Xu et al., 2024). Most of the characterized fungal STs belong to the Major Facilitator Superfamily of transporters (MFS). We could identify 79 transporter genes from the differentially expressed transcriptome data. Among them, 29 were identified as sugar transporters with highly significant fold changes and the list is shown in Table 5.2.

Table 5.2 List of differentially expressed sugar transporters identified from the transcriptome data of *P. janthinellum* under cellulose induction

geneID	Log_2 FC	Annotation
TRINITY_DN7877_c0_g2_i3	9.85	MFS monosaccharide transporter
TRINITY_DN11829_c0_g1_i1	8.48	maltose permease
TRINITY_DN7618_c0_g1_i2	7.68	Putative Hexose carrier protein
TRINITY_DN19347_c0_g1_i1	7.63	Maltose transport protein MAL11
TRINITY_DN25248_c0_g1_i1	7.20	Maltose transport protein MAL31
TRINITY_DN22208_c0_g1_i1	6.97	Putative Cellodextrin transporter cdt-c
TRINITY_DN8495_c0_g1_i4	3.35	Putative MFS transporter
TRINITY_DN23621_c0_g1_i1	4.58	Sugar transporter STL1
TRINITY_DN15975_c0_g1_i1	3.91	Putative Hexose transporter
TRINITY_DN9187_c0_g1_i1	3.68	Putative Maltose permease
TRINITY_DN8036_c0_g1_i1	3.54	Putative Hexose transporter
TRINITY_DN4780_c0_g1_i1	3.50	Putative Sugar transporter
TRINITY_DN8495_c0_g1_i6	5.81	Putative MFS transporter
TRINITY_DN21865_c0_g1_i1	3.33	Putative Hexose transporter protein -HGT-1
TRINITY_DN3418_c0_g1_i1	3.22	Hexose transporter(cdt-2)
TRINITY_DN11777_c0_g1_i1	2.76	Putative Sugar transporter
TRINITY_DN9195_c0_g2_i4	2.65	Putative MFS sugar transporter
TRINITY_DN6865_c0_g2_i2	2.59	Putative MFS sugar transporter
TRINITY_DN16822_c0_g1_i1	2.34	Sugar transporter STL1

TRINITY_DN22354_c0_g1_i1	*2.31*	*Maltose transport protein MAL11*
TRINITY_DN16217_c0_g1_i1	*2.29*	*Sugar transporter STL1*
TRINITY_DN5398_c0_g1_i2	*2.21*	*High-affinity fructose transporter ght6*
TRINITY_DN9657_c1_g1_i1	*-2.46*	*Low-affinity glucose transporter HXT1*
TRINITY_DN22576_c0_g1_i1	*-2.66*	*Putative Sugar transporter*
TRINITY_DN15733_c1_g1_i1	*-3.17*	*Probable glucose transporter rco-3*
TRINITY_DN26204_c0_g1_i1	*-4.94*	*Putative MFS monosaccharide transporter*
TRINITY_DN4295_c0_g1_i1	*-5.36*	*Putative maltose permease*
TRINITY_DN621_c0_g1_i1	*-5.36*	*Putative Sugar transport protein*
TRINITY_DN23073_c0_g1_i1	*-9.08*	*Putative Sugar transport protein [Penicillium brasilianum]*

Among the identified sugar transporters, majority showed upregulation with $\log_2$ FC values ranging from +9.85 to +2.21. Seven transporters were downregulated with $\log_2$ FC values ranging from -2.46 to -9.08. The upregulated sugar transporters included multiple putative MFS transporters, hexose/monosaccharide transporters and maltose permeases. The genes TRINITY_DN22208_c0_g1_i1 ($\log_2$ FC = +6.97) and TRINITY_DN3418_c0_g1_i1 ($\log_2$ FC = +3.22) were identified to be the Cellodextrin transporters *cdt-c* and *cdt-2* respectively. These transporters are involved in the transport of cellodextrins (the soluble products of cellulose hydrolysis) in filamentous fungi and reported to be involved in cellulase induction. Orthologs of these transporters are characterized in *N. crassa* (CDT-1 and CDT-2) and *P. oxalicum* (CdtC, CdtD and CdtG) (Li et al., 2013; Cai et al., 2014). Also, CDT-2 is known to function as a transceptor (both a transporter and a sensor) in *N. crassa* that not only imports cellodextrins but act as a sensor to these substrates triggering cellulase induction (Znameroski et al., 2013). A similar function is suggested for this gene product in *P. janthinellum*, given that it is upregulated in the early stages of cellulose induction, but requires further experimental validation.

Multiple genes annotated as *stl1* was upregulated in this study- TRINITY_DN23621_c0_g1_i1 ($\log_2$ FC =+4.58), TRINITY_DN16217_c0_g1_i1 ($\log_2$ FC =+2.29), and TRINITY_DN16822_c0_g1_i1 ($\log_2$ FC =+2.34). STL1 primarily act as an active glycerol importer and osmoregulator in *Saccharomyces cerevisiae*, however its role is not characterized in any other filamentous fungi (Klein et al., 2017). Differential expression of several maltoses permeases/in this study underscores the broader regulation of carbohydrate import pathways in response to cellulose.

High-affinity hexose transporters which are adapted to efficiently import glucose under carbon-limited conditions- *hgt-1* and *ght6* were upregulated with $\log_2$ FC values of +3.33 and

+2.21 respectively. Also a low affinity glucose transporter *hxt1* was downregulated with $\log_2$ FC= (-2.46). This type of regulation is typical of a "dual-affinity hexose transport system"- a conserved strategy used by microorganisms to adapt with fluctuations in nutrient availability in the environment (Leandro et al., 2009). These are important regulatory components which help in the metabolic shift towards lignocellulosic substrates in filamentous fungi (Wang et al., 2017). The downregulated gene *rco-3* (TRINITY_DN15733_c1_g1_i1), annotated as a probable glucose transporter, exhibited a $\log_2$ FC of -3.17. RCO-3 was earlier identified as a non-transporting high affinity glucose sensor that regulates the dual-affinity glucose transport system in *Neurospora crassa* (Li et al., 2021).

These findings indicate a complex network of sugar transporters involving both known homologs and uncharacterized proteins in *P. janthinellum* and are promising targets for improving lignocellulose degradation.

5.3.6 Identification of differentially expressed signaling genes from the transcriptome data

Cellulase/CAZyme gene regulation in filamentous fungi is coordinated by signaling pathways that sense nutrient availability signals and relay it into the effectors. The signaling mechanism is relatively less studied compared to the well-known roles of transcription factors or sugar transporters. However, signal transduction routes such as MAPK signaling pathway, GPCR-cAMP/PKA signaling pathway and Calcium-Calmodulin-Calcineurin pathway are known to be involved in cellulase regulation (de Paula et al., 2018; Hu et al., 2018; Zhao et al., 2022).

Table 5.3 List of differentially expressed signaling genes identified from the transcriptome data of *P. janthinellum* under cellulose induction

geneID	Log_2FC	Annotation
TRINITY_DN10023_c0_g5_i14	8.3	Putative protein kinase
TRINITY_DN9728_c0_g1_i4	8.01	Inositol hexaphosphate multikinase kcs1
TRINITY_DN17675_c0_g1_i1	7.97	Uncharacterized kinase
TRINITY_DN14934_c0_g1_i1	7.73	G protein-coupled receptor
TRINITY_DN9954_c1_g2_i1	7.57	Serine/threonine protein kinase
TRINITY_DN19547_c0_g1_i1	7.51	cAMP-dependent protein kinase
TRINITY_DN5814_c0_g1_i1	6.93	hypothetical protein PMG11_01990
TRINITY_DN19556_c0_g1_i1	4.90	Putative Ca2 transporting ATPase
TRINITY_DN5345_c0_g1_i1	4.88	Putative SNF-1 like protein kinase
TRINITY_DN21924_c0_g1_i1	2.86	hypothetical protein PMG11_10611
TRINITY_DN7848_c0_g1_i1	2.80	Ribosomal S6 kinase homolog psk1

TRINITY_DN7206_c0_g1_i3	*-2.90*	*Serine/threonine-protein kinase csk1*
TRINITY_DN3201_c0_g1_i1	*-2.90*	*Calcium/calmodulin-dependent protein kinase*
TRINITY_DN4750_c0_g1_i1	*-3.30*	*dis1-suppressing protein kinase dsk1*
TRINITY_DN8718_c0_g1_i1	*-4.35*	*Ras-related small GTPase*

Transcriptome profiling data revealed extensive reprogramming of signaling pathways, as shown by the differential expression of key signaling-related genes in *P. janthinellum* as shown in Table 5.3. The majority of these signaling genes are up regulated with $\log_2$ FC values as high as +8.3, signifying activation of intracellular signal transduction pathways during adaptation to cellulose. The upregulated genes include multiple kinases which are central to essential signaling cascades. A G protein coupled receptor (GPCR) ($\log_2$ FC =+7.73) and cAMP-dependent protein kinase ($\log_2$ FC =+7.51) were strongly upregulated indicating the involvement of GPCR-cAMP/PKA signaling pathway consistent with its role in cellulase induction in *N. crassa* and *T. reesei* (Schmoll and Kubicek, 2003; Collier et al., 2020). Up regulated SNF1-like protein kinase ($\log_2$ FC =+4.88) is homologous to AMPK/SNF1, which is a master regulator of carbon source adaptation, promoting the de-repression of cellulase genes in low-glucose environments (Suto and Tomita, 2001). Downregulated genes include CSK1 (TRINITY_DN7206; −2.9), Ca^{2+}/calmodulin-dependent kinase (DN3201; −2.9), Dsk1 (−3.3), and Ras-related small GTPase (DN8718; −4.35). Ras-related small GTPases are known to regulate cellulases in *T. reesei* (Zhang et al., 2012). The putative Ca2+ transporting ATPase (Log2 FC=4.9) and downregulation of calcium/calmodulin-dependent protein kinases (Log2 FC=-2.9) suggest a reconfiguration of calcium signaling pathways during cellulose induction in *P.janthinellum*.

5.2.4 Probable Cellulase Regulation mechanism in *P. janthinellum*: A Hypothetical Model

Based on the transcriptome analysis under cellulose induction, a probable cellulase regulation model is proposed for *P. janthinellum* as shown in Figure 5.4. The model involves the partial extracellular hydrolysis of cellulose into soluble cellodextrins by basal levels of secreted enzymes. These are then imported into the cell and/or sensed by upregulated Cellodextrin transporters (Cdt-c, Cdt-2) and other MFS-type transporters. This activates multiple signaling pathways, such as GPCR/cAMP-PKA, SNF1/AMPK, and calcium signaling. These signals converge to activate core transcriptional regulators, prominently ClrB. This protein functions as a master activator by directly binding to the promoters of major cellulase genes and also

regulating other key transcription factors such as XlnR and HacA. HacA, and as part of the UPR, ensures efficient protein folding and secretion, allowing the cell to handle increased enzyme output and ER stress. Carbon catabolite repression on cellulase genes is relieved through regulation of Rgt1. The integration of sugar uptake, intracellular signaling, transcription control, and ER quality control enables *P. janthinellum* to finely tune cellulase induction in response to cellulose, providing a framework for both fundamental understanding and industrial optimization of fungal enzyme production.

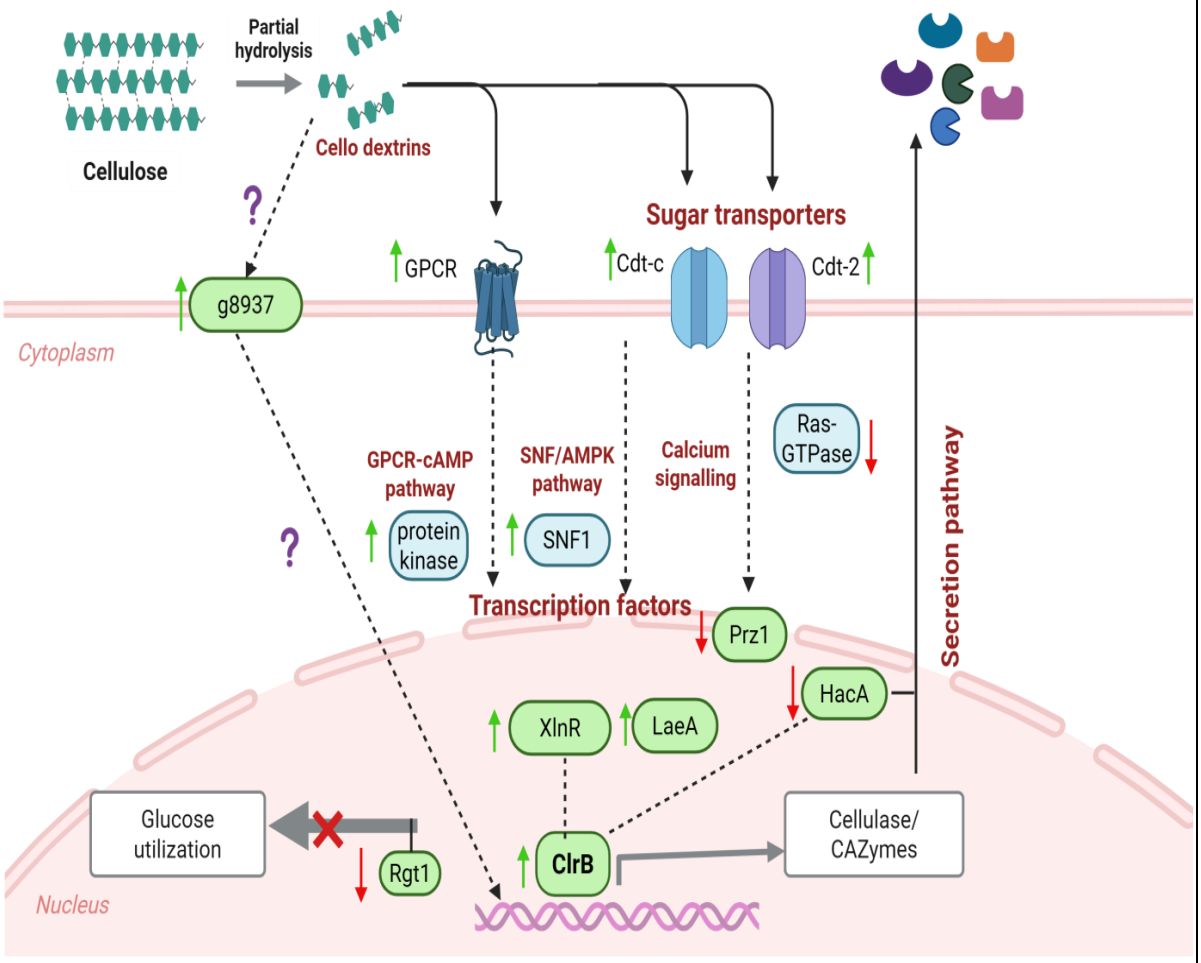


Figure 5.4 Schematic diagram showing the major cellulase regulatory elements detected through transcriptomics study in *P. janthinellum* and their speculative interrelationships (Created in https://BioRender.com)

5.4 Conclusions

In this chapter, the factors co-expressed with CAZymes under cellulose induction in *Penicillium janthinellum* NCIM1366 were identified through comprehensive transcriptome profiling. The analysis revealed a probable regulatory network encompassing sugar sensing and transport, signal transduction via diverse kinase cascades, and a hierarchical system of transcription factors. We propose a mechanism in which ClrB acts as a master activator, directly regulating major cellulase and hemicellulase genes and also controlling additional regulators such as XlnR and HacA. However, these conclusions are based on transcriptomic and *in silico* analyses, and further mutational studies are required to confirm the precise roles of these factors. Overall, this study provides a valuable framework for understanding the regulation of CAZyme expression in *P. janthinellum* and lays the groundwork for further optimization of enzyme production.

Chapter 6

Conclusions and Future Perspectives

6.1 Summary of the study

A major proportion of the overall cost factor contributing to the production of bioethanol from lignocellulosic biomass comes from the hydrolysis stage, mainly because of the high cost of the enzymes used to break down the biomass, particularly cellulases. The major producers of these enzymes are filamentous fungi from genera *Trichoderma, Aspergillus* and *Penicillium* with *Trichoderma reesei* being the current best industrial cellulase producer. Although significant enhancement in the enzyme production by these fungi has been achieved over the past years through random/rational strain improvement, the commercially feasibility of lignocellulosic biomass waste to bioethanol conversion still remains limited. A filamentous fungal strain *P. janthinellum* NCIM1366, developed through classical mutagenesis at CSIR-NCL, Pune has demonstrated efficient cellulase production for the hydrolysis of natural biomass substrates. With further advances in enzyme production strategies and targeted strain improvements, it shows considerable promise for development as an industrial strain for cellulase production. The present study is intended to elucidate the CAZyme profile and cellulase gene regulation of this relatively unexplored hypercellulolytic organism through Multi-Omics analysis towards targeted engineering of this organism to enhance its cellulase production and secretion capability.

This study presents the first secretome analysis of *Penicillium janthinellum* NCIM1366 and compares it with the well-known cellulase hyper-producing industrial strain, *Trichoderma reesei* RUT-C30. The analyses have highlighted the better hydrolytic efficiency, enzyme activity, protein production, and secretion efficiency of *P. janthinellum*, in comparison to the industrial work horse – *Trichoderma reesei* RUT-C30, which indicates its potential as a future industrial cellulase producer. To further understand its CAZymes, the complete genome of *Penicillium janthinellum* NCIM1366 was sequenced and analyzed. The genome, at 37.6 Mbp, encodes an extensive array of carbohydrate-active enzymes (CAZymes), comprising approximately 5.4% of all predicted genes; notably higher than the CAZyme content reported in other industrially relevant fungi, including *Trichoderma reesei*. This remarkable enrichment, particularly in glycoside hydrolases, underscores the organism's evolutionary adaptation toward efficient plant biomass utilization. The fungus was found to possess an extensive collection of cellulolytic genes, encompassing several endoglucanases, cellobiohydrolases, β-glucosidases, lytic polysaccharide monooxygenases (LPMOs), and hemicellulases, highlighting the genetic basis for its superior biomass hydrolyzing ability. The genome also revealed a

substantial number of transcription factors, with a striking dominance of GAL4-like Zn_2Cys_6 binuclear cluster DNA-binding domain proteins, alongside key orthologs of well-characterized cellulase regulators such as XlnR/Xyr1, CreA/Cre1, ClrB, Ace1, Ace3, and others.

In an induction study, Cellulose and Avicel® were identified as inducers with quickest cellulase response in *P. janthinellum*. Transcriptome of *P. janthinellum* 4 h post-induction with cellulose showed CAZymes with significant differential expression which accounts for 19.4 % of the total differentially expressed genes. This included 14 cellulases (4 CBH, 6 EG, 4 BGL), 15 hemicellulases and 2 LPMO's reflecting the initial response to cellulose. Proteins co-expressed with CAZymes were also identified, and the analysis revealed a probable regulatory network encompassing sugar sensing and transport, signal transduction via diverse kinase cascades, and a hierarchical system of transcription factors. We propose a mechanism in which ClrB acts as a master activator, directly regulating major cellulase and hemicellulase genes and also controlling additional regulators such as XlnR and HacA. Overall, this study provides a valuable framework for understanding the regulation of CAZyme expression in *P. janthinellum* and lays the groundwork for further optimization of enzyme production, through engineering of the fungus. These results highlight *P. janthinellum* as a very promising option for industrial use in biomass conversion and biofuel production. Further, targeted genetic modifications are expected to improve its performance even more, providing a worthy alternative for *T. reesei*, or complement it in the cellulase applications for biomass conversion.

6.2 Major conclusions

- **High extracellular proteins and enzyme production**

P. janthinellum showed superior extracellular protein production, cellulase enzyme activities and biomass hydrolysis potential compared to *T. reesei*.

- **Rich CAZyme repertoire in the genome**

The genome encoded 657 CAZyme genes including 391 glycoside hydrolases. This is substantially higher than in *T. reesei* or other *Penicillium/Aspergillus* species, indicating genetic enrichment for plant cell wall degradation.

- **Transcriptome reveals strong CAZyme induction by cellulose**

Over 11,000 transcripts were upregulated under cellulose conditions, including major cellulases (endoglucanases, cellobiohydrolases, β-glucosidases), hemicellulases, and oxidative enzymes (LPMOs).

- **ClrB as master regulator**

In-silico analysis of promoter motifs verified the presence of ClrB binding sites on nearly all upregulated cellulase and hemicellulase genes, along with the promoters of *xlnR* and *hacA*, indicating a hierarchical regulatory framework.

- **Novel transcriptional regulator with dual-domain architecture**

A highly upregulated protein ($log_2FC = 10.23$) with both a sugar transporter domain and a Zn2Cys6 transcription factor domain was discovered, suggesting a direct sugar-sensing transcriptional regulatory role.

- **Coordinated repression of alternative pathways**

Glucose-repressive TFs such as Rgt1 and UPR regulator HacA were significantly downregulated, indicating a regulatory switch toward biomass degradation.

6.3 Future Perspectives

- **Functional validation of regulatory elements**

Further studies using gene knockout or overexpression are needed to confirm the role of ClrB, the novel dual-domain regulator, and other TFs in cellulase regulation.

- **Promoter engineering and TF manipulation**

Identified high-expression TFs and strong native promoters can be used to engineer *P. janthinellum* strains with improved enzyme yields.

- **Industrial strain development**

The insights from proteomic, transcriptomic and genomic analysis can be utilized for rational metabolic engineering to enhance cellulase titers, thermostability, and performance on industrial substrates.

- **Synthetic biology approaches**

Incorporation of synthetic regulatory circuits using the characterized TFs may enable tunable, context-dependent cellulase expression.

REFERENCES

Adav, S. S., Chao, L. T., & Sze, S. K. (2012). Quantitative secretomic analysis of *Trichoderma reesei* strains reveals enzymatic composition for lignocellulosic biomass degradation. *Molecular & CellularProteomics*, *11*. https://doi.org/10.1074/mcp.M111.012419

Adsul, M. G., Bastawde, K. B., Varma, A. J., & Gokhale, D. V. (2007). Strain improvement of *Penicillium janthinellum* NCIM 1171 for increased cellulase production. *Bioresource Technology*, *98*, 1467–1473. https://doi.org/10.1016/j.biortech.2006.02.036

Adsul, M. G., Ghule, J. E., Shaikh, H., Singh, R., Bastawde, K. B., & Gokhale, D. V., et al. (2005). Enzymatic hydrolysis of delignified bagasse polysaccharides. *Carbohydrate Polymers*, *62*, 6–10. https://doi.org/10.1016/j.carbpol.2005.07.010

Adsul, M. G., Ghule, J. E., Singh, R., Shaikh, H., Bastawde, K. B., & Gokhale, D. V., et al. (2004). Polysaccharides from bagasse: Applications in cellulase and xylanase production. *Carbohydrate Polymers*, *57*, 67–72. https://doi.org/10.1016/j.carbpol.2004.04.001

Adsul, M. G., Terwadkar, A. P., Varma, A. J., & Gokhale, D. V. (2009). Cellulases from *Penicillium janthinellum* mutants: Solid-state production and their stability in ionic liquids. *BioResources*, *4*, 1670–1681. https://doi.org/10.15376/biores.4.4.1670-1681

Aghcheh, R. K., Németh, Z., Atanasova, L., Fekete, E., Paholcsek, M., Sándor, E., et al. (2014). The VELVET A orthologue VEL1 of *Trichoderma reesei* regulates fungal development and is essential for cellulase gene expression. *PLoS One*, *9*, e112799. https://doi.org/10.1371/journal.pone.0112799

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*, 420–423. https://doi.org/10.1038/s41587-019-0036-z

Amore, A., Giacobbe, S., & Faraco, V. (2013). Regulation of cellulase and hemicellulase gene expression in fungi. *Current Genomics*, *14*, 230–249. https://doi.org/10.2174/1389202911314040002

Armendáriz-Ruiz, M., Rodríguez-González, J. A., Camacho-Ruíz, R. M., & Mateos-Díaz, J. C. (2018). Carbohydrate esterases: An overview. In *Methods in Molecular Biology* (pp. 39–68). Humana Press Inc. https://doi.org/10.1007/978-1-4939-8672-9_2

Aro, N., Ilmén, M., Saloheimo, A., & Penttilä, M. (2003a). ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression. *Applied and Environmental Microbiology*, *69*, 56–65. https://doi.org/10.1128/AEM.69.1.56-65.2003

Aro, N., Saloheimo, A., Ilmén, M., & Penttilä, M. (2001). ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of *Trichoderma*

reesei. *Journal of Biological Chemistry*, *276*, 24309–24314. https://doi.org/10.1074/jbc.M003624200

Atreya, M. E., Strobel, K. L., & Clark, D. S. (2016). Alleviating product inhibition in cellulase enzyme Cel7A. *Biotechnology and Bioengineering*, *113*, 330–338. https://doi.org/10.1002/bit.25809

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, *37*. https://doi.org/10.1093/nar/gkp335

Bajpai, P. (2016). Structure of lignocellulosic biomass. In *Biotechnology for pulp and paper processing* (pp. 7–12). https://doi.org/10.1007/978-981-10-0687-6_2

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*, 455–477. https://doi.org/10.1089/cmb.2012.0021

Berlin, A., Maximenko, V., Gilkes, N., & Saddler, J. (2007). Optimization of enzyme complexes for lignocellulose hydrolysis. *Biotechnology and Bioengineering*, *97*, 287–296. https://doi.org/10.1002/bit.21238

Betts, W. B., Dart, R. K., Ball, A. S., & Pedlar, S. L. (1991). Biosynthesis and structure of lignocellulose. In *Lignocellulose biotechnology* (pp. 139–155). https://doi.org/10.1007/978-1-4471-3470-1_7

Biely, P. (2012). Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances*, *30*, 1575–1588. https://doi.org/10.1016/j.biotechadv.2012.04.010

Bischof, R., Fourtis, L., Limbeck, A., Gamauf, C., Seiboth, B., & Kubicek, C. P. (2013). Comparative analysis of the *Trichoderma reesei* transcriptome during growth on the cellulase inducing substrates wheat straw and lactose. *Biotechnology for Biofuels*, *6*. https://doi.org/10.1186/1754-6834-6-127

Bischof, R. H., Ramoni, J., & Seiboth, B. (2016a). Cellulases and beyond: The first 70 years of the enzyme producer *Trichoderma reesei*. *Microbial Cell Factories*, *15*. https://doi.org/10.1186/s12934-016-0507-6

Bischof, R. H., Ramoni, J., & Seiboth, B. (2016b). Cellulases and beyond: The first 70 years of the enzyme producer *Trichoderma reesei*. *Microbial Cell Factories*, *15*. https://doi.org/10.1186/s12934-016-0507-6

Borin, G. P., Sanchez, C. C., de Santana, E. S., Zanini, G. K., dos Santos, R. A. C., de Oliveira Pontes, A., et al. (2017). Comparative transcriptome analysis reveals different strategies for degradation of steam-exploded sugarcane bagasse by *Aspergillus niger* and *Trichoderma reesei*. *BMC Genomics*, *18*. https://doi.org/10.1186/s12864-017-3857-5

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*, 248–254.

Brakhage, A. A., Andrianopoulos, A., Kato, M., Steidl, S., Davis, M. A., Tsukagoshi, N., et al. (1999). HAP-like CCAAT-binding complexes in filamentous fungi: Implications for biotechnology. *Fungal Genetics and Biology*, *27*, 243–252. https://doi.org/10.1006/fgbi.1999.1136

Cai, P., Gu, R., Wang, B., Li, J., Wan, L., Tian, C., et al. (2014). Evidence of a critical role for cellodextrin transporter 2 (CDT-2) in both cellulose and hemicellulose degradation and utilization in *Neurospora crassa*. *PLoS One*, *9*, e89330. https://doi.org/10.1371/journal.pone.0089330

Campos Antoniêto, A. C., Graciano de Paula, R., Santos Castro, L. dos, Silva-Rocha, R., Felix Persinoti, G., & Nascimento Silva, R. (2015). *Trichoderma reesei* CRE1-mediated carbon catabolite repression in response to sophorose through RNA sequencing analysis. *Current Genomics*, *17*, 119–131. https://doi.org/10.2174/1389202917666151116212901

Chen, H. (2014). *Biotechnology of lignocellulose*. https://doi.org/10.1007/978-94-007-6898-7

Chen, H., & Chen, H. (2014). Chemical composition and structure of natural lignocellulose. In *Biotechnology of lignocellulose: Theory and practice* (pp. 25–71). Springer Netherlands. https://doi.org/10.1007/978-94-007-6898-7_2

Chen, X., Luo, Y., Yu, H., Sun, Y., Wu, H., Song, S., et al. (2014). Transcriptional profiling of biomass degradation-related genes during *Trichoderma reesei* growth on different carbon sources. *Journal of Biotechnology*, *173*, 59–64. https://doi.org/10.1016/j.jbiotec.2014.01.011

Chen, Y., Wu, C., Fan, X., Zhao, X., Zhao, X., Shen, T., et al. (2020). Engineering of *Trichoderma reesei* for enhanced degradation of lignocellulosic biomass by truncation of the cellulase activator ACE3. *Biotechnology for Biofuels*, *13*, 62. https://doi.org/10.1186/s13068-020-01701-3

Cherry, J. R., Fidantsef, A. L., Bischof, R. H., Ramoni, J., Seiboth, B., Cherry, J. R., et al. (2003). Directed evolution of industrial enzymes: An update. *Current Opinion in Biotechnology*, *14*, 438–443. https://doi.org/10.1016/S0958-1669(03)00099-5

Chikhi, R., & Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, *30*, 31–37. https://doi.org/10.1093/bioinformatics/btt310

Collier, L. A., Ghosh, A., & Borkovich, K. A. (2020). Heterotrimeric G-protein signaling is required for cellulose degradation in *Neurospora crassa*. *MBio*, *11*, e02419-20. https://doi.org/10.1128/MBIO.02419-20

Coradetti, S. T., Craig, J. P., Xiong, Y., Shock, T., Tian, C., & Glass, N. L. (2012a). Conserved and essential transcription factors for cellulase gene expression in ascomycete

fungi. *Proceedings of the National Academy of Sciences*, *109*, 7397–7402. https://doi.org/10.1073/pnas.1200785109

Coradetti, S. T., Craig, J. P., Xiong, Y., Shock, T., Tian, C., & Glass, N. L. (2012b). Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *Proceedings of the National Academy of Sciences*, *109*, 7397–7402. https://doi.org/10.1073/pnas.1200785109

Coradetti, S. T., Xiong, Y., & Glass, N. L. (2013). Analysis of a conserved cellulase transcriptional regulator reveals inducer-independent production of cellulolytic enzymes in *Neurospora crassa*. *MicrobiologyOpen*, *2*, 595–609. https://doi.org/10.1002/mbo3.94

Craig, J. P., Coradetti, S. T., Starr, T. L., & Glass, N. L. (2015). Direct target network of the *Neurospora crassa* plant cell wall deconstruction regulators clr-1, clr-2, and xlr-1. *MBio*, *6*. https://doi.org/10.1128/MBIO.01452-15

Dahmen, N., Lewandowski, I., Zibek, S., & Weidtmann, A. (2019). Integrated lignocellulosic value chains in a growing bioeconomy: Status quo and perspectives. *GCB Bioenergy*, *11*, 107–117. https://doi.org/10.1111/gcbb.12586

De Paula, R. G., Antoniêto, A. C. C., Carraro, C. B., Lopes, D. C. B., Persinoti, G. F., Peres, N. T. A., et al. (2018). The duality of the MAPK signaling pathway in the control of metabolic processes and cellulase production in *Trichoderma reesei*. *Scientific Reports*, *8*, 1–16. https://doi.org/10.1038/s41598-018-33383-1

Derntl, C., MacH, R. L., & MacH-Aigner, A. R. (2019). Fusion transcription factors for strong, constitutive expression of cellulases and xylanases in *Trichoderma reesei*. *Biotechnology for Biofuels*, *12*, 231. https://doi.org/10.1186/s13068-019-1575-8

Derntl, C., Rassinger, A., Srebotnik, E., Mach, R. L., & Mach-Aigner, A. R. (2015). Xpp1 regulates the expression of xylanases, but not of cellulases in *Trichoderma reesei*. *Biotechnology for Biofuels*, *8*, 1–11. https://doi.org/10.1186/S13068-015-0298-8

Druzhinina, I. S., & Kubicek, C. P. (2017). Genetic engineering of *Trichoderma reesei* cellulases and their production. *Microbial Biotechnology*, *10*, 1485–1499. https://doi.org/10.1111/1751-7915.12726

Ellilä, S., Fonseca, L., Uchima, C., Cota, J., Goldman, G. H., Saloheimo, M., et al. (2017). Development of a low-cost cellulase production process using *Trichoderma reesei* for Brazilian biorefineries. *Biotechnology for Biofuels*, *10*, 1–17.

Fernández, R. G., & Novo, J. V. J. (2013). Proteomic protocols for the study of filamentous fungi. In *Laboratory Protocols in Fungal Biology* (pp. 299–308). Springer New York. https://doi.org/10.1007/978-1-4614-2356-0_24

Ferreira Filho, J. A., Horta, M. A. C., Beloti, L. L., dos Santos, C. A., & de Souza, A. P. (2017). Carbohydrate-active enzymes in *Trichoderma harzianum*: A bioinformatic

analysis bioprospecting for key enzymes for the biofuels industry. *BMC Genomics*, *18*. https://doi.org/10.1186/s12864-017-4181-9

Foreman, P. K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N. S., et al. (2003). Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *Journal of Biological Chemistry*, *278*, 31988–31997. https://doi.org/10.1074/jbc.M304750200

Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. O., Jaffe, D., et al. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*, *422*, 859–868. https://doi.org/10.1038/nature01554

Gao, D., Uppugundla, N., Chundawat, S. P. S., Yu, X., Hermanson, S., Gowda, K., et al. (2011). Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. *Biotechnology for Biofuels*, *4*. https://doi.org/10.1186/1754-6834-4-5

Ghose, T. K. (1987). Measurement of cellulase activities. *Pure and Applied Chemistry*, *59*, 257. https://doi.org/10.1351/pac198759020257

Glass, N. L., Schmoll, M., Cate, J. H. D., & Coradetti, S. (2013). Plant cell wall deconstruction by ascomycete fungi. *Annual Review of Microbiology*, *67*, 477–498. https://doi.org/10.1146/annurev-micro-092611-150044

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644–652. https://doi.org/10.1038/nbt.1883

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, *27*, 1017–1018. https://doi.org/10.1093/bioinformatics/btr064

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., et al. (2014). MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research*, *42*. https://doi.org/10.1093/nar/gkt1183

Guo, X., An, Y., Liu, F., Lu, F., & Wang, B. (2022). Lytic polysaccharide monooxygenase—A new driving force for lignocellulosic biomass degradation. *Bioresource Technology*, *362*, 127803. https://doi.org/10.1016/j.biortech.2022.127803

Gupta, A., & Verma, J. P. (2015). Sustainable bio-ethanol production from agro-residues: A review. *Renewable and Sustainable Energy Reviews*, *41*, 550–567. https://doi.org/10.1016/j.rser.2014.08.032

Gyalai-Korpos, M., Nagy, G., Mareczky, Z., Schuster, A., Réczey, K., & Schmoll, M. (2010). Relevance of the light signaling machinery for cellulase expression in *Trichoderma reesei* (Hypocrea jecorina). *BMC Research Notes*, *3*, 330. https://doi.org/10.1186/1756-0500-3-330

Häkkinen, M., Sivasiddarthan, D., Aro, N., Saloheimo, M., & Pakula, T. M. (2015). The effects of extracellular pH and of the transcriptional regulator PACI on the transcrip-

tome of *Trichoderma reesei*. *Microbial Cell Factories*, *14*, 1–15. https://doi.org/10.1186/S12934-015-0247

Hall, M., Bansal, P., Lee, J. H., Realff, M. J., & Bommarius, A. S. (2010). Cellulose crystallinity - A key predictor of the enzymatic hydrolysis rate. *FEBS Journal*, *277*, 1571–1582. https://doi.org/10.1111/j.1742-4658.2010.07585.x

Havukainen, S., Valkonen, M., Koivuranta, K., & Landowski, C. P. (2020). Studies on sugar transporter CRT1 reveal new characteristics that are critical for cellulase induction in *Trichoderma reesei*. *Biotechnology for Biofuels*, *13*, 1–20. https://doi.org/10.1186/S13068-020-01797-7

Herpoël-Gimbert, I., Margeot, A., Dolla, A., Jan, G., Mollé, D., Lignon, S., et al. (2008). Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. *Biotechnology for Biofuels*, *1*, 1–12. https://doi.org/10.1186/1754-6834-1-18

Horn, S. J., Vaaje-Kolstad, G., Westereng, B., & Eijsink, V. G. H. (2012). Novel enzymes for the degradation of cellulose. *Biotechnology for Biofuels*, *5*. https://doi.org/10.1186/1754-6834-5-45

Hu, Y., Liu, Y., Hao, X., Wang, D., Akhberdi, O., Xiang, B., et al. (2018). Regulation of the Gα-cAMP/PKA signaling pathway in cellulose utilization of *Chaetomium globosum*. *Microbial Cell Factories*, *17*, 160. https://doi.org/10.1186/S12934-018-1008-6

Huang, T., Niu, S., Zhang, F., Wang, B., Wang, J., Liu, G., et al. (2024). Correlating gene expression levels with transcription factor binding sites facilitates identification of key transcription factors from transcriptome data. *Frontiers in Genetics*, *15*, 1511456. https://doi.org/10.3389/fgene.2024.1511456

Ilmén, M., Saloheimo, A., Onnela, M. L., & Penttilä, M. E. (1997). Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. *Applied and Environmental Microbiology*, *63*, 1298–1306. https://doi.org/10.1128/aem.63.4.1298-1306.1997

Ilmén, M., Thrane, C., & Penttilä, M. (1996). The glucose repressor gene cre1 of *Trichoderma*: Isolation and expression of a full-length and a truncated mutant form. *Molecular and General Genetics*, *251*, 451–460. https://doi.org/10.1007/bf02172374

Ivanova, C., Bååth, J. A., Seiboth, B., & Kubicek, C. P. (2013). Systems analysis of lactose metabolism in *Trichoderma reesei* identifies a lactose permease that is essential for cellulase induction. *PLoS One*, *8*, 1–10. https://doi.org/10.1371/journal.pone.0062631

Janusz, G., Pawlik, A., Świderska-Burek, U., Polak, J., Sulej, J., Jarosz-Wilkołazka, A., et al. (2020). Laccase properties, physiological functions, and evolution. *International Journal of Molecular Sciences*, *21*. https://doi.org/10.3390/ijms21030966

Jayasekara, S., & Ratnayake, R. (2019). Microbial cellulases: An overview and applications. *Cellulose*. https://doi.org/10.5772/intechopen.84531

Johnson, E. (2016). Integrated enzyme production lowers the cost of cellulosic ethanol. *Biofuels, Bioproducts and Biorefining*, *10*, 164–174. https://doi.org/10.1002/bbb.1634

Kamagata, Y., Yachi, M., Kurasawa, T., Suto, M., Sasaki, H., Takao, S., et al. (1991). Cellulase induction by cellobiose-octaacetate in *Penicillium purpurogenum*. *Journal of Fermentation and Bioengineering*, *72*, 217–220. https://doi.org/10.1016/0922-338X(91)90221-2

Klaubauf, S., Narang, H. M., Post, H., Zhou, M., Brunner, K., Mach-Aigner, A. R., et al. (2014). Similar is not the same: Differences in the function of the (hemi-)cellulolytic regulator XlnR (Xlr1/Xyr1) in filamentous fungi. *Fungal Genetics and Biology*, *72*, 73–81. https://doi.org/10.1016/j.fgb.2014.07.007

Klein, M., Swinnen, S., Thevelein, J. M., & Nevoigt, E. (2017). Glycerol metabolism and transport in yeast and fungi: established knowledge and ambiguities. *Environmental Microbiology*, *19*, 878–893. https://doi.org/10.1111/1462-2920.13617

Kubicek, C. P., Messner, R., Gruber, F., Mandels, M., & Kubicek-Pranz, E. M. (1993). Triggering of cellulase biosynthesis by cellulose in *Trichoderma reesei*: Involvement of a constitutive, sophorose-inducible, glucose-inhibited β-diglucoside permease. *Journal of Biological Chemistry*, *268*, 19364–19368. https://doi.org/10.1016/s0021-9258(19)36523-8

Kubicek, C. P., Mikus, M., Schuster, A., Schmoll, M., & Seiboth, B. (2009). Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. *Biotechnology for Biofuels*, *2*. https://doi.org/10.1186/1754-6834-2-19

Kuhad, R. C., Gupta, R., & Singh, A. (2011). Microbial cellulases and their industrial applications. *Enzyme Research*, *2011*. https://doi.org/10.4061/2011/280696

Kumar, R., Singh, S., & Singh, O. V. (2008). Bioconversion of lignocellulosic biomass: Biochemical and molecular perspectives. *Journal of Industrial Microbiology & Biotechnology*, *35*, 377–391. https://doi.org/10.1007/s10295-008-0327-8

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*, 680–685. https://doi.org/10.1038/227680a0

Lairson, L. L., Henrissat, B., Davies, G. J., & Withers, S. G. (2008). Glycosyl transferases: Structures, functions, and mechanisms. *Annual Review of Biochemistry*, *77*, 521–555. https://doi.org/10.1146/annurev.biochem.76.061005.092322

Leandro, M. J., Fonseca, C., & Gonçalves, P. (2009). Hexose and pentose transport in ascomycetous yeasts: An overview. *FEMS Yeast Research*, *9*, 511–525. https://doi.org/10.1111/j.1567-1364.2009.00509.x

Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., & Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*, *6*. https://doi.org/10.1186/1754-6834-6-41

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 1–16. https://doi.org/10.1186/1471-2105-12-323

Li, C., Pang, A. P., Yang, H., Lv, R., Zhou, Z., Wu, F. G., et al. (2019). Tracking localization and secretion of cellulase spatiotemporally and directly in living *Trichoderma reesei*. *Biotechnology for Biofuels*, *12*, 200. https://doi.org/10.1186/s13068-019-1538-0

Li, J., Liu, G., Chen, M., Li, Z., Qin, Y., & Qu, Y. (2013). Cellodextrin transporters play important roles in cellulase induction in the cellulolytic fungus *Penicillium oxalicum*. *Applied Microbiology and Biotechnology*, *97*, 10479–10488. https://doi.org/10.1007/s00253-013-5301-3

Li, J., Liu, Q., Li, J., Lin, L., Li, X., Zhang, Y., et al. (2021). RCO-3 and COL-26 form an external-to-internal module that regulates the dual-affinity glucose transport system in *Neurospora crassa*. *Biotechnology for Biofuels*, *14*, 33. https://doi.org/10.1186/s13068-021-01877-2

Li, J., Wiebenga, A., Lipzen, A., Ng, V., Tejomurthula, S., Zhang, Y., et al. (2023a). Comparative genomics and transcriptomics analyses reveal divergent plant biomass-degrading strategies in fungi. *Journal of Fungi*, *9*. https://doi.org/10.3390/jof9080860

Li, N., Kunitake, E., Aoyama, M., Ogawa, M., Kanamaru, K., Kimura, M., et al. (2016). McmA-dependent and -independent regulatory systems governing expression of ClrB-regulated cellulase and hemicellulase genes in *Aspergillus nidulans*. *Molecular Microbiology*, *102*, 810–826. https://doi.org/10.1111/mmi.13493

Li, N., Qiu, Z., Cai, W., Shen, Y., Wei, D., Chen, Y., et al. (2023b). The Ras small GTPase RSR1 regulates cellulase production in *Trichoderma reesei*. *Biotechnology for Biofuels and Bioproducts*, *16*. https://doi.org/10.1186/s13068-023-02341-z

Li, W. C., Huang, C. H., Chen, C. L., Chuang, Y. C., Tung, S. Y., & Wang, T. F. (2017). *Trichoderma reesei* complete genome sequence, repeat-induced point mutation, and partitioning of CAZyme gene clusters. *Biotechnology for Biofuels*, *10*. https://doi.org/10.1186/s13068-017-0825-x

Liu, S., Quan, L., Yang, M., Wang, D., & Wang, Y. Z. (2024). Regulation of cellulase production via calcium signalling in *Trichoderma reesei* under PEG8000 stress. *Applied Microbiology and Biotechnology*, *108*, 178. https://doi.org/10.1007/s00253-023-12901-w

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*. https://doi.org/10.1093/nar/gkt1178

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*. https://doi.org/10.1186/s13059-014-0550-8

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*, 955–964. https://doi.org/10.1093/nar/25.5.955

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience*, *1*. https://doi.org/10.1186/2047-217X-1-18

Lynd, L. R., Weimer, P. J., Van Zyl, W. H., & Isak, S. (2002). Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, *66*, 506–577. https://doi.org/10.1128/MMBR.66.3.506

Mach-Aigner, A. R., Pucher, M. E., Steiger, M. G., Bauer, G. E., Preis, S. J., & Mach, R. L. (2008). Transcriptional regulation of xyr1, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. *Applied and Environmental Microbiology*, *74*, 6554–6562. https://doi.org/10.1128/AEM.01143-08

Mach, R. L., & Zeilinger, S. (2003). Regulation of gene expression in industrial fungi: *Trichoderma*. *Applied Microbiology and Biotechnology*, *60*, 515–522. https://doi.org/10.1007/s00253-002-1162-x

Magwaza, B., Amobonye, A., & Pillai, S. (2024). Microbial β-glucosidases: Recent advances and applications. *Biochimie*, *225*, 49–67. https://doi.org/10.1016/j.biochi.2024.05.009

Mäkelä, M. R., Donofrio, N., & De Vries, R. P. (2014). Plant biomass degradation by fungi. *Fungal Genetics and Biology*, *72*, 2–9. https://doi.org/10.1016/j.fgb.2014.08.010

Malherbe, S., & Cloete, T. E. (2002). Lignocellulose biodegradation: Fundamentals and applications. *Reviews in Environmental Science and Biotechnology*, *1*, 105–114. https://doi.org/10.1023/A:1020858910646

Mandels, M., & Weber, J. (1969). The production of cellulases. In *Cellulases and their applications* (pp. 23–391). American Chemical Society. https://doi.org/10.1021/ba-1969-0095.ch023

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*, 10. https://doi.org/10.14806/ej.17.1.200

Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., et al. (2008). Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnology*, *26*, 553–560. https://doi.org/10.1038/nbt1403

Mattam, A. J., Chaudhari, Y. B., & Velankar, H. R. (2022a). Factors regulating cellulolytic gene expression in filamentous fungi: An overview. *Microbial Cell Factories*, *21*, 1–19. https://doi.org/10.1186/s12934-022-01764-x

Messner, R., & Kubicek, C. P. (1991). Carbon source control of cellobiohydrolase I and II formation by *Trichoderma reesei*. *Applied and Environmental Microbiology*, *57*, 630–635. https://doi.org/10.1128/aem.57.3.630-635.1991

Miura, N., & Ueda, M. (2018). Evaluation of unconventional protein secretion by *Saccharomyces cerevisiae* and other fungi. *Cells*, *7*, 128. https://doi.org/10.3390/cells7090128

Montenecourt, B. S., & Eveleigh, D. E. (1979). Selective screening methods for the isolation of high yielding cellulase mutants of *Trichoderma reesei*. In H. S. Gautier & P. Huisman (Eds.), *Hydrolysis of cellulose: Mechanisms of enzymatic and acid hydrolysis* (Advances in Chemistry, No. 181, pp. 289–301). American Chemical Society. https://doi.org/10.1021/ba-1979-0181.ch014

Nitta, M., Furukawa, T., Shida, Y., Mori, K., Kuhara, S., Morikawa, Y., et al. (2012). A new Zn(II) 2Cys6-type transcription factor BglR regulates β-glucosidase expression in *Trichoderma reesei*. *Fungal Genetics and Biology*, *49*, 388–397. https://doi.org/10.1016/j.fgb.2012.02.009

Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., et al. (2015). Fungal cellulases. *Chemical Reviews*, *115*, 1308–1448. https://doi.org/10.1021/cr500351c

Percival Zhang, Y. H., Himmel, M. E., & Mielenz, J. R. (2006). Outlook for cellulase improvement: Screening and selection strategies. *Biotechnology Advances*, *24*, 452–481. https://doi.org/10.1016/j.biotechadv.2006.03.003

Peterson, R., & Nevalainen, H. (2012). *Trichoderma reesei* RUT-C30 - Thirty years of strain improvement. *Microbiology*, *158*, 58–68. https://doi.org/10.1099/mic.0.054031-0

Portnoy, T., Margeot, A., Seidl-Seiboth, V., Le Crom, S., Ben Chaabane, F., Linke, R., et al. (2011). Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in *Trichoderma reesei* strains producing high and low levels of cellulase. *Eukaryotic Cell*, *10*, 262–271. https://doi.org/10.1128/EC.00208-10

Punt, P. J., Dingemanse, M. A., Kuyvenhoven, A., Soede, R. D. M., Pouwels, P. H., & van den Hondel, C. A. M. J. J. (1990). Functional elements in the promoter region of the *Aspergillus nidulans* gpdA gene encoding glyceraldehyde-3-phosphate dehydrogenase. *Gene*, *93*, 101–109. https://doi.org/10.1016/0378-1119(90)90142-E

Rajasree, K. P., Mathew, G. M., Pandey, A., & Sukumaran, R. K. (2013). Highly glucose tolerant β-glucosidase from *Aspergillus unguis*: NII 08123 for enhanced hydrolysis of biomass. *Journal of Industrial Microbiology and Biotechnology*, *40*, 967–975. https://doi.org/10.1007/s10295-013-1291-5

Raulo, R., Kokolski, M., & Archer, D. B. (2016). The roles of the zinc finger transcription factors XlnR, ClrA and ClrB in the breakdown of lignocellulose by *Aspergillus niger*. *AMB Express*, *6*, 1–12. https://doi.org/10.1186/s13568-016-0177-0

Rauscher, R., Würleitner, E., Wacenovsky, C., Aro, N., Stricker, A. R., Zeilinger, S., et al. (2006). Transcriptional regulation of xyn1, encoding xylanase I, in *Hypocrea jecorina*. *Eukaryotic Cell*, *5*, 447–456. https://doi.org/10.1128/EC.5.3.447-456.2006

Reese, E. T., & Mandels, M. (1984). Rolling with the times: Production and applications of *Trichoderma reesei* cellulase in Third Annual Marvin Johnson Memorial Lecture Presented by Mary Mandels in Kansas City, Missouri 15 September 1982. In *Enzymes and microbial technology* (pp. 1–20). Elsevier. https://doi.org/10.1016/b978-0-12-040307-3.50006-8

Ries, L., Pullan, S. T., Delmas, S., Malla, S., Blythe, M. J., & Archer, D. B. (2013). Genome-wide transcriptional response of *Trichoderma reesei* to lignocellulose using RNA sequencing and comparison with *Aspergillus niger*. *BMC Genomics*, *14*. https://doi.org/10.1186/1471-2164-14-541

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Ryu, D. D. Y., & Mandels, M. (1980). Cellulases: Biosynthesis and applications. *Enzyme and Microbial Technology*, *2*, 91–102. https://doi.org/10.1016/0141-0229(80)90063-0

Saini, J. K., Saini, R., & Tewari, L. (2015). Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: Concepts and recent developments. *3 Biotech*, *5*, 337–353. https://doi.org/10.1007/s13205-014-0246-5

Saloheimo, M., Kuja-Panula, J., Ylösmäki, E., Ward, M., & Penttilä, M. (2002). Enzymatic properties and intracellular localization of the novel *Trichoderma reesei* β-glucosidase BGLII (Cel1A). *Applied and Environmental Microbiology*, *68*, 4546–4553. https://doi.org/10.1128/AEM.68.9.4546-4553.2002

Saloheimo, M., & Pakula, T. M. (2012a). The cargo and the transport system: Secreted proteins and protein secretion in *Trichoderma reesei* (Hypocrea jecorina). *Microbiology*, *158*, 46–57. https://doi.org/10.1099/mic.0.053132-0

Saloheimo, M., & Pakula, T. M. (2012b). The cargo and the transport system: Secreted proteins and protein secretion in *Trichoderma reesei* (Hypocrea jecorina). *Microbiology*, *158*, 46–57. https://doi.org/10.1099/mic.0.053132-0

Schalamun, M., Beier, S., Hinterdobler, W., Wanko, N., Schinnerl, J., Brecker, L., et al. (2023). MAPkinases regulate secondary metabolism, sexual development and light dependent cellulase regulation in *Trichoderma reesei*. *Scientific Reports*, *13*, 1–17. https://doi.org/10.1038/s41598-023-28938-w

Schmoll, M. (2018). Regulation of plant cell wall degradation by light in *Trichoderma*. *Fungal Biology and Biotechnology*, *5*, 1–20. https://doi.org/10.1186/s40694-018-0052-7

Schmoll, M., & Hinterdobler, W. (2022). Tools for adapting to a complex habitat: G-protein coupled receptors in *Trichoderma*. *Progress in Molecular Biology and Translational Science*, *193*, 65–97. https://doi.org/10.1016/bs.pmbts.2022.06.003

Schmoll, M., & Kubicek, C. P. (2003). Regulation of *Trichoderma* cellulase formation: Lessons in molecular biology from an industrial fungus. A review. *Acta Microbiologica*

et Immunologica Hungarica, *50*, 125–145. https://doi.org/10.1556/amicr.50.2003.2-3.3

Schülein, M. (1988). Cellulases of *Trichoderma reesei*. In *Methods in Enzymology* (Vol. 160, pp. 234–245). https://doi.org/10.1016/0076-6879(88)60125-x

Seiboth, B., Karimi, R. A., Phatale, P. A., Linke, R., Hartl, L., Sauer, D. G., et al. (2012). The putative protein methyltransferase LAE1 controls cellulase gene expression in *Trichoderma reesei*. *Molecular Microbiology*, *84*, 1150. https://doi.org/10.1111/j.1365-2958.2012.08083.x

Seidl, V., Gamauf, C., Druzhinina, I. S., Seiboth, B., Hartl, L., & Kubicek, C. P. (2008). The *Hypocrea jecorina* (*Trichoderma reesei*) hypercellulolytic mutant RUT C30 lacks a 85 kb (29 gene-encoding) region of the wild-type genome. *BMC Genomics*, *9*. https://doi.org/10.1186/1471-2164-9-327

Shallom, D., & Shoham, Y. (2003). Microbial hemicellulases. *Current Opinion in Microbiology*, *6*, 219–228. https://doi.org/10.1016/S1369-5274(03)00056-0

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Singhania, R. R., Saini, J. K., Saini, R., Adsul, M., Mathur, A., Gupta, R., et al. (2014). Bioethanol production from wheat straw via enzymatic route employing *Penicillium janthinellum* cellulases. *Bioresource Technology*, *169*, 490–495. https://doi.org/10.1016/j.biortech.2014.07.011

Singhvi, M. S., Adsul, M. G., & Gokhale, D. V. (2011). Comparative production of cellulases by mutants of *Penicillium janthinellum* NCIM 1171 and its application in hydrolysis of Avicel and cellulose. *Bioresource Technology*, *102*, 6569–6572. https://doi.org/10.1016/j.biortech.2011.01.014

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*, 637–644. https://doi.org/10.1093/bioinformatics/btn013

Stappler, E., Walton, J. D., Beier, S., & Schmoll, M. (2017). Abundance of secreted proteins of *Trichoderma reesei* is regulated by light of different intensities. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.02586

Stricker, A. R., Grosstessner-Hain, K., Würleitner, E., & Mach, R. L. (2006). Xyr1 (Xylanase Regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. *Eukaryotic Cell*, *5*, 2128–2137. https://doi.org/10.1128/ec.00211-06

Stricker, A. R., Mach, R. L., & De Graaff, L. H. (2008). Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecori-*

na (*Trichoderma reesei*). *Applied Microbiology and Biotechnology*, *78*, 211–220. https://doi.org/10.1007/s00253-007-1322-0

Su, X., Schmitz, G., Zhang, M., Mackie, R. I., & Cann, I. K. O. (2012). Heterologous gene expression in filamentous fungi. Elsevier. https://doi.org/10.1016/b978-0-12-394382-8.00001-0

Sukumaran, R. K., Singhania, R. R., & Pandey, A. (2005). Microbial cellulases—Production, applications and challenges. *Applied Microbiology and Biotechnology*.

Suto, M., & Tomita, F. (2001). Induction and catabolite repression mechanisms of cellulase in fungi. *Journal of Bioscience and Bioengineering*, *92*, 305–311. https://doi.org/10.1016/S1389-1723(01)80231-0

Tisch, D., & Schmoll, M. (2010a). Light regulation of metabolic pathways in fungi. *Applied Microbiology and Biotechnology*, *85*, 1259–1277. https://doi.org/10.1007/s00253-009-2320-1

Tisch, D., & Schmoll, M. (2010b). Light regulation of metabolic pathways in fungi. *Applied Microbiology and Biotechnology*, *85*, 1259–1277. https://doi.org/10.1007/s00253-009-2320-1

Todd, R. B., Zhou, M., Ohm, R. A., Leeggangers, H. A. C. F., Visser, L., & de Vries, R. P. (2014). Prevalence of transcription factors in ascomycete and basidiomycete fungi. *BMC Genomics*, *15*. https://doi.org/10.1186/1471-2164-15-214

Trauger, S. A., Webb, W., & Siuzdak, G. (2002). Peptide and protein analysis with mass spectrometry. *Spectroscopy*, *16*, 15–28. https://doi.org/10.1155/2002/320152

Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sørlie, M., et al. (2010). An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science*, *330*, 219–222. https://doi.org/10.1126/science.1192231

Van Dyk, J. S., & Pletschke, B. I. (2012). A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes—Factors affecting enzymes, conversion and synergy. *Biotechnology Advances*, *30*, 1458–1480. https://doi.org/10.1016/j.biotechadv.2012.03.002

Van Peij, N. N. M. E., Visser, J., & De Graaff, L. H. (1998). Isolation and analysis of xlnR, encoding a transcriptional activator coordinating xylanolytic expression in *Aspergillus niger*. *Molecular Microbiology*, *27*, 131–142. https://doi.org/10.1046/j.1365-2958.1998.00666.x

Wang, B., Li, J., Gao, J., Cai, P., Han, X., & Tian, C. (2017). Identification and characterization of the glucose dual-affinity transport system in *Neurospora crassa*: Pleiotropic roles in nutrient transport, signaling, and carbon catabolite repression. *Biotechnology for Biofuels*, *10*, 1–22. https://doi.org/10.1186/S13068-017-0705-4

Wang, L., Lv, X., Cao, Y., Zheng, F., Meng, X., Shen, Y., et al. (2019). A novel transcriptional regulator RXE1 modulates the essential transactivator XYR1 and cellulase gene

expression in *Trichoderma reesei*. *Applied Microbiology and Biotechnology*, *103*, 4511–4523. https://doi.org/10.1007/S00253-019-09739-6

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*, 57–63. https://doi.org/10.1038/nrg2484

Wang, Z., Yang, R., Lv, W., Zhang, W., Meng, X., & Liu, W. (2022). Functional characterization of sugar transporter CRT1 reveals differential roles of its C-terminal region in sugar transport and cellulase induction in *Trichoderma reesei*. *Microbiology Spectrum*, *10*. https://doi.org/10.1128/spectrum.00872-22

Xiong, Y., Sun, J., & Glass, N. L. (2014). VIB1, a link between glucose signaling and carbon catabolite repression, is essential for plant cell wall degradation by *Neurospora crassa*. *PLoS Genetics*, *10*. https://doi.org/10.1371/journal.pgen.1004500

Xu, L., Li, J., Gonzalez Ramos, V. M., Lyra, C., Wiebenga, A., Grigoriev, I. V., et al. (2024). Genome-wide prediction and transcriptome analysis of sugar transporters in four ascomycete fungi. *Bioresource Technology*, *391*, 130006. https://doi.org/10.1016/j.biortech.2023.130006

Yan, S., Xu, Y., & Yu, X. W. (2021a). From induction to secretion: A complicated route for cellulase production in *Trichoderma reesei*. *Bioresources and Bioprocessing*, *8*. https://doi.org/10.1186/s40643-021-00461-8

Yan, S., Xu, Y., & Yu, X. W. (2021b). From induction to secretion: A complicated route for cellulase production in *Trichoderma reesei*. *Bioresources and Bioprocessing*, *8*, 1–15. https://doi.org/10.1186/S40643-021-00461-8

Yennamalli, R. M., Rader, A. J., Kenny, A. J., Wolt, J. D., & Sen, T. Z. (2013). Endoglucanases: Insights into thermostability for biofuel applications. *Biotechnology for Biofuels*. https://doi.org/10.1186/1754-6834-6-136

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). DbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, *40*. https://doi.org/10.1093/nar/gks479

Zeilinger, S., Ebner, A., Marosits, T., Mach, R., & Kubicek, C. P. (2001). The *Hypocrea jecorina* HAP 2/3/5 protein complex binds to the inverted CCAAT-box (ATTGG) within the cbh2 (cellobiohydrolase II-gene) activating element. *Molecular Genetics and Genomics*, *266*, 56–63. https://doi.org/10.1007/s004380100518

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*, 821–829. https://doi.org/10.1101/gr.074492.107

Zhang, J., Chen, Y., Wu, C., Liu, P., Wang, W., & Wei, D. (2019). The transcription factor ACE3 controls cellulase activities and lactose metabolism via two additional regulators in the fungus *Trichoderma reesei*. *Journal of Biological Chemistry*, *294*, 18435. https://doi.org/10.1074/jbc.ra119.008497

Zhang, J., Zhang, Y., Zhong, Y., Qu, Y., & Wang, T. (2012). Ras GTPases modulate morphogenesis, sporulation and cellulase gene expression in the cellulolytic fungus *Trichoderma reesei*. *PLoS One*, *7*, e48786. https://doi.org/10.1371/journal.pone.0048786

Zhang, W., Kou, Y., Xu, J., Cao, Y., Zhao, G., Shao, J., et al. (2013). Two major facilitator superfamily sugar transporters from *Trichoderma reesei* and their roles in induction of cellulase biosynthesis. *Journal of Biological Chemistry*, *288*, 32861–32872. https://doi.org/10.1074/jbc.m113.505826

Zhang, X.-Z., & Zhang, Y.-H. P. (2013). Cellulases: Characteristics, sources, production, and applications. In *Bioprocessing Technologies in Biorefinery for Sustainable Production of Fuels, Chemicals, and Polymers* (pp. 131–146). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118642047.ch8

Zhang, Y. H. P., & Lynd, L. R. (2004). Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnology and Bioengineering*, *88*, 797–824. https://doi.org/10.1002/bit.20282

Zhao, K., Liu, Z., Li, M., Hu, Y., Yang, L., Song, X., et al. (2022). Drafting *Penicillium oxalicum* calcineurin-CrzA pathway by combining the analysis of phenotype, transcriptome, and endogenous protein–protein interactions. *Fungal Genetics and Biology*, *158*, 103652. https://doi.org/10.1016/j.fgb.2021.103652

Zhao, S., Zhang, T., Hasunuma, T., Kondo, A., Zhao, X. Q., & Feng, J. X. (2024). Every road leads to Rome: Diverse biosynthetic regulation of plant cell wall-degrading enzymes in filamentous fungi *Penicillium oxalicum* and *Trichoderma reesei*. *Critical Reviews in Biotechnology*, *44*(7), 1241–1261. https://doi.org/10.1080/07388551.2023.2280810

Zhou, Q., Xu, J., Kou, Y., Lv, X., Zhang, X., Zhao, G., Zhang, W., Chen, G., & Liu, W. (2012). Differential involvement of β-glucosidases from *Hypocrea jecorina* in rapid induction of cellulase genes by cellulose and cellobiose. *Eukaryotic Cell*, *11*(11), 1371–1381. https://doi.org/10.1128/ec.00170-12

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, *29*, 2669–2677. https://doi.org/10.1093/bioinformatics/btt476

Annexure I: Secretome profile of *Trichoderma reesei* RUT-C30

Accession	Pep-tide count	Unique pep-tides	Max fold change	Highest mean condition	Lowest mean condition	Mass	Description	Normal-ized Abun-dance Tr-Cel 01_2020 0102_TR C_01	Normalized Abundance Tr-Cel 02_202001 02_TRC_02	Normal mal-ized Abun-dance Tr-Glu 01_202 00102_ TRG_0 1	Normal-ized Abun-dance Tr-Glu 02_202 00102_ TRG_02
A0A024SFG8;A7J2C 6;G0RHJ4	8	1	2273	TRC	TRG	39374	Acetyl esterase OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_136770 PE=4 SV=1	26131.45	23291.35	11.60	10.14
A0A024SDN0;A0A1 D9CUY1;A0A2H2ZVZ 6;G0RFZ5;P34825	1	1	12583	TRC	TRG	50159	Elongation factor 1-alpha OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_122704 PE=3 SV=1	1989.46	1631.73	0.16	0.13

P62694;A0A024RXP8;A0A2H2ZK41;G0RVK1;G9NTY1;P13860	27	27	419	TRC	TRG	55442	Exoglucanase 1 OS=Hypocrea jecorina OX=51453 GN=cbh1 PE=1 SV=1	416914.22	359272.79	948.06	903.11
Q99034;A0A024SJR7;G0R6T6;A0A024S0D5;A0A2H2ZRV5;G0R6X2	4	4	2522	TRC	TRG	31552	Acetylxylan esterase OS=Hypocrea jecorina OX=51453 GN=axe1 PE=1 SV=1	54120.18	44535.47	20.53	18.59
A0A1L7H884;A0A1L7H894;A0A2H2Z5N2;B2CZF9;G0RUP7;P36217;P48793	1	1	Infinity	TRC	TRG	20888	Endo-1_4-beta-xylanase (Fragment) OS=Hypocrea jecorina OX=51453 GN=Xyn2 PE=2 SV=1	5147.53	3913.26	0.00	0.00
A0A2H3A3Q5;G0R947;P36218	2	2	299	TRC	TRG	24544	Endo-1_4-beta-xylanase OS=Trichoderma parareesei OX=858221 GN=A9Z42_0081650 PE=3 SV=1	35898.05	29921.06	106.95	113.11
A0A024SJK4;A0A2H3A6A8;G0R6T7;Q7Z9M9	2	2	35	TRC	TRG	33389	CBM1 domain-containing protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_121449 PE=4 SV=1	3458.27	3864.06	100.30	106.70

A0A024RZP7;A0A2H2ZLZ4;G0RVK3;Q9P8D0	1	1	26748	TRC	TRG	53121	Swollenin OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_104220 PE=4 SV=1	21546.28	18952.88	0.98	0.54
A0A024SNB7;G0RKH9;P07981;A0A2H2Z4J6;Q5BMS5;E1AFV7	11	11	184310	TRC	TRG	49463	Endoglucanase EG-1 OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=egl1 PE=1 SV=1	79837.25	74435.38	0.65	0.18
G0RG23;A0A024SDM6;A0A2H2Z0L4;Q92458	31	31	13279	TRC	TRG	87533	Glycoside hydrolase family 3 OS=Hypocrea jecorina (strain QM6a) OX=431241 GN=TRIREDRAFT_121127 PE=4 SV=1	41017.80	34413.55	3.36	2.32
A0A2H2ZS58	2	2	47	TRC	TRG	70280	Uncharacterized protein OS=Trichoderma parareesei OX=858221 GN=A9Z42_0001730 PE=4 SV=1	71040.28	61252.30	1360.58	1473.29

Q99036;G0RC85	3	2	108	TRC	TRG	47851	Mannan endo-1_4-beta-mannosidase A OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=man1 PE=1 SV=1	1530.83	1243.49	12.28	13.51
A0A024RXP9;A0A2H2ZPA1;G0RXH1	5	4	405	TRC	TRG	57296	1_3-beta-glucanosyltransferase OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_103899 PE=3 SV=1	3397.89	3052.28	6.82	9.09
A0A024RV01;G0RV92;A0A2H2YV98	5	5	88	TRC	TRG	48747	Cellulosome enzyme OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_93498 PE=4 SV=1	3084.95	3020.18	38.78	30.27

A0A024SM10;A0A2H2ZWJ0;G0R6T8;O14405	3	3	987	TRC	TRG	35967	Endoglucanase-4 OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_139633 PE=4 SV=1	3142.39	2492.81	3.35	2.36
A0A2H2ZCU1;A0A024RVF3;G0RX84	6	6	391	TRC	TRG	19012	CFEM domain-containing protein OS=Trichoderma parareesei OX=858221 GN=A9Z42_0055850 PE=4 SV=1	62998.86	54258.53	120.33	179.25
P07982;A0A024SH20;G0RB67;Q2F8H3;A0A2H2ZP44	8	8	50334	TRC	TRG	44911	Endoglucanase EG-II OS=Hypocrea jecorina OX=51453 GN=egl2 PE=1 SV=1	95621.83	98188.53	1.10	2.75
A0A024S437;A0A2H3A4B7;G0RQL9	1	1	53	TRC	TRG	22889	CFEM domain-containing protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_25723 PE=4 SV=1	2003.73	1835.84	42.33	29.81

A0A024SBP4;A0A2H2ZPV6;G0RSZ4	6	6	2259	TRC	TRG	24446	Uncharacterized protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_111096 PE=4 SV=1	14269.18	17357.31	4.49	9.51
A0A024SE00	1	1	1083	TRC	TRG	38456	Uncharacterized protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_7543 PE=4 SV=1	3063.67	2724.69	1.50	3.85
G0RN29	1	1	104	TRC	TRG	73423	Predicted protein OS=Hypocrea jecorina (strain QM6a) OX=431241 GN=TRIREDRAFT_108674 PE=4 SV=1	910.98	811.62	10.63	5.94
G0RGX3;A0A024SCI8;A0A2H2ZX60	3	2	4	TRC	TRG	148621	Predicted protein OS=Hypocrea jecorina (strain QM6a) OX=431241 GN=TRIREDRAFT_106626 PE=4 SV=1	4529.97	4130.63	1094.17	931.83

P07987;A0A024SH76;G0RB58;A0A2H3A139;D3YNY1;Q9HEY8;W8GHL8	21	21	18845	TRC	TRG	50338	Exoglucanase 2 OS=Hypocrea jecorina OX=51453 GN=cbh2 PE=1 SV=1	255359.20	257832.20	21.57	5.66
A0A2H2ZT16;A0A024SI56;G0RA78	4	4	1	TRG	TRC	32183	Methyltransferase OS=Trichoderma parareesei OX=858221 GN=A9Z42_0087730 PE=4 SV=1	2898.44	2972.96	4004.17	3871.88
A0A2H2ZU97;A0A024S2S5;G0RP19	3	3	2	TRC	TRG	217070	Uncharacterized protein OS=Trichoderma parareesei OX=858221 GN=A9Z42_0092140 PE=4 SV=1	634.93	569.38	273.01	292.10
G0RH70;A0A024SEL9;A0A2H2ZN37	4	4	2	TRC	TRG	30901	Uncharacterized protein OS=Hypocrea jecorina (strain QM6a) OX=431241 GN=TRIREDRAFT_121408 PE=4 SV=1	108446.61	109019.96	64968.68	58907.26

A0A024SMA7;A0A2H3A2Q7;G0RDL2	1	1	5373	TRC	TRG	29787	Uncharacterized protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_95954 PE=4 SV=1	6944.92	6770.97	2.55	0.00
A0A024S7E6;G0RQ16	1	1	3	TRC	TRG	29370	GH16 domain-containing protein (Fragment) OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_66752 PE=4 SV=1	2040.05	1989.80	859.79	715.36
A0A024SM86;A0A2H2ZLK2;G0RKR2	2	2	92	TRC	TRG	24445	Uncharacterized protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_121745 PE=4 SV=1	6923.04	5354.16	38.61	94.19

A0A024S7X2;A0A2H2ZTU2;G0RJB7;P78976	1	1	2	TRG	TRC	21603	Small COPII coat GTPase sar1 OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_8999 PE=3 SV=1	617.19	684.15	1130.39	1211.47
A0A024RX79;A0A2H2YVL3;G0RXE9	2	2	7	TRG	TRC	65661	Subtilisin-like protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_116864 PE=4 SV=1	41.52	31.26	303.65	213.79
A0A024S5P6;A0A2H2ZIH2;G0RP66;P43317;W8GPQ7	1	1	7407	TRC	TRG	25323	Endo-1_4-beta-glucanase V OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_25940 PE=4 SV=1	24976.14	20196.92	0.29	5.81
G0RCE9;A0A024SGR2;A0A2H3A0R9	5	4	2	TRC	TRG	125391	Predicted protein OS=Hypocrea jecorina (strain QM6a) OX=431241 GN=TRIREDRAFT_75450 PE=4 SV=1	93364.98	86040.98	49959.91	43659.92

A0A024SGY9;A0A2H3A176;G0RBE4	2	2	1	TRG	TRC	85399	Eukaryotic translation initiation factor 3 subunit B OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=PRT1 PE=3 SV=1	1573.90	1667.13	2349.68	2196.16
A0A024S0K1;A0A024S1T5;A0A2H2ZLP0;A0A2H2ZN46;A2VEB6;E5F5K8;E5F5K9;E5F5L3;E5F5L4;E5F5L5;E5F5L6;E5F5M1;E5F5M7;E5F5M9;E5F5N1;G0RSA2;G0RUB2;G9FLF6;Q65YQ7	2	2	53	TRG	TRC	46036	Glyco_18 domain-containing protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_124526 PE=3 SV=1	16.95	5.47	480.69	713.40
A0A024SNI3;G0RK84	1	1	2	TRG	TRC	161363	Uncharacterized protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_95059 PE=4 SV=1	483.88	673.23	1273.75	1230.26

A0A2H3A8G3	8	1	85	TRC	TRG	39929	Lipase G-D-S-L OS=Trichoderma parareesei OX=858221 GN=A9Z42_0091020 PE=4 SV=1	1760.63	1385.87	3.40	33.83
A0A2H2ZNG2	4	4	4	TRG	TRC	93793	ATP-dependent DNA helicase PIF1 OS=Trichoderma para-reesei OX=858221 GN=PIF1 PE=3 SV=1	118.94	105.77	609.09	320.05
G0RJ22;A0A024S7P0;A0A2H2ZEP4	3	3	1	TRC	TRG	72103	Predicted protein OS=Hypocrea jecorina (strain QM6a) OX=431241 GN=TRIREDRAFT_77795 PE=4 SV=1	20589.90	18246.60	15234.62	13113.18
A0A024SAF4;A0A2H2ZZH3;G0RLL5	3	3	1	TRG	TRC	43334	Glycoside hydrolase OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_110434 PE=4 SV=1	261.51	266.71	341.21	429.06

A0A024SM18;A0A2H2ZDP7;G0RKT7	1	1	4	TRG	TRC	21584	Uncharacterized protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_69766 PE=4 SV=1	146.51	121.43	779.96	265.20
A0A024SEG0;A0A2H2ZQS3;G0RHL9	2	1	2	TRG	TRC	58776	Splicing factor 3A subunit 3 OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_136759 PE=4 SV=1	540.11	585.24	1073.39	717.08
A0A024S726;G0RQ52	1	1	2	TRC	TRG	67265	Uncharacterized protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_101539 PE=4 SV=1	3992.38	3699.52	2977.77	1840.89

A0A024SFE2;A0A2H2ZGW1;G0RG61	3	3	1	TRC	TRG	162571	ABC transporter OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_75459 PE=3 SV=1	20694.07	18763.36	17835.33	15480.50
A0A2H2ZGX4	5	5	1	TRC	TRG	87696	Uncharacterized protein OS=Trichoderma parareesei OX=858221 GN=A9Z42_0028830 PE=4 SV=1	15957.50	13560.99	12949.29	10967.19
A0A2H2ZVU7;A0A024SDA9;G0RL92	2	2	1	TRC	TRG	139328	Pre-rRNA processing protein Rrp12 OS=Trichoderma parareesei OX=858221 GN=A9Z42_0005020 PE=4 SV=1	1391.09	1330.94	1267.09	1063.80
A0A2H2ZLT6;A0A024RZU0;G0RVX8	2	2	2	TRG	TRC	105926	Uncharacterized protein OS=Trichoderma parareesei OX=858221 GN=A9Z42_0078500 PE=3 SV=1	248.30	306.26	336.50	619.33

A0A024S109;A0A2H2ZI87;G0RTC0	7	7	1	TRC	TRG	42058	ECM33-like protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_38236 PE=4 SV=1	663.27	584.44	579.72	543.41
G0RT03;A0A024SBQ1	2	2	46	TRG	TRC	190127	Predicted protein OS=Hypocrea jecorina (strain QM6a) OX=431241 GN=TRIREDRAFT_123493 PE=4 SV=1	16.82	17.21	1551.95	30.03
A0A024SEY9	1	1	2	TRC	TRG	7705	Uncharacterized protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_77182 PE=4 SV=1	932.46	1059.26	368.66	932.76
A0A2H2ZF40;G0RKF3	1	1	1	TRC	TRG	59226	Uncharacterized protein OS=Trichoderma parareesei OX=858221 GN=A9Z42_0068410 PE=4 SV=1	665.45	723.55	35.38	972.25

A0A024S457;G0RQN8	1	1	3	TRG	TRC	88762	Uncharacterized protein OS=Hypocrea jecorina (strain ATCC 56765 / BCRC 32924 / NRRL 11460 / Rut C-30) OX=1344414 GN=M419DRAFT_84016 PE=4 SV=1	62.69	55.43	363.32	37.48
A0A2H2ZV70;A0A024S9J8;G0RMY0	2	2	1	TRG	TRC	137416	Uncharacterized protein OS=Trichoderma parareesei OX=858221 GN=A9Z42_0051510 PE=4 SV=1	217.70	190.90	399.59	161.82

Annexure II: Secretome profile of *Penicillium janthinellum* NCIM1366

Accession	Peptide count	Unique peptides	Max fold change	Highest mean condition	Lowest mean condition	Mass	Description	Normalized Abundance Pj-Cel 01_20200102_TRC_01	Normalized Abundance Pj-Cel 02_20200102_TRC_02	Normalized Abundance Pj-Glu 01_20200102_TRG_01	Normalized Abundance Pj-Glu 02_20200102_TRG_02
S7ZRD6;C9EI49;F6L7A1;U3N6C8;A0A0S2N073;B2LSE2;B3GS64	7	3	Infinity	PJC	PJG	58144.1559	Glucanase OS=Penicillium oxalicum (strain 114-2 / CGMCC 5302) OX=933388 GN=PDE_07945 PE=3 SV=1	28347.40	28308.68	0.00	0.00
A0A0F7TXT6;A0A1Q5ULE5;A0A1S9RDR1;B8Q961;A0A0A2JSU0	4	4	Infinity	PJC	PJG	45475.9853	Putative Endo-1_4-beta-D-glucanase OS=Penicillium brasilianum OX=104259 GN=PMG11_10007 PE=3 SV=1	22909.89	22991.33	0.00	0.00
A0A1V6YI05	2	1	Infinity	PJC	PJG	104466.651	Uncharacterized protein OS=Penicillium nalgiovense OX=60175 GN=PENNAL_c0020G04807 PE=4 SV=1	2799.01	2809.33	0.00	0.00
A0A1Q5TDS9	2	2	110463.54	PJC	PJG	180216.4071	BTB/POZ domain-containing protein 1 OS=Penicillium subrubescens OX=1316194 GN=PENSUB_9457 PE=4 SV=1	3472.87	3443.37	0.04	0.02
A0A1Q5T9M4	2	2	Infinity	PJC	PJG	50197.7283	Uncharacterized protein OS=Penicillium subrubescens OX=1316194 GN=PENSUB_10425 PE=4 SV=1	8813.01	9055.11	0.00	0.00
A0A135LPM4	3	2	4166.36	PJC	PJG	120759.8102	Cytochrome P450 OS=Penicillium patulum OX=5078 GN=PGRI_064880 PE=4 SV=1	84347.66	87267.34	20.48	20.71

A0A093UWI5;A0A093UXI4;A0A093VD53;B6QTV9;A0A0F7U1Y3;A0A1Q5U5B4;A0A1S9RYQ3;A0A2H3IJD6;B8MP33;S8B4L4	3	2	20155.02	PJC	PJG	107073.2847	Sodium/potassium-transporting ATPase subunit alpha-1 OS=Talaromyces marneffei PM1 OX=1077442 GN=GQ26_0270740 PE=3 SV=1	1940.39	1963.29	0.11	0.08
A0A1Q5UFZ3;A0A1V6T7J7	6	4	2396803.81	PJC	PJG	48514.6971	Glucanase OS=Penicillium subrubescens OX=1316194 GN=PENSUB_3088 PE=3 SV=1	8475.35	9007.99	0.00	0.00
A0A1Q5UH86;A0A1V6TMY4;W6R7Z3;A0A0G2UM08;A0A0G4PQY1;A0A167SNQ6;A0A1V6TE85;A0A1V6V332;A0A1V6YTI1;B6HE71;Q5S1P9;B8MK69	14	4	28184.23	PJC	PJG	57684.5604	Glucanase OS=Penicillium subrubescens OX=1316194 GN=PENSUB_2616 PE=3 SV=1	10045.34	10486.13	0.39	0.33
T1WI91	2	1	1209577.68	PJC	PJG	60545.8639	Glucanase OS=Penicillium granulatum OX=395885 GN=cbh1 PE=2 SV=1	733.13	690.87	0.00	0.00
A0A0U1M963	1	1	Infinity	PJG	PJC	120441.8002	Valyl-tRNA synthetase OS=Talaromyces islandicus OX=28573 GN=PISL3812_09153 PE=3 SV=1	0.00	0.00	107.71	103.04
A0A1Q5UI44;A0A0A2JRC1;A0A0A2L5I7;A0A0G4NZZ9;A0A0M8NPY9;A0A1V6QYU6;A0A1V6SRY1;K9G2R5;K9GUL5	2	1	Infinity	PJC	PJG	42252.3882	Glucanase OS=Penicillium subrubescens OX=1316194 GN=PENSUB_2291 PE=3 SV=1	11186.76	12302.22	0.00	0.00
A0A0F7KIJ2;A0A1Q5UIL1;O94221	2	2	Infinity	PJC	PJG	48172.7888	Alpha-galactosidase OS=Talaromyces purpureogenus OX=1266744 PE=3 SV=1	4917.55	5427.71	0.00	0.00

A0A0A2L7T5;A0A1V6RZ76	1	1	Infinity	PJG	PJC	82419.2109	Dynamin OS=Penicillium italicum OX=40296 GN=PITC_010890 PE=3 SV=1	0.00	0.00	77.09	72.82
A0A088DLG0	8	1	20103.03	PJC	PJG	57925.891	Glucanase (Fragment) OS=Penicillium canescens OX=5083 GN=cbhI PE=3 SV=1	2466.59	2747.35	0.14	0.12
A0A1V6P1A2	2	2	762.84	PJC	PJG	92421.452	DRMBL domain-containing protein OS=Penicillium decumbens OX=69771 GN=PENDEC_c024G04321 PE=4 SV=1	3057.40	3034.33	4.17	3.81
A0A0U1LLY3	1	1	Infinity	PJC	PJG	90354.9513	Uncharacterized protein OS=Talaromyces islandicus OX=28573 GN=PISL3812_01370 PE=4 SV=1	7826.54	8990.96	0.00	0.00
B8MM77	1	1	Infinity	PJC	PJG	61764.7025	Purine permease_ putative OS=Talaromyces stipitatus (strain ATCC 10500 / CBS 375.48 / QM 6759 / NRRL 1006) OX=441959 GN=TSTA_098460 PE=4 SV=1	1013.18	1132.31	0.00	0.00
A0A0N0RZT8;A0A1V6PZ34	1	1	10715.60	PJC	PJG	61412.3202	Fungal_trans domain-containing protein OS=Penicillium nordicum OX=229535 GN=ACN38_g1741 PE=4 SV=1	1801.60	1999.66	0.15	0.21
A0A0F7TNK0	1	1	14838.95	PJC	PJG	54123.3625	Uncharacterized protein OS=Penicillium brasilianum OX=104259 GN=PMG11_05680 PE=4 SV=1	975.09	1056.95	0.02	0.12
A0A0F7TP54;A0A1Q5UCC0;A0A1S9S1T2;S7ZGU9	1	1	Infinity	PJC	PJG	33972.7253	Uncharacterized protein OS=Penicillium brasilianum OX=104259 GN=PMG11_02960 PE=4 SV=1	1583.78	1840.99	0.00	0.00
A0A093V3D6;A0A1Q5THY0;B6QI81	1	1	Infinity	PJG	PJC	65886.6217	Glucoamylase OS=Talaromyces marneffei PM1 OX=1077442 GN=GQ26_0160970 PE=3 SV=1	0.00	0.00	236.79	211.04

A0A1Q5UKX2;A0A0F7VJX3;A0A1V6SJT6	2	2	3.17	PJC	PJG	36258.4583	Endo-chitosanase OS=Penicillium subrubescens OX=1316194 GN=PENSUB_1236 PE=3 SV=1	1346.73	1334.43	426.81	418.35
B8LTF8	3	3	10368.15	PJC	PJG	71617.6234	Dipeptidyl peptidase IV_ putative OS=Talaromyces stipitatus (strain ATCC 10500 / CBS 375.48 / QM 6759 / NRRL 1006) OX=441959 GN=TSTA_064970 PE=4 SV=1	157149.99	162820.72	16.92	13.94
B8MGB4	1	1	Infinity	PJG	PJC	69801.1356	C6 transcription factor_ putative OS=Talaromyces stipitatus (strain ATCC 10500 / CBS 375.48 / QM 6759 / NRRL 1006) OX=441959 GN=TSTA_013370 PE=4 SV=1	0.00	0.00	68.73	76.48
A0A1V6V0J6	2	2	73827.83	PJC	PJG	103842.0059	Uncharacterized protein OS=Penicillium coprophilum OX=36646 GN=PENCOP_c002G03145 PE=4 SV=1	126801.39	131198.56	1.49	2.00
A0A0G4PWZ1	2	2	837.04	PJC	PJG	62200.9349	Phosphate transporter OS=Penicillium camemberti FM 013 OX=1429867 GN=PCAMFM013_S057g000005 PE=3 SV=1	4542.70	4271.24	5.70	4.83
W6QDH0	1	1	4079.76	PJC	PJG	61888.2305	Zn(2)-C6 fungal-type DNA-binding domain OS=Penicillium roqueforti (strain FM164) OX=1365484 GN=PROQFM164_S01g001458 PE=4 SV=1	6944.17	8371.12	1.73	2.03
W6PTR4	1	1	20005.73	PJC	PJG	76524.0114	Acetoacetyl-CoA synthase OS=Penicillium roqueforti (strain FM164) OX=1365484 GN=PROQFM164_S01g000975 PE=4 SV=1	2543.89	2553.32	0.25	0.00

A0A1V6TA46	1	1	33488.36	PJC	PJG	10420.5836	Uncharacterized protein OS=Penicillium flavigenum OX=254877 GN=PENFLA_c012G05911 PE=4 SV=1	3473.59	4114.97	0.00	0.23
K9G0W3;K9FDB1;K9FVQ7;K9GC76	3	2	297.75	PJC	PJG	21667.4854	HTH CENPB-type domain-containing protein OS=Penicillium digitatum (strain PHI26 / CECT 20796) OX=1170229 GN=PDIG_28900 PE=4 SV=1	6026.68	6844.63	23.30	19.93
S7ZQC9	4	3	14.19	PJC	PJG	205235.2566	Uncharacterized protein OS=Penicillium oxalicum (strain 114-2 / CGMCC 5302) OX=933388 GN=PDE_07891 PE=4 SV=1	9578.31	10175.23	726.21	666.28
A0A0A2L603	3	2	10.84	PJC	PJG	209031.546	Helicase_ C-terminal OS=Penicillium italicum OX=40296 GN=PITC_082440 PE=4 SV=1	89129.71	82998.44	7678.19	8207.93
Q06886	12	4	23487.09	PJC	PJG	57877.032	Exoglucanase 1 OS=Penicillium janthinellum OX=5079 GN=cbh1 PE=2 SV=1	344039.19	346375.50	17.82	11.57
A0A0F7TSC9;A0A1Q5UH83;A0A1S9RT40;A0A2H3J2E4;B8LZT2;G9DBG4	2	2	426.15	PJC	PJG	51394.0049	Glucanase OS=Penicillium brasilianum OX=104259 GN=PMG11_06607 PE=3 SV=1	13991.67	11219.90	32.49	26.67
A0A2H3IHI9	2	2	1320.81	PJC	PJG	38745.2496	Uncharacterized protein OS=Penicillium sp. 'occitanis' OX=290292 GN=PENO1_090540 PE=4 SV=1	3918.23	4040.92	3.59	2.43
A0A1V6SPB5;A0A2H3IFK0;B8M856	1	1	15.60	PJG	PJC	83835.9336	Uncharacterized protein OS=Penicillium steckii OX=303698 GN=PENSTE_c028G08968 PE=4 SV=1	23.25	22.62	389.72	325.89

A0A1Q5UIA5;A0A088STL5;A0A0A2LID0;A0A0G4NWG3;A0A0G4PHA5;A0A117NLW4;A0A135L9H9;A0A135LXP4;A0A1V6UHP5;W6PXL5	6	4	20979.67	PJC	PJG	78177.8512	Glucanase OS=Penicillium subrubescens OX=1316194 GN=PENSUB_2329 PE=3 SV=1	7309.05	6523.67	0.66	0.00
A0A1V6R394	2	1	1.79	PJC	PJG	60514.8574	Condensation domain-containing protein OS=Penicillium solitum OX=60172 GN=PENSOL_c018G01093 PE=4 SV=1	89078.82	86038.84	48233.12	49812.60
A0A1V6PVF6	3	3	2.28	PJC	PJG	33822.4702	S-methyl-5'-thioadenosine phosphorylase OS=Penicillium antarcticum OX=416450 GN=PENANT_c030G09407 PE=3 SV=1	94561.30	94410.77	42820.77	39979.38
A0A1Q5UIH7;A0A088S9G7;A0A0A2JQK5;A0A0F7TT43;A0A101MFU7;A0A1S9RU93;A0A1V6N644;H9TUB0;K9FWI1;K9GSU4;S7ZX22	3	3	86.29	PJC	PJG	65253.3804	Endoglucanase B OS=Penicillium subrubescens OX=1316194 GN=PENSUB_1985 PE=3 SV=1	1205.73	1307.17	17.30	11.82
A0A1S9RZR9;A0A0F7TWR3	3	3	23.09	PJC	PJG	143134.6085	Lipid A export ATP-binding/permease protein msbA OS=Penicillium brasilianum OX=104259 GN=PEBR_02635 PE=4 SV=1	4550.06	4980.30	234.03	178.74
B8MG06	1	1	40.64	PJC	PJG	80577.6966	Uncharacterized protein OS=Talaromyces stipitatus (strain ATCC 10500 / CBS 375.48 / QM 6759 / NRRL 1006) OX=441959 GN=TSTA_009910 PE=4 SV=1	826.68	825.04	16.93	23.72

A0A1Q5UPK8;A0A0F7TUC2;A0A1B1MQM8;A0A1S9RYE7;S8B0B4	5	5	64.56	PJC	PJG	38636.582	Uncharacterized protein OS=Penicillium subrubescens OX=1316194 GN=PENSUB_14178 PE=3 SV=1	4239.99	3022.93	61.19	51.31
A0A093V670;A0A0A2IL19;A0A0A2KKJ9;A0A0F7U3Z6;A0A0G4PS76;A0A0M8PDT1;A0A0U1LTI1;A0A117NKA2;A0A135LN00;A0A167QIU1;A0A1F5LW18;A0A1Q5UAD4;A0A1S9REU8;A0A1V6NLL5;A0A1V6P748;A0A1V6PU96;A0A1V6RBV4;A0A1V6SDD1;A0A1V6SMU8;A0A1V6TJA2;A0A1V6UP83;A0A1V6XNB7;B6GZF2;B6QHT9;B8MIM8;K9FUJ8;K9GDI2;S7ZTZ1;W6R1D9	1	1	12.45	PJC	PJG	108085.5071	Long chain acyl-CoA synthetase 7_ peroxi-somal OS=Talaromyces marneffei PM1 OX=1077442 GN=GQ26_0470430 PE=4 SV=1	5070.74	5331.42	369.62	465.99
B8MED7	2	2	112.94	PJC	PJG	21275.368	Uncharacterized protein OS=Talaromyces stipitatus (strain ATCC 10500 / CBS 375.48 / QM 6759 / NRRL 1006) OX=441959 GN=TSTA_016420 PE=3 SV=1	2662.67	2270.76	26.65	17.03
A0A1Q5UEG0	3	3	698.30	PJC	PJG	50398.4505	Uncharacterized protein OS=Penicillium subrubescens OX=1316194 GN=PENSUB_3678 PE=4 SV=1	4988.35	9502.02	11.08	9.67

A0A117NS52	3	3	2.62	PJC	PJG	231443.6984	Uncharacterized protein OS=Penicillium freii OX=48697 GN=ACN42_g631 PE=4 SV=1	21084.18	20354.84	7520.68	8304.71
A0A1V6SXC4	4	4	2.33	PJC	PJG	140725.7728	HECT domain-containing protein OS=Penicillium steckii OX=303698 GN=PENSTE_c018G08715 PE=4 SV=1	9819.67	9444.46	4324.30	3959.56
S8AMY9	2	2	1.83	PJC	PJG	13386.0859	Uncharacterized protein OS=Penicillium oxalicum (strain 114-2 / CGMCC 5302) OX=933388 GN=PDE_02206 PE=4 SV=1	104247.92	103435.70	58577.26	54635.52
A0A088S933;A0A1Q5UEK4;S8AP92;A0A2H3IGC2;B8MEX2	2	1	40.11	PJC	PJG	30033.8776	Endo-1_4-beta-xylanase OS=Penicillium oxalicum OX=69781 PE=3 SV=1	782.10	884.49	25.02	16.53
A0A1S9S0X0	2	1	2710.11	PJC	PJG	86566.6587	Uncharacterized protein OS=Penicillium brasilianum OX=104259 GN=PEBR_01371 PE=4 SV=1	1500.79	1623.11	1.15	0.00
S7ZCW4	2	2	24.28	PJC	PJG	54148.4122	Putative exo-beta-1_3-galactanase OS=Penicillium oxalicum (strain 114-2 / CGMCC 5302) OX=933388 GN=PDE_01468 PE=3 SV=1	1968.97	1832.27	60.20	96.35

Annexure III: Hemicellulase genes identified from the genome of *P.janthinellum* NCIM1366

Gene	Annotation	CAZy family
g515_t1	Arabinan endo-1,5-alpha-L-arabinosidase (EC 3.2.1.99)	GH43_6
g549_t1	Putative beta-1,4-mannanase	GH26
g748_t1	Beta-xylanase (EC 3.2.1.8)	GH10
g916_t1	Arabinogalactan endo-beta-1,4-galactanase (EC 3.2.1.89)	GH53
g938_t1	Alpha-galactosidase (EC 3.2.1.22) (Melibiase)	GH27
g1331_t1	Probable xyloglucan-specific endo-beta-1,4-glucanase A	GH12
g1786_t1	Alpha-L-arabinofuranosidase 2 (EC 3.2.1.55)	GH51_2
g1956_t1	Probable alpha-N-arabinofuranosidase A	GH51_2
g2256_t1	Putative beta-mannosidase	GH2
g2435_t1	Beta-xylosidase	GH43_14
g2530_t1	Arabinan endo-1,5-alpha-L-arabinosidase (EC 3.2.1.99)	GH43_6
g2952_t1	Putative exo-beta-1,3-galactanase	GH43_25
g3156_t1	Arabinan endo-1,5-alpha-L-arabinosidase (EC 3.2.1.99)	GH43_6
g3248_t1	Beta-galactosidase (EC 3.2.1.23)	GH35
g3613_t1	Exo-arabinanase GH93	GH93
g3650_t1	Alpha-galactosidase (EC 3.2.1.22) (Melibiase)	GH27
g4026_t1	Putative Alpha-L-arabinofuranosidase	GH62
g4028_t1	Putative endo-beta-1,4-xylanase	GH30_7
g4225_t1	Putative Alpha-L-arabinofuranosidase axhA	GH62
g4267_t1	Putative Beta-xylosidase	GH43_13
g4488_t1	Similar to Probable beta-galactosidase A acc. no. A2QAN3	GH35
g4942_t1	Putative Beta-1,3-mannanase	GH5_31
g4996_t1	Beta-D-xylosidase (EC 3.2.1.37)	GH43_1
g5063_t1	Beta-galactosidase (EC 3.2.1.23)	GH35
g5259_t1	Alpha-L-arabinofuranosidase	GH54
g5356_t1	Arabinan endo-1,5-alpha-L-arabinosidase (EC 3.2.1.99)	GH43_6
g5362_t1	Putative Alpha-galactosidase C	GH36
g5394_t1	Alpha-L-arabinofuranosidase 2 (EC 3.2.1.55)	GH51_2
g6413_t1	Alpha-L-arabinofuranosidase 3	GH43_29
g6445_t1	Probable arabinan endo-1,5-alpha-L-arabinosidase C (EC 3.2.1.99) (Endo-1,5-alpha-L-arabinanase C) (ABN C)	GH43_6
g6737_t1	Probable alpha-N-arabinofuranosidase B	GH54
g6834_t1	Probable alpha-L-arabinofuranosidase B (ABF B) (Arabinosidase B) (EC 3.2.1.55)	GH54
g6974_t1	Probable xyloglucan-specific endo-beta-1,4-glucanase A	GH12
g6976_t1	Alpha-L-arabinofuranosidase, putative	GH62
g7193_t1	Xylanase GH30	GH30_7
g7196_t1	Non-reducing end alpha-L-arabinofuranosidase	GH43_14

g8062_t1	Probable alpha-L-arabinofuranosidase B (ABF B) (Arabinosidase B) (EC 3.2.1.55)	GH54
g8174_t1	Endo-1,4-beta-xylanase (EC 3.2.1.8)	GH11
g8569_t1	Endo-1,4-beta-xylanase A (Xylanase A) (EC 3.2.1.8) (1,4-beta-D-xylan xylanohydrolase A)	GH10
g8709_t1	Probable alpha-galactosidase G (EC 3.2.1.22) (Melibiase G)	GH36
g9001_t1	Beta-galactosidase (EC 3.2.1.23)	GH35
g9112_t1	Arabinan endo-1,5-alpha-L-arabinosidase (EC 3.2.1.99)	GH43_6
g9222_t1	Beta-xylosidase, putative	GH3
g9307_t1	Arabinosidase, putative	GH43_22
g9309_t1	Alpha-glucuronidase (EC 3.2.1.139)	GH67
g9310_t1	Putative endo-beta-1,4-xylanase	GH30_7
g9622_t1	Alpha-galactosidase (EC 3.2.1.22) (Melibiase)	GH27
g9824_t1	Putative Exo-beta-1,4-galactanase	GH35
g9978_t1	Beta-galactosidase (EC 3.2.1.23)	GH35
g10103_t1	Putative beta-xylosidase/alpha-L-arabinofuranosidase	GH43_14
g10248_t1	Endo-1,4-beta-xylanase (EC 3.2.1.8)	GH11
g9192_t1	Uncharacterized protein	GH12
g4027_t1	Putative Cellobiohydrolase 1	GH43_36
g11631_t1	Endo-1,4-beta-xylanase (EC 3.2.1.8)	GH11
g3484_t1	Acetylxylan esterase 2 (EC 3.1.1.72) (AXE II)	CE5
g7103_t1	Acetylxylan esterase A (EC 3.1.1.72) (AXE I) (Acetylxylan esterase 1)	CE1
g10614_t1	Beta-xylosidase GH3	GH3
g11582_t1	Putative beta-xylosidase/alpha-L-arabinofuranosidase	GH43_14
g236_t1	Uncharacterized protein	GH51_2
g10741_t1	Uncharacterized protein	GH51_2
g11479_t1	Putative Alpha-L-arabinofuranosidase	GH51_1
g11889_t1	Uncharacterized protein	GH36
g966_t1	Alpha-galactosidase (EC 3.2.1.22) (Melibiase)	GH27
g4069_t1	Uncharacterized protein	GH2
g5070_t1	Uncharacterized protein	GH2
g6883_t1	Uncharacterized protein	GH2
g7438_t1	Uncharacterized protein	GH35
g10385_t1	Alpha-galactosidase (EC 3.2.1.22) (Melibiase)	GH27
g10832_t1	Uncharacterized protein	GH42

Annexure IV

AcSIR Course Work

No.	Course No.	Title
1	BIO-NIIST-1-0001	Biostatistics
2	BIO-NIIST-1-0002	Bioinformatics
3	BIO-NIIST-1-0003	Basic Chemistry
4	BIO-NIIST-1-0004	Research Methodology, Communication/ Ethics/ Safety
5	BIO-NIIST-2-4101	Biotechniques and Instrumentation
6	BIO-NIIST-2-4102	Protein Sciences and Proteomics
7	BIO-NIIST-2-4104	Basic Molecular Biology
8	BIO-NIIST-3-4101	Seminar Course
9	BIO-NIIST-3-4104	Bioprocess Technology
10	BIO-NIIST-3-4105	Enzymology & Enzyme Technology
11	BIO-NIIST-4-0001	Project proposal
12	BIO-NIIST-4-0002	Review article
13	BIO-NIIST-4-0003	CSIR-800 (Societal Program)

ABSTRACT

Name of the Student: Athiraj S R **Registration No.:** 10BB16A39012
Faculty of Study: Biological Sciences **Year of Submission:** 2025
AcSIR academic centre/CSIR Lab: CSIR-NIIST, Thiruvananthapuram
Name of the Supervisor: Dr. Rajeev K. Sukumaran
Title of the thesis: Multi-Omics studies on the hypercellulolytic mutant *Penicillium Janthinellum* NCIM 1366 for elucidating its CAZyme profile and cellulase gene regulation

---

The economic production of the second generation (2G) bioethanol is largely hindered by the high cost of biomass hydrolysis enzymes. The current industrial standard for cellulase production is *Trichoderma reesei* RUT-C30; however its low β-glucosidase output largely hinders the hydrolysis efficiency. This thesis establishes *Penicillium janthinellum* NCIM1366 as potent alternative organism with higher enzyme activities, protein production and secretion efficiency. The genome of *P. janthinellum*, at 37.6 Mbp, encodes an extensive array of carbohydrate-active enzymes (CAZymes), comprising approximately 5.4% of all predicted genes-notably higher than the CAZyme content reported in other industrially relevant fungi. A large repertoire of cellulolytic genes, including multiple endoglucanases, cellobiohydrolases, β-glucosidases, lytic polysaccharide monooxygenases (LPMOs), and hemicellulases in the genome collectively highlight the genetic basis for the superior biomass hydrolysis. Cellulose-induced transcriptome of *P. janthinellum* with cellulose showed CAZymes with significant differential expression which accounts for 19.4 % of the total differentially expressed genes 4h post-induction. The study also revealed a probable regulatory network encompassing sugar sensing and transport, signal transduction via diverse kinase cascades, and a hierarchical system of transcription factors. We propose a mechanism in which ClrB acts as a master activator, directly regulating major cellulase and hemicellulase genes and also controlling additional regulators such as XlnR and HacA. The insights from proteomic, transcriptomic and genomic analysis can be utilized for rational metabolic engineering to enhance cellulase titers, thermostability, and performance on industrial substrates.

Research Output

1. List of publications emanating from the thesis work

a. **Sreeja-Raju, A.**, Christopher, M., Kooloth-Valappil, P. *et al. Penicillium janthinellum* NCIM1366 shows improved biomass hydrolysis and a larger number of CAZymes with higher induction levels over *Trichoderma reesei* RUT-C30. *Biotechnol Biofuels* 13, 196 (2020). https://doi.org/10.1186/s13068-020-01830-9.

2. List of publications not related to the thesis

a. Christopher, M., **Sreeja-Raju, A.**, Kooloth-Valappil, P. *et al.* Cellulase Hyper-Producing Fungus *Penicillium janthinellum* NCIM 1366 Elaborates a Wider Array of Proteins Involved in Transport and Secretion, Potentially Enabling a Diverse Substrate Range (2023). *Bioenerg. Res.* **16**, 61–73. https://doi.org/10.1007/s12155-022-10407-3.

b. Kooloth-Valappil, P., Christopher, M., **Sreeja-Raju, A**., Mathew, R.M., Kuni-Parambil, R., Abraham, A., Sankar, M., Pandey, A. and Sukumaran, R.K. (2021). Draft genome of the glucose tolerant β-glucosidase producing rare *Aspergillus unguis* reveals complete cellulolytic machinery with multiple beta-glucosidase genes. Fungal Genetics and Biology, 151. https://doi.org/10.1016/j.fgb.2021.103551

c. Christopher, M., Kooloth-Valappil, P., **Sreeja-Raju, A.**, and Sukumaran, R. K. (2021). Repurposing proteases: An in-silico analysis of the binding potential of extracellular fungal proteases with selected viral proteins. Bioresour. Technol. Reports 15. https://doi.org/10.1016/j.biteb.2021.100756

d. Sukumaran, R.K., Christopher, M., Kooloth-Valappil, P., **Sreeja-Raju, A**., Mathew, R.M., Sankar, M., Puthiyamadam, A., Adarsh, V.-P., Aswathi, A., Rebinro, V., Abraham, A., Pandey, A. (2021). Addressing challenges in production of cellulases for biomass hydrolysis: Targeted interventions into the genetics of cellulase-producing fungi. Bioresour. Technol. 329. https://doi.org/10.1016/j.biortech.2021.124746

3. Book chapters

a. Sukumaran, R.K., Christopher, M., **Sreeja-Raju, A.,** Sankar, M., Kooloth-Valappil, P., Gnanaraj, V.R., Puthiyamadam, A., Mathew, R.M., Adarsh, V.-P., 2022. Utilization of Agro-Industrial Wastes for Biofuel Generation, in: Valorization of Agro-Industrial Byproducts Sustainable Approaches for Industrial Transformation. CRC Press, pp. 285-314. https://doi.org/10.1201/9781003125679

b. Sukumaran RK, Sankar M, Adarsh VP, Mathew RM, **Sreeja-Raju AR**, Athulya, Neetha PS, Pahy B, Gnanaraj VR. 2024. Technological Advancements in Enzyme Production for 2G Ethanol. In: Suresha GS, Krishnappa G, Palanichamy M, Mahadeva Swamy HK, Kuppusamy H, Govindakurup H.(eds) Value Addition and Product Diversification in Sugarcane. Springer, Singapore. https://doi.org/10.1007/978-981-97-7228-5_18

c. Mathew RM, Sankar M, **Sreeja-Raju A**, Adarsh VP, Athulya, Puthiyamadam A, Christopher M, Jose J, Rebinro VG, Sukumaran RK. 2024. Fungal Production β-Glucosidases. In: Singhania et al (eds). Biomass Hydrolyzing Enzymes: Basics, Advancements, and Applications. CRC Press, Boca Raton, FL, USA. Pp. 48-62. https://doi.org/10.1201/9781003335313

4. Workshops attended

a. One Day Workshop on Road Map to Transcriptome Interpretation, Cochin University of Science and Technology (CUSAT), Kochi, India, December 20, 2018.

5. List of conference presentations

a. **Athira Raj SR,** Amith Abraham, Meera Christopher, Prajeesh KV, Digambar V Gokhale, Ashok Pandey, Rajeev K Sukumaran, Genome sequencing and functional annotation of the fungus *Penicillium janthinellum* provide insights into its cellulolytic system. Poster preseted at International Conference on New Horizons in Biotechnology (NHBT), Thiruvananthapuram, November 20-24, 2019.

b. **Athira Raj SR**, Amith Abraham, Meera Christopher, Prajeesh KV, Digambar V Gokhale, Ashok Pandey, Rajeev K Sukumaran, Transcriptome analysis identifies candidate genes involved in regulation of cellulases in *Penicillium janthinellum* NCIM-1366. Poster preseted at International Conference on Biotechnology for Sustainable Agriculture, Environment and Health (BASEH), Jaipur, India, April 4-8, 2021.

Abstracts for conference presentations

I. Genome sequencing and functional annotation of the fungus *Penicillium janthinellum* provide insights into its cellulolytic system

Athira Raj SR[1,2], Amith Abraham[2], Meera Christopher, [1,2] Prajeesh KV[1,2]
Digambar V Gokahle[3], Ashok Pandey[4], Rajeev K Sukumaran[2]*

[1]*Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India*
[2]*Biofuels and Biorefineries section, Microbial processes and Technology Division, CSIR-National Institute for Interdisciplinary Science and Technology, Industrial Estate P.O., Thiruvananthapuram, India*
[3]*CSIR-National Chemical Laboratory, Pune, India*
[4]*Centre for Innovation and Translational Research, CSIR-Indian Institute of Toxicology Research, Lucknow, India*

***** *Corresponding Author, Email: rajeevs@niist.res.in , Tel.: +91 4712515368*

Penicillium janthinellum NCIM1366 (*P.janthinellum*) is a filamentous fungus commonly associated with soil and decaying organic matters and it is known to be a hyper-producer of cellulases. It produces higher enzyme titers and outperforms the industrial cellulase producer *Trichoderma reesei* RUT-C30 in biomass hydrolysis. In order to identify the apparent reasons for its notable hydrolytic efficiency and to understand the features involved in cellulase regulation, the whole genome sequencing of *P.janthinellum* was performed using NGS technology. The complete genome of *P.janthinellum* has a size of 37.6 Mb, predicted to encode 12003 proteins. The genome showed an abundance of CAZymes (844) among which 25 % are glycoside hydrolases. The major cellulase degrading enzymes were identified from the genome, which includes endoglucanases, cellobiohydrolases, beta-glucosidases, LPMO, swollenin etc. The orthologs of most of the known cellulase regulators were also identified from the genome. These results would be useful understanding and modulating the cellulase expression in the fungus. Many *T. reesei* genes encoding carbohydrate-active enzymes are known to be distributed non-randomly in clusters, which enable a coordinated regulation and is considered to be the reason for its high cellulase producing efficiency. An assembled genome of *P.janthinellum* is expected to provide such insights into its hyper-cellulolytic activity that outperforms *T .reesei* and the regulation of cellulase genes in the fungus.

Keywords: *cellulase, Penicillium janthinellum, cellulase regulation, Penicillium genome, Trichoderma reesei*

II. **Transcriptome analysis identifies candidate genes involved in regulation of cellulases in *Penicillium janthinellum* NCIM-1366**

Athira Raj SR[1,2], Meera Christopher[1,2], Prajeesh KV[1,2], Amith Abraham[3]
Digambar V Gokhale[4], Ashok Pandey[5], Rajeev K Sukumaran[2*]

[1]*Academy of Scientific and Innovative Research (AcSIR), Ghaziabad-201002, India*
[2]*Biofuels and Biorefineries section, Microbial processes and Technology Division, CSIR-National Institute for Interdisciplinary Science and Technology, Industrial Estate P.O., Thiruvananthapuram, India*
[3]*Department of Chemical Engineering, Hanyang University, Seoul, Republic of Korea*
[4]*CSIR-National Chemical Laboratory, Pune, India*
[5]*Centre for Innovation and Translational Research, CSIR-Indian Institute of Toxicology Research, Lucknow, India*
***** *Corresponding Author, Tel.: +91 4712515368 Email: rajeevs@niist.res.in*

Cellulase enzyme system in filamentous fungi is under the control of a cellulase regulon which fine tunes the cellulase production according to the requirements of the organism and environmental conditions. These aspects have been studied elaborately in *Trichoderma reesei* RUT-C30, but the complete mechanism is still elusive. *Penicillium janthinellum* NCIM-1366 is a filamentous fungus known for its ability to secrete high titre of cellulase enzymes which outperform *T. reesei* in the hydrolysis of rice straw and sugarcane bagasse making it a potent industrial strain for biomass hydrolyzing enzyme production. Identification of the targets for genetic modifications intended to further improve cellulase production in *P. janthinellum* is hindered by the limited information on cellulase regulation in this fungus. In the current study, we have sequenced the transcriptome of the fungus using NGS under conditions of cellulose induction and compared against non-induced (glucose grown) culture, which identified a total of 22023 unigenes. Most of the major cellulase genes were upregulated during cellulose induction. Analysis of differentially expressed transcription factors identified homologs of known cellulase regulators which included CreA, XlnR, ClrB, PacC, AceA, LaeA and AraR. Three transcription factors which did not have a known homolog in other fungi, also showed very high expression levels compared to the control. Further studies are expected to reveal the role of these factors in regulating cellulases in this fungus, which will advance the knowledge on cellulase regulation in this fungus, and may serve as genetic targets for strain improvement.

Keywords: Penicillium janthinellum, transcriptome, cellulase, biofuel, Cre1

Biotechnology for Biofuels

RESEARCH Open Access

Penicillium janthinellum NCIM1366 shows improved biomass hydrolysis and a larger number of CAZymes with higher induction levels over *Trichoderma reesei* RUT-C30

AthiraRaj Sreeja-Raju[1,2], Meera Christopher[1,2], Prajeesh Kooloth-Valappil[1,2], Rajasree Kuni-Parambil[3], Digambar Vittal Gokhale[4], Meena Sankar[1,2], Amith Abraham[5], Ashok Pandey[6] and Rajeev K. Sukumaran[1,2*]

Abstract

Background: Major cost of bioethanol is attributed to enzymes employed in biomass hydrolysis. Biomass hydrolyzing enzymes are predominantly produced from the hyper-cellulolytic mutant filamentous fungus *Trichoderma reesei* RUT-C30. Several decades of research have failed to provide an industrial grade organism other than *T. reesei*, capable of producing higher titers of an effective synergistic biomass hydrolyzing enzyme cocktail. *Penicillium janthinellum* NCIM1366 was reported as a cellulase hyper producer and a potential alternative to *T. reesei*, but a comparison of their hydrolytic performance was seldom attempted.

Results: Hydrolysis of acid or alkali-pretreated rice straw using cellulase enzyme preparations from *P. janthinellum* and *T. reesei* indicated 37 and 43% higher glucose release, respectively, with *P. janthinellum* enzymes. A comparison of these fungi with respect to their secreted enzymes indicated that the crude enzyme preparation from *P. janthinellum* showed 28% higher overall cellulase activity. It also had an exceptional tenfold higher beta-glucosidase activity compared to that of *T. reesei*, leading to a lower cellobiose accumulation and thus alleviating the feedback inhibition. *P. janthinellum* secreted more number of proteins to the extracellular medium whose total concentration was 1.8-fold higher than *T. reesei*. Secretome analyses of the two fungi revealed higher number of CAZymes and a higher relative abundance of cellulases upon cellulose induction in the fungus.

Conclusions: The results revealed the ability of *P. janthinellum* for efficient biomass degradation through hyper cellulase production, and it outperformed the established industrial cellulase producer *T. reesei* in the hydrolysis experiments. A higher level of induction, larger number of secreted CAZymes and a high relative proportion of BGL to cellulases indicate the possible reasons for its performance advantage in biomass hydrolysis.

Keywords: *Penicillium janthinellum*, *Trichoderma reesei*, Cellulase, CAZymes, Secretome, Bioethanol

Background

Lignocellulosic biomass is mainly composed of three polymers—cellulose, hemicellulose, and lignin. Among them, cellulose is the most abundant component which consists of 35–50% of plant dry weight followed by hemicellulose (20–35%) and then lignin (5–30%) [1]. Degradation of lignocelluloses is generally carried out by

*Correspondence: rajeevs@niist.res.in
[1] Biofuels and Biorefineries Section, Microbial Processes and Technology Division, CSIR-National Institute for Interdisciplinary Science and Technology, Industrial Estate, Pappanamcode, Thiruvananthapuram 695019, India
Full list of author information is available at the end of the article

a set of enzymes including ligninases, hemicellulases, cellulases and other accessory enzymes. Their relative proportions and quantities can determine the efficiency of hydrolysis [2]. The major cost involved in bioethanol production from lignocellulosic biomass is contributed by the hydrolysis step, owing to the cost of production of these enzymes [3]. Thus, reducing the cost of enzymes is critical for making the process economical. The major sources of lignocellulose-degrading enzymes are filamentous fungi, mainly the genera *Trichoderma*, *Aspergillus* and *Penicillium* [4]. With cellulose being the major component, cellulases play an important role in the degradation process. Cellulases are mainly divided into three major groups based on their mode of action on cellulose: (1) endoglucanases, which randomly cleave internal β-1, 4 linkages in cellulose chain generating free ends; (2) cellobiohydrolases, which act in a processive manner on either reducing or non-reducing ends of the cellulose chain, releasing cellobiose as the major product and (3) beta-glucosidases that hydrolyze cellobiose into glucose [5].

Among the filamentous fungi, *Trichoderma reesei* is the most studied cellulase producer and considered as the model organism for cellulase research. It was originally isolated from the Solomon Islands during the Second World War, and the isolate was named *T. reesei* QM6a as part of the culture collection at the US Army Quarter Master Research and Development Center at Natick, Massachusetts. Attempts to improve the cellulase production from *T. reesei* QM6a by several generations of mutagenesis resulted in an enzyme hyper-secreting mutant named as *T. reesei* RUT-C30 [6]. Currently, *T. reesei* RUT-C30 is the predominantly used industrial cellulase producer, because of its ability to produce a high titer of cellulases and a gene regulation mechanism that is highly adapted for cellulose utilization [7]. Nevertheless, the cost of cellulases is still a major concern, with only a handful of companies commercially manufacturing biomass hydrolysis enzymes, and transportation and storage of the enzyme itself adding significantly to the cost [8]. Availability of an alternate cellulase source, better than *T. reesei* in terms of higher yields, an effective ratio of different glucanases to effect improved biomass hydrolysis performance, and/or better production economics as determined by higher specific activities, shorter fermentation times, etc., would be highly advantageous to the 2G-ethanol industry, especially in countries where commercial production of efficient biomass hydrolyzing cellulases are not available.

Penicillium janthinellum NCIM 1171 is a filamentous soil fungus known for its efficient cellulase production and hydrolysis efficiency [9, 10]. Classical mutagenesis studies conducted at CSIR-NCL, Pune had yielded three mutants of *P. janthinellum* NCIM 1171 with enhanced cellulase production and they were named as EMS-UV-8 (NCIM1366), EU-21 and EU2D-21 [11–13]. The mutant EMS-UV-8, named as isolate NCIM 1366 was reported to have higher beta-glucosidase (BGL) activity compared to other mutants [11]. While it's enzyme was not the best in terms of hydrolyzing pure cellulose [11], it was efficient in the hydrolysis of pretreated rice straw (unpublished results).The strain was also successfully used for bioethanol production from pretreated wheat straw [14]. Since the mutant strain produced cellulases efficient in hydrolysis of natural biomass substrates, *P. janthinellum* NCIM1366 was chosen as the model strain to be compared with *T. reesei* RUT-C30, the best known industrial producer of cellulase. Secretome analyses performed using liquid chromatography tandem mass spectrometry (LC–MS/MS) revealed higher number of CAZymes in *P. janthinellum* compared to *T. reesei,* and a higher relative abundance of cellulases upon induction using cellulose, which may explain the higher activity and better biomass hydrolytic performance of enzyme preparation from the fungus.

Results

P. janthinellum secretes higher amount of proteins compared to *T. reesei*

Cellulose is a large polymer and utilization of it requires secretion of enzyme by microorganisms to process it outside the cell, so that the simple sugars derived from its breakdown can be taken inside. The total secreted protein concentration in presence of cellulose is indicative of the efficiency of the fungus in utilizing the polymer, as efficient cellulose digestion typically requires a milieu of different enzyme activities, in addition to cellulases. *T. reesei* showed a maximum protein secretion of 0.28 mg/ml on the 6th day of growth, whereas *P. janthinellum* secreted the maximum protein on 10th day of growth which was ~ 1.8 times higher than *T. reesei* (Fig. 1a). At all the time points tested, extracellular protein concentration was higher in *P. janthinellum*. SDS-PAGE of the extracellular fractions from both fungi under uninduced (glucose grown) and induced (cellulose grown) conditions indicated significant elevation in secreted proteins upon cellulose induction (Fig. 1b). It was also observed that visibly, a greater number of extracellular proteins were secreted by *P. janthinellum.*

P. janthinellum shows lesser cellobiose accumulation in the hydrolysis medium which is indicative of its better beta-glucosidase activity

Cellobiose, the intermediate product of enzymatic cellulose hydrolysis, is produced through action of exoglucanases and is the substrate for cellobiase/

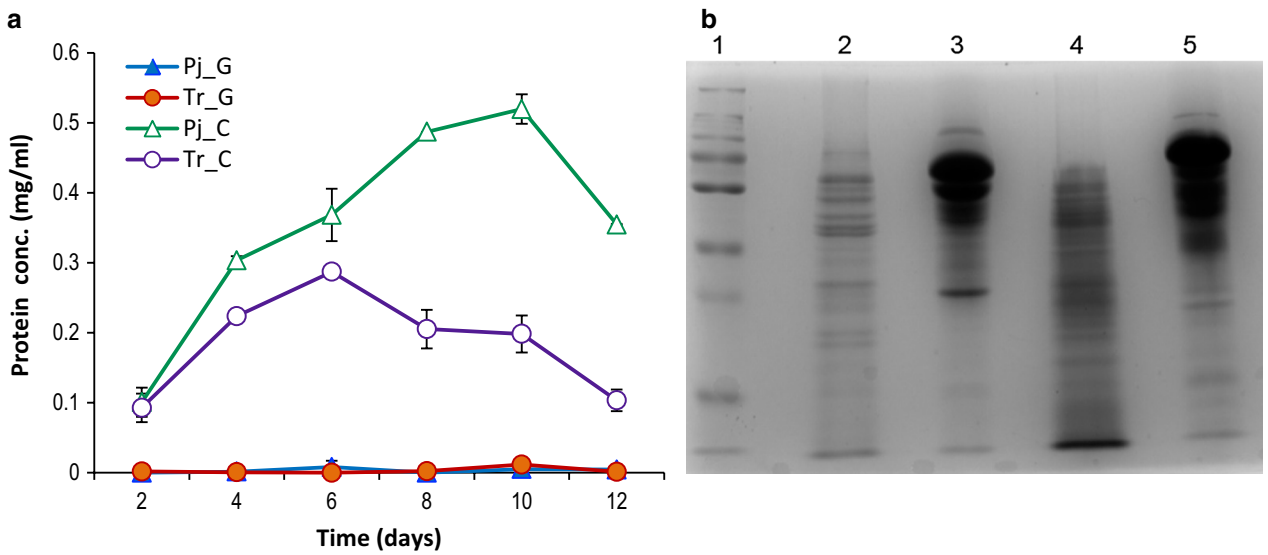


Fig. 1 Extracellular protein production by *T. reesei* and *P. janthinellum* in response to induction by cellulose. **a** Extracellular protein concentrations in *T. reesei* (Tr) and *P. janthinellum* (Pj) at different time points when cultivated in glucose and in cellulose. **b** SDS-PAGE showing differences in protein production by the two fungi. Lane information—1) Page Ruler ® pre-stained protein ladder, 2) Tr grown in glucose 3) Tr. grown in cellulose as sole C source 4) Pj grown in glucose 5) Pj. grown in cellulose as sole C source

beta-glucosidase. Cellobiose accumulation can lead to product inhibition of upstream enzymes (endoglucanases and cellobiohydrolases) thus slowing down the whole hydrolytic process [15, 16]. *T. reesei* is known to have limited cellobiase/beta-glucosidase (BGL) activity, and while it may be advantageous for the organism in tight regulation of cellulose metabolism while growing on natural substrates, it is a disadvantage for the biomass hydrolyzing enzyme cocktails produced using the fungus and often the *T. reesei* enzyme's lack of BGL activity is compensated by addition of BGL enzyme from other organisms. In this study, it was observed that the cellobiose accumulation in the hydrolysis mixtures was higher for the *T. reesei* enzyme compared to *P. janthinellum,* indicating an incomplete digestion due to the inherent low BGL activity of the former (Fig. 2). For acid-pretreated rice straw, there was little or no cellobiose accumulation observed during the 24 h of hydrolysis, in the case of *P. janthinellum* enzyme, while about 5 mg/ml cellobiose was observed consistently from 4th hour onwards in the case of *T. reesei*. In alkali pretreated rice straw hydrolysis, the cellobiose concentration increased from 2.42 mg/ml in 4 h to 7.45 mg/ml in 24 h for *T. reesei* while the maximum cellobiose accumulation in the case of *P. janthinellum* was only 0.76 mg/ml at 12 h after which it again decreased. This indicated the efficient removal of cellobiose, from the reaction medium by *P. janthinellum* enzyme, which could be accounted for by the almost tenfold higher beta-glucosidase activity in the fungus. The results are an indicative of an optimum enzyme cocktail from a single fungus that outperforms the conventional cellulase producer.

Cellulases from *P. janthinellum* perform better than *T. reesei* cellulases in the hydrolysis of pretreated biomass

Both the dilute acid and dilute alkali pretreated rice straw were hydrolyzed better by *P. janthinellum* cellulases compared to enzymes from *T. reesei,* indicated by a significantly higher glucose release (Fig. 3a and b). At 24 h, glucose release by *T. reesei* and *P. janthinellum* cellulases from acid-pretreated rice straw were 12.94 ± 0.8 mg/ml and 17.69 ± 0.47 mg/ml, respectively, the latter showing a 37% higher glucose release. Similar results were observed for alkali-pretreated rice straw, where *P. janthinellum* enzyme released 27.24 ± 0.22 mg/ml of glucose which was 43% higher than the *T. reesei* cellulase. Also, the glucose release for both the acid and alkali pretreated biomasses was higher with *P. janthinellum* cellulase at all the measured time points. Total sugar release from acid-pretreated rice straw, calculated as the sum of the concentrations of glucose, xylose, arabinose and mannose in the hydrolysis mixture, was 17.88 ± 0.5 mg/ml for *T. reesei* enzyme and 22.41 ± 0.23 mg/ml for *P. janthinellum* enzyme. For alkali-pretreated rice straw, the hydrolysate sugar contents were 29.49 ± 0.57 mg/ml and 32.94 ± 0.87 mg/ml, respectively, for *T. reesei* and *P. janthinellum* enzymes (Fig. 3a and b). *P. janthinellum* enzyme released 25% higher glucose and 11% higher total

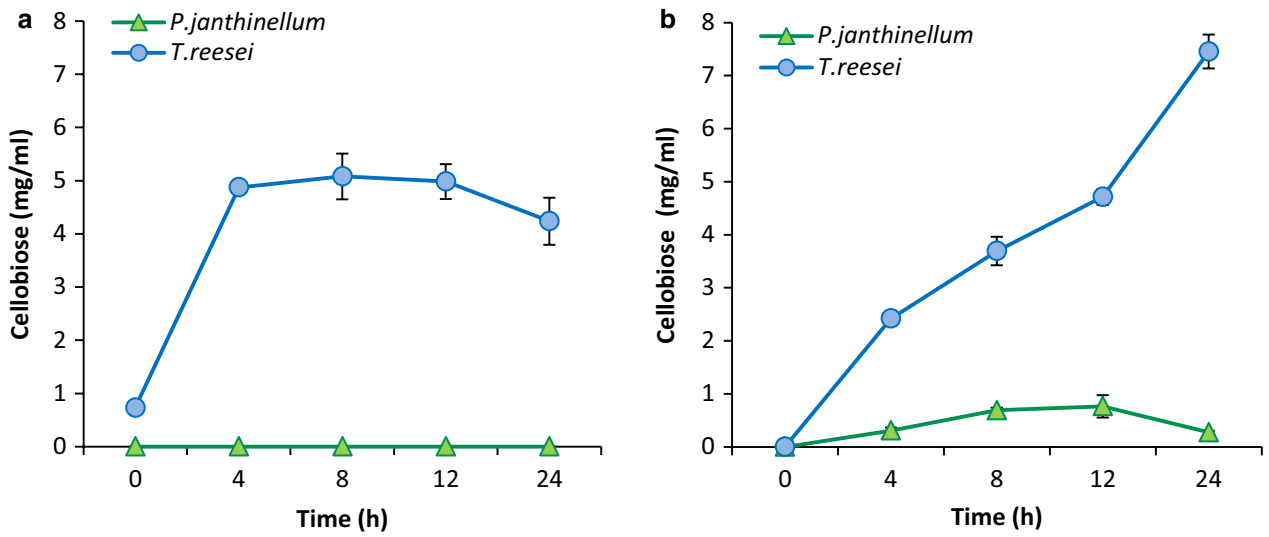


Fig. 2 Cellobiose accumulation during hydrolysis of biomass using *T. reesei* or *P. janthinellum* enzymes. **a** Acid-pretreated rice straw, **b** alkali-pretreated rice straw

sugars from alkali pretreated rice straw sugars compared to *T. reesei* enzyme.

The choice of substrate for the hydrolysis reaction often affects the efficiency of glucose release, and enzyme performance could be different on different biomass substrates. Two different biomasses with significantly different properties, viz. sugar cane bagasse and Eucalyptus leaves pretreated using acid or alkali, were used as substrates for testing hydrolytic efficiency of enzymes from both the fungi. Both the biomasses, regardless of the method of pretreatment were hydrolyzed better by *P. janthinellum* enzyme, indicated by higher glucose release. For acid-pretreated sugarcane bagasse, the glucose release at 24 h was 24.19 ± 0.34 for *P. janthinellum* enzyme and 15.54 ± 0.76 for *T. reesei* enzyme, while in the case of alkali-pretreated biomass, it was 22.16 ± 0.35 and 17.87 ± 0.91, respectively, for *P. janthinellum* and *T. reesei*. Similar results were obtained for eucalyptus leaves, for which the glucose release across treatments were less compared to other biomass types (Fig. 3c and d).

P. janthinellum produces higher enzyme titers compared to *T. reesei*

To know how each of the major components of cellulolytic system contribute to the hydrolytic efficiency of *P. janthinellum* cellulase cocktail, standard cellulase assays were performed, on the enzymes produced by the fungi. Extracellular enzyme production in this case was carried out using the same medium and under identical conditions of growth. Secreted enzymes from both fungi were analyzed for the total cellulase, endoglucanase and beta-glucosidase activities. Both *T. reesei* and *P. janthinellum* showed maximum cellulase activity on the 10th day, but the FPAse activity of *P. janthinellum* (0.83 FPU/ml) was 28% higher than that of *T. reesei* (0.65 FPU/ml) (Fig. 4a). Peak endoglucanase activity of 21.72 IU/ml was shown by *P. janthinellum* on the 12th day, whereas *T. reesei* showed maximum activity (15.55 IU/ml) at 10th day (Fig. 4b). *T. reesei* showed an endoglucanase activity of 12.93 IU/ml even at 6th day, but the levels were raised only upto 15.55 IU/ml on 10th day and were not sustained at further time points, probably indicating a feedback inhibition through glucose accumulation. *P. janthinellum* on the contrary, had a lower initial endoglucanase activity (9.55 IU/ml) which steadily increased to 21.72 IU/ml on 12th day and showed an ascending trend. The largest difference in enzyme activity between the two fungi was observed in the case of beta-glucosidase (BGL) activity. Highest BGL activity in the case of *T. reesei* was 10.15 U/ml. *P. janthinellum* also showed the highest BGL activity (95.42 U/ml) on the 10th day (Fig. 4c). Also, the fungus produced 24.88 U/ml activity on the 2nd day, where *T. reesei* could elaborate only 1.68 U/ml. These results are remarkable as the beta-glucosidase activity at peak levels by the two fungi is different by an almost tenfold margin. Also, it becomes evident that the expression of BGL sets in early in the *P. janthinellum* which would allow it to hydrolyze cellulose faster and prevent cellobiose accumulation, which in turn may help to overcome an early setting in of feedback inhibition. The results were also confirmed by a zymogram analysis which showed a prominent BGL

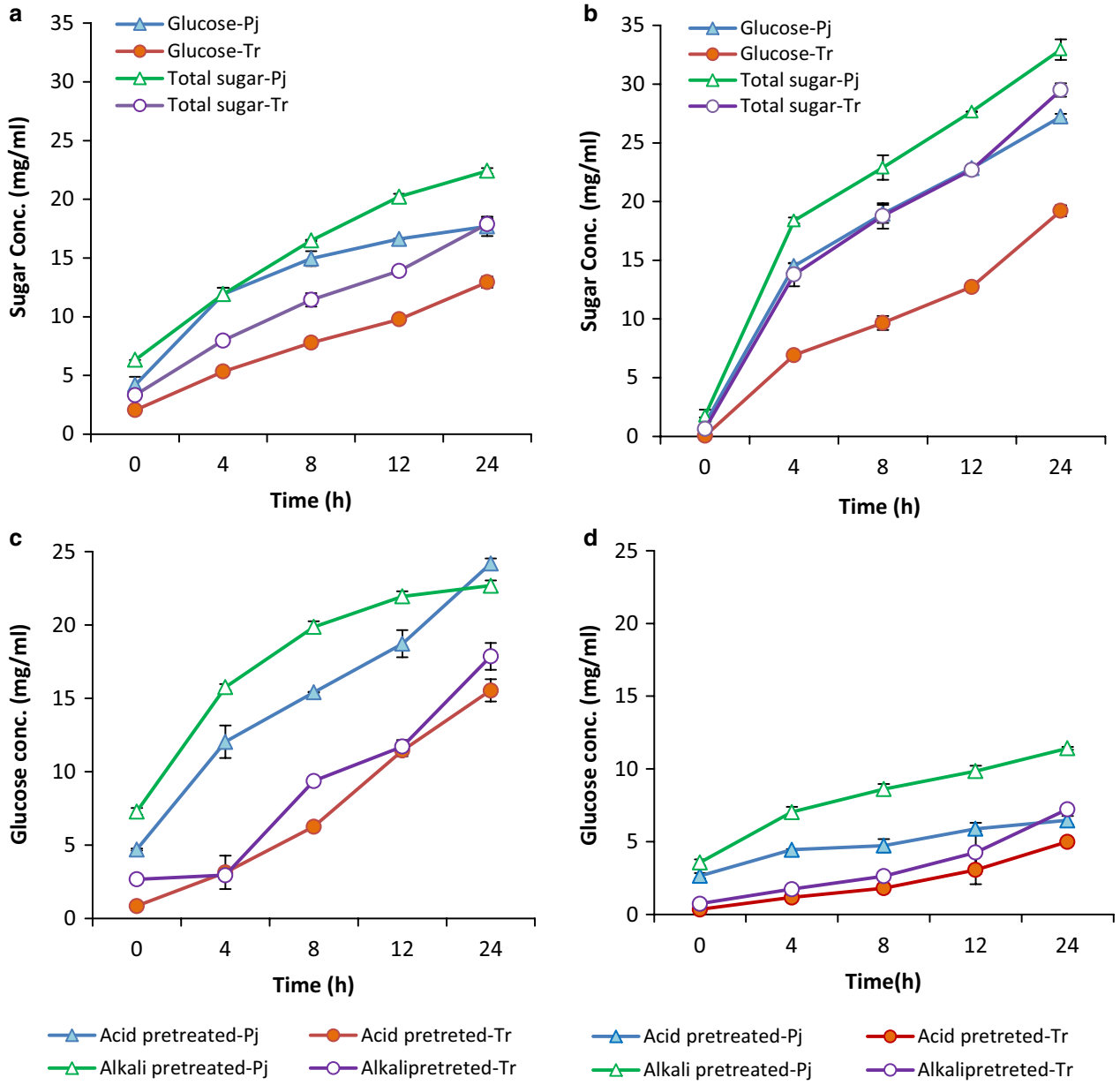


Fig. 3 Sugar release on enzymatic hydrolysis of pretreated biomass by cellulases from *T. reesei* RUT C30 and *P. janthinellum* NCIM 1366. Glucose and total sugar release on enzymatic hydrolysis using *T. reesei* (Tr) and *P. janthinellum* (Pj) cellulases from **a** acid-pretreated rice straw and **b** alkali-pretreated rice straw. Glucose release on hydrolysis using Tr or Pj cellulases from **c** sugarcane bagasse, **d** eucalyptus leaves

activity band in *P. janthinellum*, whereas the *T. reesei* BGL band was barely visible (Fig. 4d).

Comparative secretome analysis of cellulose-induced cultures confirm secretion of a relatively larger number of CAZymes and lignocellulose active enzymes by *P. janthinellum*

The observed cellulase activity and hydrolysis activity are contributed by the extracellular enzymes in both organisms. As both cultures showed maximum filter paper activity on 10th day of inoculation in cellulose medium, it was speculated that the maximum repertoire of enzymes are secreted at that time point. For comparison glucose was selected as the non-inducing carbon source. The secreted proteins from both cultures, either grown with glucose as carbon source or upon induction with cellulose on the 10th day of growth, were identified and quantitatively analyzed

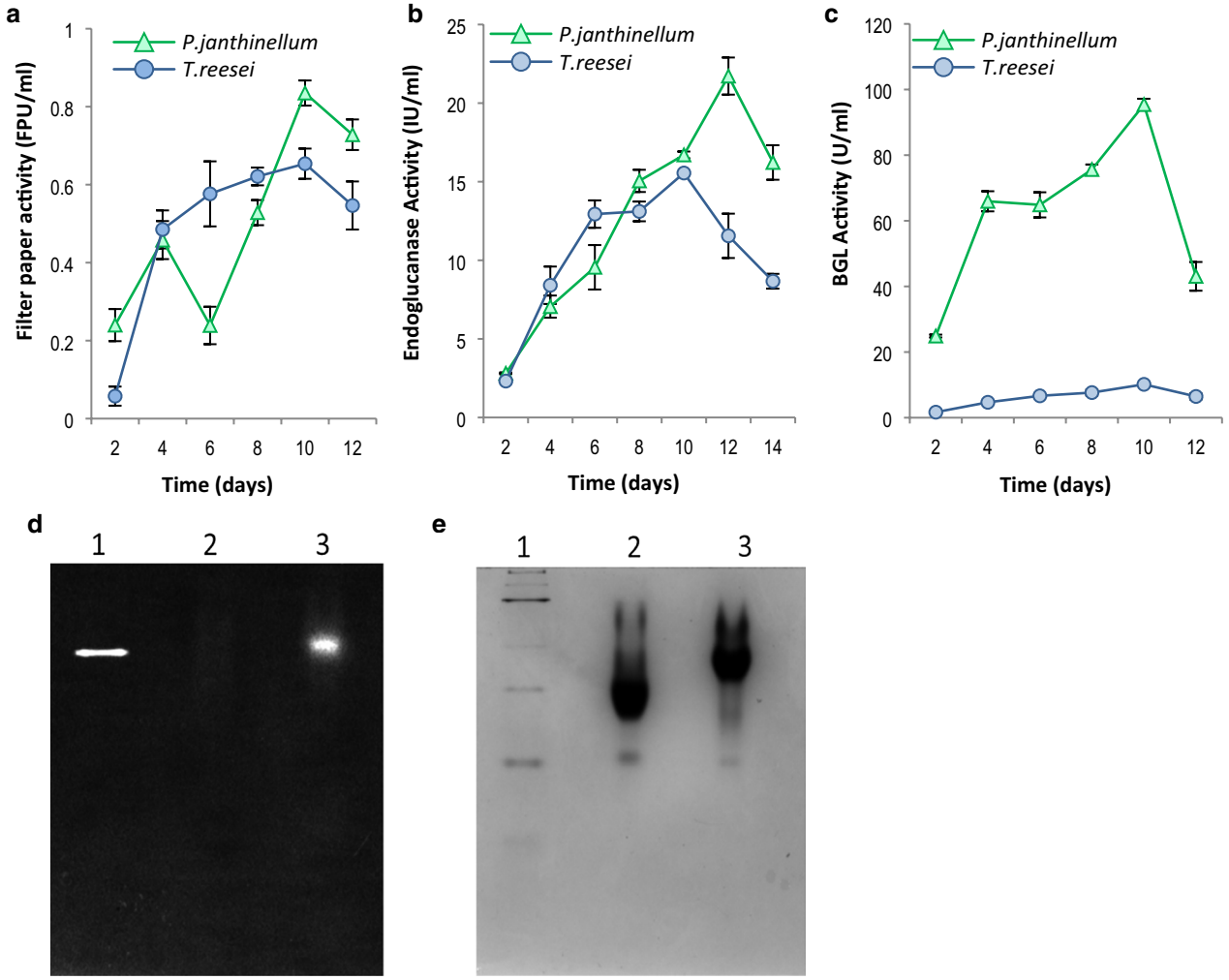


Fig. 4 Extracellular cellulase production by *T. reesei* and *P. janthinellum* in Mandels and Weber medium with cellulose as sole carbon source. **a** Total cellulase activity (filter paper units), **b** endoglucanase activity (CMCase), **c** beta-glucosidase (BGL) activity. **d** Zymogram analysis showing BGL activity as MUG fluorescence. **e** Native PAGE showing protein profile of *T. reesei* and *P. janthinellum* secreted enzymes (Lane1—NativeMark™ unstained protein standard from Invitrogen, Lane2—*T. reesei* enzyme extract, Lane3—*P. janthinellum* enzyme extract for both **d** and **e**

by liquid chromatography tandem mass spectrometry (LC–MS/MS) analysis. Additional file 1: Table S1, and Additional file 2: Table S2 lists all the proteins, their Uniprot accession number, molecular weights, number of unique peptides, normalized abundance in glucose and cellulose grown cultures and fold change of upon induction for both *T. reesei* RUT-C30 and *P. janthinellum* NCIM1366 cultures, respectively. Our analysis detected a total of 53 proteins from *T. reesei* and 85 proteins from *P. janthinellum* in the 10th day secretome. The distribution of proteins according to their biological function is shown in Fig. 5. Among them, 27 proteins from *T. reesei* (51%) and 29 proteins from *P. janthinellum* (34%) were predicted to have an N terminal signal peptide using SignalP5.0 server. The identification of proteins without a signal peptide in the secretome could be indicative of the presence of cell lysis, cell death, or secretion through unconventional mechanisms [17]. Figure 6 shows the top 10 highly expressed proteins, as measured by the normalized abundance of their peptides in cellulose-induced cultures, compared to the control (grown in glucose). Most of the highly expressed proteins from both organisms were directly involved in the lignocellulose degradation.

Since it appeared from the forgoing studies that the reason for better overall cellulolytic activity and hydrolytic efficiency of *P. janthinellum* could be its secretion of a larger number of proteins, most of which are known to be involved in lignocellulose hydrolysis, it was

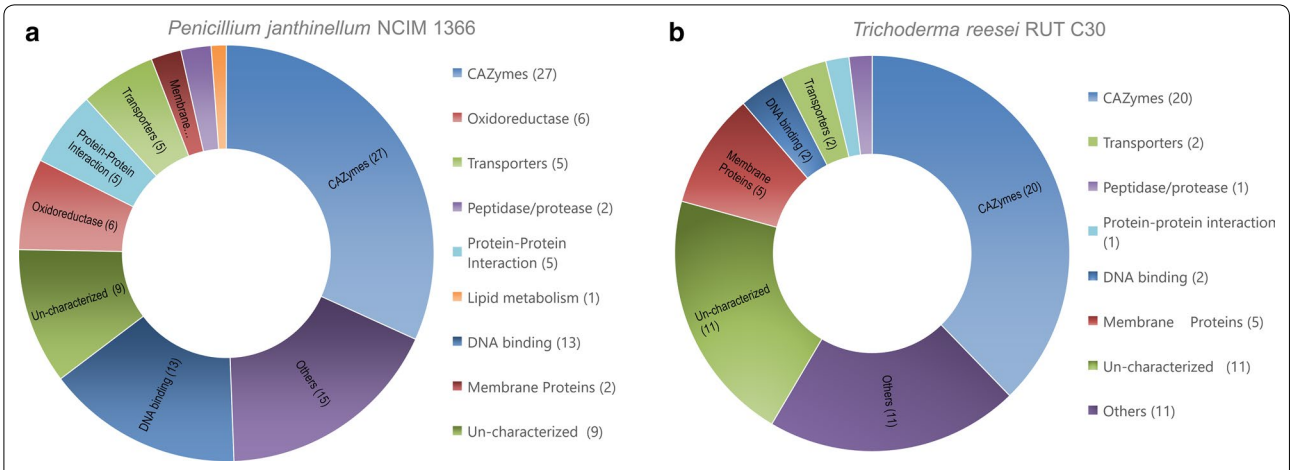


Fig. 5 Functional categories of proteins secreted on cellulose induction by *P. janthinellum* and *T. reesei*. **a** *P. janthinellum*, **b** *T. reesei*

speculated that the organism could elaborate more cellulolytic enzymes and/or accessory proteins compared to the industrial workhorse—*T. reesei*. An analysis of the CAZymes in the total secretomes was performed to understand their distribution in the extracellular proteins of both fungi. Among the 53 secreted proteins detected in *T. reesei*, 20 were identified as CAZymes (Fig. 5a) and the number of CAZymes identified in the 85 identified secreted proteins of *P. Janthinellum* were 27 (Fig. 5b). The distribution of proteins among different CAZy families and the distribution of glycosyl hydrolase (GH) family proteins in both the fungi are shown in Fig. 7. CAZymes from *P. janthinellum* were mostly GH family proteins except one in GT family. The CAZymes from *T. reesei* RUT-C30 were distributed to more CAZy families which included GH (glycoside hydrolases), CE (carbohydrate esterases), AA (auxiliary activities) and CBM (carbohydrate-binding module). In the case of GH family proteins, *P. janthinellum* secretome had almost double the number of different Glycoside Hydrolases compared to *T. reesei*. GH family proteins from *P. janthinellum* spanned over 12 GH subfamilies while for *T. reesei* it was 11 GH subfamilies. GH subfamilies 5, 6, 7 and 11 were detected in both secretomes while GH subfamilies 3, 4, 16, 17, 30 and 72 were detected only in *T. reesei* and GH subfamilies 2, 15, 27, 28, 36, 43, 55 and 75, were detected only in *P. janthinellum*.

Table 1 shows the list of CAZymes identified from the secretomes of *P. janthinellum* and *T. reesei*. Among the CAZymes detected, a total of 17 enzymes which are directly involved in cellulose hydrolysis were detected, of which 3 were common to both fungi, which were cellobiohydrolase1 (CBH1) (Uniprot accession: P62694, A0A088DLG0), cellobiohydrolase2 (CBH2) (P07987, F1CHI2) and endoglucanase 1 (EG-1) (A0A024SNB7, A0A0F7TSC9). The two cellobiohydrolases showed higher abundance in *P. janthinellum* and the endoglucanase showed higher abundance in *T. reesei*. Apart from the EG-1, two other endoglucanases, EG-II (P07982) and EG-V (A0A024S5P6), are identified from *T. reesei.* In *P. janthinellum*, 10 cellobiohydrolases and 2 endoglucanases were identified from the common cellulases. However, it may be noted that multiple peptide tags may be matching the same *P. janthinellum* NCIM 1366 gene sequence in reality, which may not be captured on analyzing against the genome (s) of other *Penicillium* species' genomes as is the case here. While *P. janthinellum* exhibited 10 times the BGL activity of *T. reesei*, no beta-glucosidases were identified in both the fungi. Therefore, there was no way to confirm if the higher BGL activity obtained experimentally for *P. janthinellum* correlates to a higher amount of the corresponding protein in the secretome. It was previously observed that the BGL proteins had high specific activities and minute quantities can give high hydrolytic efficiencies, even though their proteins were undetectable by conventional means.

There were 9 enzymes involved in hemicellulose degradation identified from the secretome of *T. reesei,* while 6 were identified in *P. janthinellum,* of which 2 were common with *T. reesei*. Chitin-degrading enzymes were also identified from both secretomes but pectin-degrading enzymes were identified only in *P. janthinellum*. Accessory activities known to aid cellulose hydrolysis in *T. reesei*, swollenin (A0A024RZP7) and lytic polysaccharide monooxygenase (A0A024SM10)

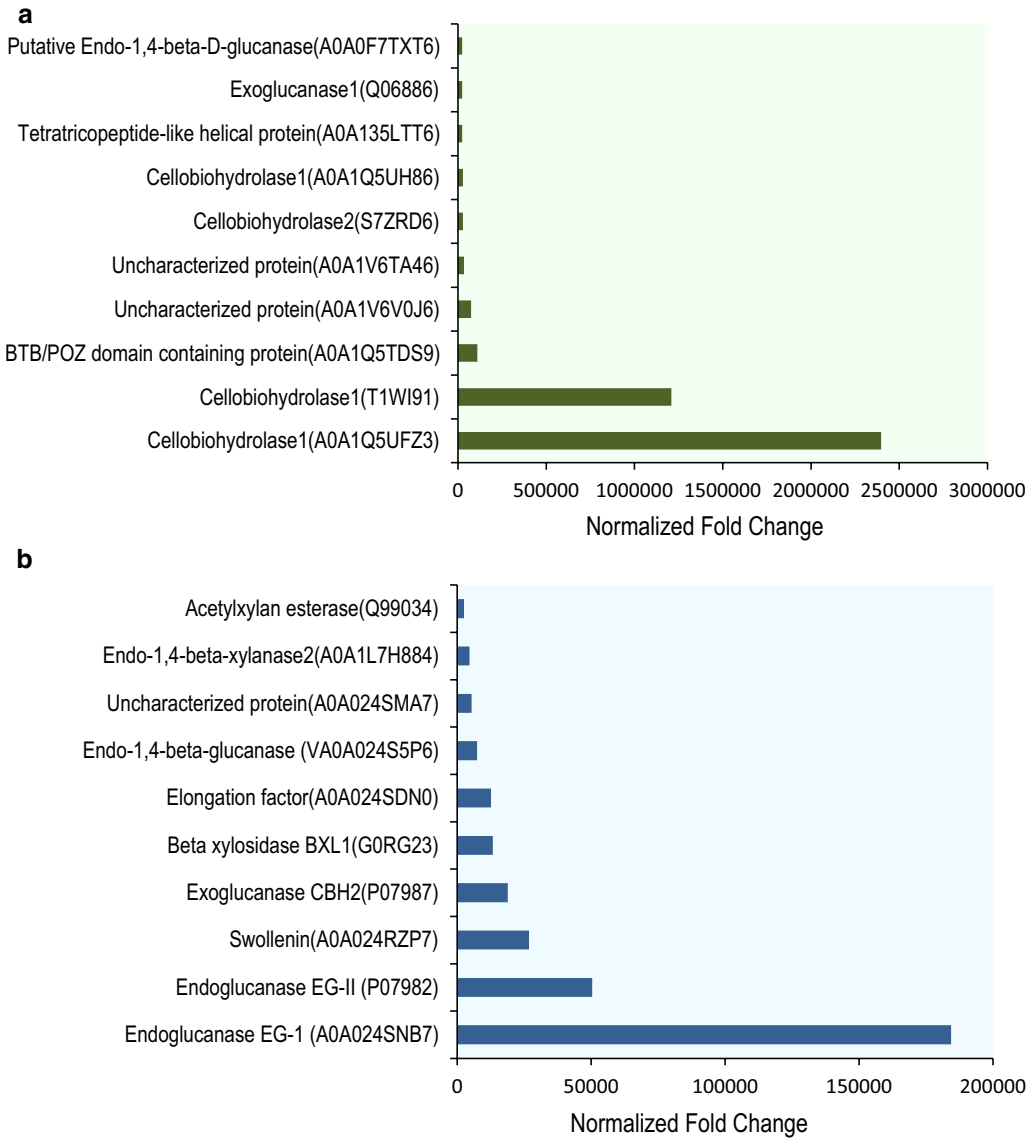


Fig. 6 Most highly expressed proteins in the secretome of *P. janthinellum* and *T. reesei,* on cellulose induction. **a** *P. janthinellum*, **b** *T. reesei*

were identified in the secretome of *T. reesei*, while these activities were not detected in *P. janthinellum*. In general, the relative abundance of most of the CAZymes were high in *P. janthinellum* compared to *T. reesei* and one of the cellobiohydrolases (A0A1Q5UFZ3) from GH7 family showed a very high relative abundance of 2.3 million. Though *P. janthinellum* did not show the accessory enzymes/activities in its secretome, it does not necessarily mean that the fungus lacks them, and further confirmations from the genome analysis is awaited.

Discussion

Lignocellulose-degrading enzymes are critical in biomass conversion to biofuels and filamentous fungi are typically used for the production of these enzymes because of their ability to synthesize and secrete a wide array of plant cell wall-degrading enzymes [18]. *T. reesei* RUT-C30 is the most widely used fungus for cellulase production, despite it having a lower titer of beta-glucosidases and much lesser number of CAZymes compared to certain other fungi. This is primarily due to the fact that *T. reesei* produces the highest known titers of enzymes, the extracellular protein concentrations reaching as high 100 g/L [19]; and there is a wealth of information

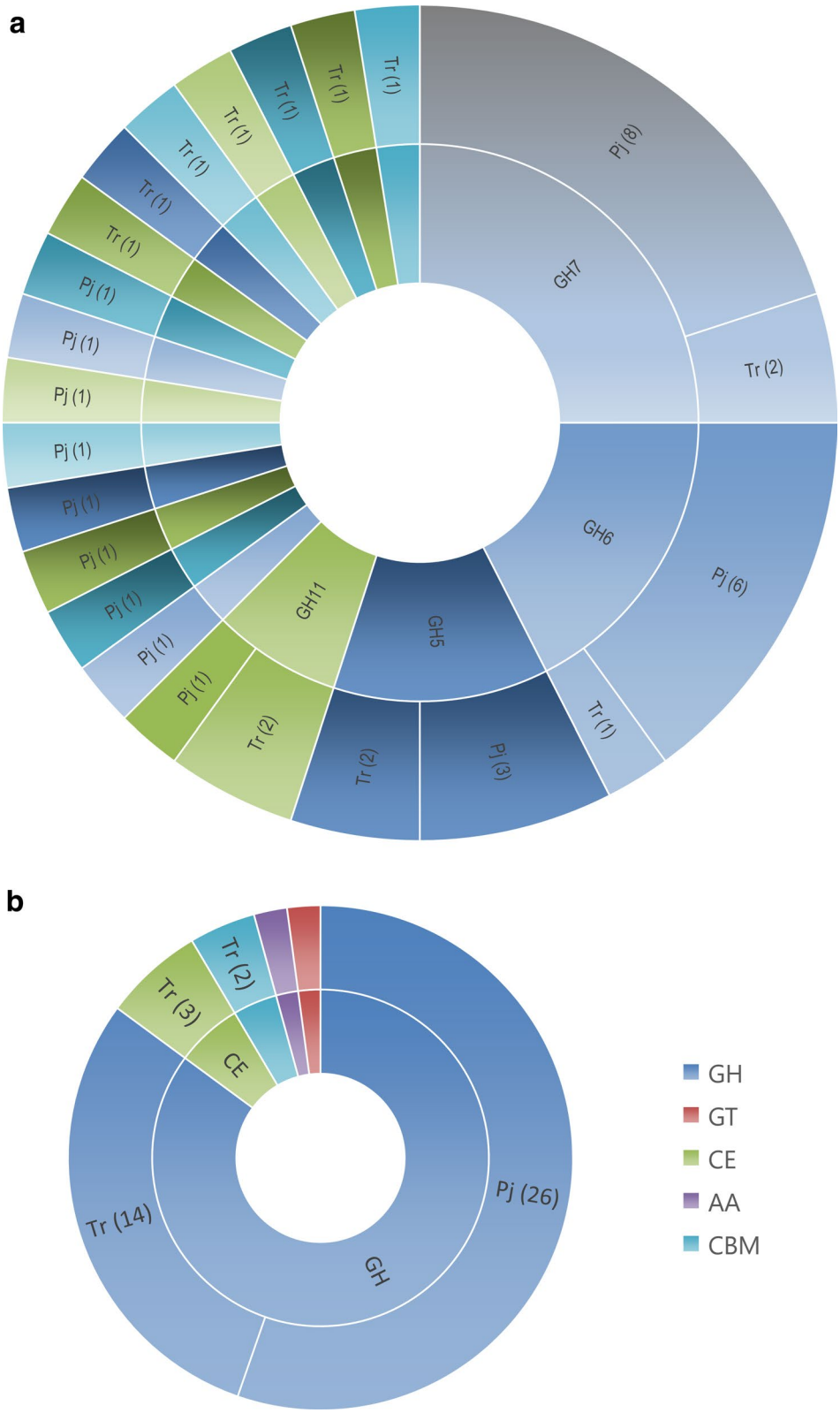


Fig. 7 Distribution of secreted proteins of the fungi in different CAZyme families, and the glycosyl hydrolase family representation in the secretome. **a** Secreted protein distribution in CAZyme families, **b** distribution of glycosyl hydrolases in subfamilies. Tr, *T. reesei;* Pj, *P. janthinellum*

Table 1 CAZymes identified from the secretomes of *P. janthinellum* and *T. reesei*

CAZy family	Enzyme	Uniprot accession(s)	Tr-Glu	Tr-Cel	Tr-FC	Pj-Glu	Pj-Cel	Pj-FC	Presence
Cellulose hydrolysis/metabolism									
G H5	Endoglucanase EG-II	P07982	1.92	96905	50333	–	–	–	Tr
GH5	Endoglucanase B	A0A1Q5UIH7	–	–	–	14.5	1256	86	Pj
GH5	Putative endo-1,4-beta-D-glucanase	A0A0F7TXT6	–	–	–	0	22950	22950	Pj
GH6	Exoglucanase 6A	A0A1Q5UI44	–	–	–	0	11744	11744	Pj
GH6	Exoglucanase CBH2	P07987, F1CHI2	13.6	256595	18844	0.24	4989	20,734	Pj & Tr
GH6	Cellobiohydrolase	A0A1Q5UIA5	–	–	–	0.33	6916	20958	Pj
GH6	Cellobiohydrolase	A0A0F7TT91	–	–	–	–	–	–	Pj
GH6	Cellobiohydrolase	A0A1F5LUH2	–	–	–	–	–	–	Pj
GH7	Exoglucanase CBH1	P62694, A0A088DLG0	925	388093	419	0.13	2607	20054	Pj & Tr
GH7	Glucanase	S7ZRD6	–	–	–	0	28328	28328	Pj
GH7	Cellobiohydrolase1	A0A1Q5UFZ3	–	–	–	0.0036	8742	2396804	Pj
GH7	Cellobiohydrolase1	A0A1Q5UH86	–	–	–	0.36	10266	28516	Pj
GH7	Cellobiohydrolase1	T1WI91	–	–	–	0.00058	711	1225862	Pj
GH7	Exoglucanase1	Q06886	–	–	–	14.6	345207	23487	Pj
GH7	Cellobiohydrolase	A0A0F7TQ87	–	–	–	–	–	–	Pj
GH7	Endoglucanase EG-1	A0A024SNB7, A0A0F7TSC9	0.42	77136	183567	29.5	12605	426	Pj & Tr
GH45	Endo-1,4-beta-glucanase V	A0A024S5P6	3	22586	7407	–	–	–	Tr]
Hemicellulose hydrolysis/metabolism									
GH2	Beta-mannosidase	A0A2H3IHX4	–	–	–	11.8	240	20	Pj
GH3	Beta-xylosidase BXL1	G0RG23	2.8	37715	13278	–	–	–	Tr
GH5	Mannan endo-1_4-beta-mannosidase	Q99036, A0A1Q5SW99	12.8	1387	107	24.4	1891	77	Pj & Tr
GH11	Endo-1,4-beta-xylanase2	A0A1L7H884, A0A088S933	0	4530	4530	20	833	40	Pj & Tr
GH11	Endo-1,4-beta-xylanase1	A0A2H3A3Q5	110	32909	299	–	–	–	Tr
GH17	Beta-1,3-endoglucanase	A0A024SAF4	385	264	1.4	–	–	–	Tr
GH27	Alpha-galactosidase	A0A0F7KIJ2	–	–	–	0	5173	5173	Pj
GH30	Endo-beta-1,4-xylanase4	A0A024RV01	34.5	3052	88	–	–	–	Tr
GH43	Putative exo-beta-1,3-galactanase	S7ZCW4	–	–	–	78.2	1900	24	Pj
GH55	Exo-beta-1,3-glucanase	S8BDR6	–	–	–	117	13.9	8.4	Pj
CE05	Acetylxylan esterase	Q99034	19.5	49211	2522	–	–	–	Tr
CE16	Acetyl esterase	A0A024SFG8	10.8	24711	2273	–	–	–	Tr
CE16	Acetyl esterase	A0A2H3A8G3	18.6	1573	84	–	–	–	Tr
Chitin hydrolysis/metabolism									
GH18	Chitinase2	A0A024S0K1	11.2	597	53	–	–	–	Tr
GH75	Endo-chitosanase	A0A1Q5UKX2	–	–	–	422	1340	3.1	Pj
GH84	Putative bifunctional alpha-glucuronidase/N-acetyl beta-glucosaminidase	S7ZNU5	–	–	–	534	1572	2.9	Pj
Starch hydrolysis/metabolism									
GH15	Glucoamylase	A0A093V3D6	-	–	–	223	0	0.0044	Pj
Pectin									
GH28	Endo-polygalacturonase	A0A1Q5UPK8	–	–	–	56	3631	64	Pj
Accessory activities/other enzymes									
AA9	Polysaccharide monooxygenase CEL61A	A0A024SM10	2.85	2817	987	–	–	–	Tr
CBM1	Swollenin	A0A024RZP7	0.75	20249	26748	–	–	–	Tr
CBM1	CBM1 domain containing protein	A0A024SJK4	104	3661	35	–	–	–	Tr
GH16	Glucanosyl transferase	A0A024S7E6	787	2014	2	–	–	–	Tr
GH36	Putative galactinol sucrose galactosyl transferase	A0A1V6YI05	–	–	–	0	2804	2804	Pj
GH72	1_3-Beta-glucanosyl transferase	A0A024RXP9	7.9	3225	405	–	–	–	Tr
GT31	Glycosyl transferase	B8MG06	–	–	–	20	825	40	Pj

Table 1 (continued)

Tr, *Trichoderma reesei* RUT C30, Pj, *Penicillium janthinellum* NCIM 1366,

Tr-Cel, Pj-Cel, *T. reesei* or *P. janthinellum* grown with cellulose as carbon source (induced),

Tr-Glu, Pj-Glu, *T. reesei* or *P. janthinellum* grown with glucose as carbon source (uninduced)

accumulated on its genetics and gene regulation through works spanning several decades [20]. However, there are still efforts targeted at improving its enzyme production [21, 22] as the cost of cellulases cannot yet be considered as economical for biorefinery operations. *P. janthinellum* NCIM 1366 is a mutant strain developed at the National Chemical Laboratory, Pune, India, through classical mutagenesis and which exhibited enhanced cellulase production compared to the parent strain NCIM 1171 [11]. The extracellular enzyme preparation from the strain was found in this study to be more efficient than the *T. reesei* enzyme in the hydrolysis of pretreated rice straw. The cellulase system of the fungus is relatively unexplored and the present study aimed to study the cellulases from *P. janthinellum* and compare it with established cellulase hyper producer *Trichoderma reesei* RUT-C30.

Experiments were designed to assess the efficiencies of both cellulase preparations for the hydrolysis of rice straw, pretreated using the most common methods of dilute acid or dilute alkali treatment at high temperature. The enzyme from *P. janthinellum* hydrolyzed the pretreated rice straw biomass better, indicated by the higher glucose release. Interestingly, the glucose release from acid and alkali pretreated rice straw was, respectively, 37% and 43% higher compared to *T. reesei*. Total sugar release was also higher for *P. janthinellum* enzyme extract. Similar to rice straw, *P. janthinellum* enzyme released higher amount of glucose from sugarcane bagasse and eucalyptus leaves. The results were surprising for a "new" cellulase producer to outperform the established industrial producer. Hence the extracellular enzyme preparations from both fungi were analyzed for their major component activities. These included endoglucanases (EGs), cellobiohydrolases (CBHs), and β-glucosidases (BGLs), which act in synergism to hydrolyze cellulose [23]. Since the parameters like media composition, pH, carbon source used, etc., can influence the quantity and variety of the cellulase components produced by fungi [24], the basic mineral salts medium of Mandels and Weber [25] with cellulose as the sole carbon source was used for both the organisms to obtain un-biased data. The results showed higher activity for all the three major components with peak activity on 10th day of incubation. Among the enzymes, the largest difference in activity was observed for beta-glucosidase (BGL), which was tenfold higher in *P. janthinellum.* It is already recognized that the extracellular enzymes of *T. reesei* strains are limited in BGL activity for effective biomass hydrolysis [26]. Thus it may be speculated that the higher BGL activity may be one of the significant factors which can contribute to the higher hydrolytic efficiency shown by *P. janthinellum* cellulase. This was also supported by the fact that the hydrolysis using *T. reesei* cellulase accumulated higher concentration of cellobiose in the medium, indicative of an incomplete hydrolysis. Thus, unlike *T. reesei* enzyme which has to be supplemented with external BGL for hydrolysis reaction [27], *P. janthinellum* enzyme preparations may be used without the need for any blending or with minimal addition of synergistic BGL preparation (s). In addition to the major cellulases, the total extracellular proteins were also 1.8 times higher for *P. janthinellum* and the qualitative analysis by SDS-PAGE showed more number of proteins in the gel complementing this finding.

Proteomic approaches have been widely used in filamentous fungi for the identification of both intracellular and extracellular proteins [28]. The genome of *T. reesei* QM6a, which is the parent strain of RUT-C30 was first sequenced in 2008 giving insight into its CAZyme system [29]. *T. reesei* is known to encode at least 10 cellulases, 16 hemicellulases and a total of around 400 CAZymes in its genome. But the composition of secretome varies depending on the carbon source used, culture conditions or experimental parameters. The first proteome analysis of *T. reesei* RUT-C30 identified a total of 22 proteins using lactose as carbon source [30]. Another study, using different carbon sources identified 230 extracellular proteins and 90 CAZymes [31]. In the present study, using a minimal mineral salt medium under identical conditions, a total of 53 proteins were identified from *T. reesei* secretome, while *P. janthinellum* secreted 85 different proteins. As expected, most of the proteins identified from both the fungi were related to biomass degradation. More number of CAZymes was identified from *P. janthinellum* secretome. CAZymes from *T. reesei* included 2 cellobiohydrolases, 3 endoglucanases, 9 hemicellulases and the accessory activities—swollenin and lytic polysaccharide monooxygenase (LPMO). CAZymes from *P. janthinellum* were grouped into 12 cellobiohydrolases, 3 endoglucanases, and 6 hemicellulases. No beta-glucosidases were identified from both secretomes to support the extremely higher beta-glucosidase activity shown by *P. janthinellum*. However, the number of cellobiohydrolases and their relative abundance was very high in *P. janthinellum*. The proteins identified from the secretome may not be a complete representation of all the CAZymes

secreted by the organism, as the study used only a single time point and pure cellulose as sole carbon source. The highest differentially expressed protein from *T. reesei* was the GH7 family endoglucanase EG-1, which showed 183,567-fold increase in expression upon cellulose induction. However, CBH1 is known to be the major secreted protein of *T. reesei* on cellulose induction [32]. The difference in this study might be a result of the culture conditions and/or the time point of analysis. It could also result from the processing of samples where the insoluble cellulose fraction, which could bind the enzyme, was removed to obtain the supernatant used for analyses. The normalized fold difference shown by the most highly expressed protein from *P. janthinellum* was almost 2 million, and this was a cellobiohydrolase from GH7 family. While *T. reesei* secreted a wider variety of enzymes involved in lignocellulose hydrolysis, it was the *P. janthinellum* that secreted more glycosyl hydrolases and especially very high level of exoglucanases. The study provides preliminary information on the presence of all major cellulolytic and hemicellulolytic activities in the fungus and a very high induction in presence of cellulose, which could account for its enhanced hydrolytic performance.

Conclusions

Here, we provide the first ever secretome analysis of *Penicillium janthinellum* NCIM1366 and its comparison with the established cellulase hyper producing industrial strain—*Trichoderma reesei* RUT-C30. The analyses have highlighted the better hydrolytic efficiency, enzyme activity, protein production and secretion efficiency of *P. janthinellum,* which indicates its potential as future industrial cellulase producer. Further exploration and a deeper understanding on the reasons of its better cellulase production warrants genome and transcriptome level studies on the fungus which is progressing and we aim to reveal soon. Further targeted genetic modifications are expected to improve its performance even more, providing a worthy alternative for *T. reesei*, or complement it in the cellulase applications for biomass conversion.

Methods

Microorganism and growth medium

Penicillium janthinellum NCIM1366 was kindly provided by National Culture Collection of Industrial Microorganisms (NCIM), CSIR-National Chemical Laboratory, Pune, India, and *Trichoderma reesei* RUT-C30 culture was a kind gift from Prof George Szakacs, Technical University of Budapest. The cultures were grown and maintained in potato dextrose agar (PDA) slant. For enzyme production, spores were collected from 20-day-old PDA slants of *P. janthinellum* NCIM1366 and 5-day-old slants of *T. reesei* RUT-C30.

Enzyme production

For enzymatic hydrolysis, cellulase enzyme production was carried out under submerged fermentation using optimized media for both organisms. Both media were modified from the original Mandels and Weber medium [25], optimized for growth and cellulase production of the respective organisms. The enzyme production medium for *P. janthinellum* contained (in g/L): KH_2PO_4 (2.0), $CaCl_2.2H_2O$ (0.3), urea (0.3), $MgSO_4.7H_2O$ (0.3), $(NH4)_2SO_4$ (1.4), peptone (0.75), yeast extract (0.25), Tween-80 (0.5) and trace elements: $FeSO_4.7H_2O$ (0.005), $MnSO_4.H_2O$ (0.0016), $ZnSO_4.7H_2O$ (0.0014), and $CoCl_2.6H_2O$ (0.002) with the pH of medium adjusted to 5.5. Cellulose (1% w/v) and wheat bran (2.5% w/v) were used as carbon sources and a spore suspension containing 1×10^5 spores/ml was used as inoculum at 1% (v/v) level. For *T. reesei*, the production medium contained (in g/L): KH_2PO_4 (2.0), $(NH_4)_2HPO_4$ (2.1), yeast extract (2), NaCl (0.5), $CaCl_2.2H_2O$ (0.3), urea (0.3), $MgSO_4.7H_2O$ (0.3), Tween 80 (0.5) and trace elements: $FeSO_4.7H_2O$ (0.005), $MnSO_4.H_2O$ (0.0016), $ZnSO_4.7H_2O$ (0.0014) and $CoCl_2.6H_2O$ (0.002) with pH of the medium adjusted to 7.2. The carbon sources used were 0.1% lactose, 2% cellulose and 1.5% wheat bran and the medium was inoculated at 1% (v/v) level with a 1×10^6 spores/ml spore suspension. Cultivation was carried out at 30 ± 2 °C and 200 rpm agitation. The extracellular crude enzyme from both cultures was collected after 10 days of incubation, and was assayed for total cellulase activity. Enzyme activity was expressed in filter paper units (FPU).

Comparison of total cellulase activity, endoglucanase activity, and beta-glucosidase activity

Both organisms were cultivated in the basic Mandels and Weber medium [25] (composition in g/L: KH_2PO_4 (2.0), $CaCl_2.2H_2O$ (0.3), urea (0.3), $MgSO_4.7H_2O$ (0.3), $(NH4)_2SO_4$ (1.4), peptone (0.75), yeast extract (0.25), Tween-80 (0.5) and trace elements (g/L): $FeSO_4.7H_2O$ (0.005), $MnSO_4.H_2O$ (0.0016), $ZnSO_4.7H_2O$ (0.0014), and $CoCl_2.6H_2O$ (0.002) with pH of the medium adjusted to 5.0. Cellulose (1% w/v) was used as the sole carbon source. The inoculum size used was 1% v/v of a 1×10^5 spores/ml suspension for both the cultures. Cultivation was carried out at 30 ± 2 °C and 200 rpm agitation. Samples were collected from the 2nd day onwards, every 48 h. Total cellulase activity (filter paper activity) and endoglucanase (CMCase) activity was determined by the IUPAC method [33] and beta-glucosidase activity was determined as described by Rajasree et al. [34]. One unit of beta-glucosidase activity is defined as the amount

of enzyme required to liberate 1 µg *p*-nitro phenol from pNPG (4-nitrophenyl β-D-glucopyranoside) per milliliter.

Zymogram analysis of extracellular proteins

Native poly-acrylamide gel electrophoresis (PAGE) of the extracellular enzyme and methyl umbelliferyl β-D-glucopyranoside (MUG) staining was performed as described by Rajasree et al. [34].

Comparison of hydrolytic efficiencies and cellobiose accumulation

Different lignocellulosic biomass, pretreated similarly using either dilute acid or alkali, was used for the hydrolysis studies. For acid pretreatment, biomass (20% w/v) was mixed with 10% w/w of H_2SO_4 and was pretreated for 1 h at 120 ± 2 °C. The biomass was cooled to room temperature and a slurry was made by adding 2 × volume of water. The pH of the slurry was adjusted to 6.0 by adding 10 N NaOH. Solid–liquid separation was performed using a nylon sieve and the biomass was washed twice with tap water. The biomass was used directly after correction of moisture or air dried at room temperature and stored until used. For alkali pretreatment, 20% w/v rice straw and 10% w/w NaOH was mixed and pretreated at 120 ± 2 °C for 1 h. Water (2 × volume) was added and the pH of the pretreated slurry was adjusted to 6 by adding 10 N H_2SO_4. The biomass was then processed as described above.

The hydrolysis reactions (20 ml reaction volume) were carried out in screw-capped glass conical flasks with the following conditions: 10% w/w of dry biomass, 10 FPUs/g enzyme loading and 0.05% w/w surfactant (Tween 80) loading. 0.5% v/v of a commercial penicillin/streptomycin mixture (Himedia, India) was added to prevent any bacterial contamination. The hydrolysis reaction was carried out at 50 °C for 24 h and samples were collected at 0, 4, 8, 12 and 24 h of hydrolysis. The amount of glucose, xylose, arabinose, mannose and cellobiose in the samples was determined using the HPLC as described previously [35].

Comparison of extracellular protein production and secretome analysis

Both organisms were grown in basic Mandels and Weber medium [25] with either glucose or cellulose at 1% (w/v) level as sole carbon sources, as described above. Samples were collected at 48-h intervals starting from the 2nd day. Total secreted proteins present in the samples were estimated by the Bradford method [36] with BSA as standard. SDS-PAGE was carried out as described by Laemmli [37], using 12% acrylamide gels. Proteins were visualized by staining with Coomassie Brilliant Blue R-250.

For secretome analyses, extracellular proteins were collected by centrifugation for 10 min at 4 °C and 13,400 × g after 10 days of incubation. The supernatants were further clarified by filtration through 1µ Glass microfiber filter. The collected proteins were dialyzed against 50 mM ammonium bicarbonate buffer and concentrations were normalized. Samples were digested using trypsin following the standard protocol [38]. The proteomic profiling was performed in duplicates by liquid chromatography tandem mass spectrometry (LC–MS/MS) at the Mass Spectrometry & Proteomics Core facility of Rajiv Gandhi Centre for Biotechnology, Trivandrum, India. The protein samples were subjected to in-solution trypsin digestion using sequence grade trypsin (Sigma Aldrich, India). The LC/MS/MS analyses of the tryptic peptides were performed in a SYNAPT G2 high definition mass spectrometer (Waters, UK), connected to a NanoACQUITY UPLC® chromatographic system (Waters, UK) for the separation of the peptides. The LC–MS/MS acquired raw data were analyzed by Progenesis QI for Proteomics V3.0 (NonLinear Dynamics, Waters, UK) for protein identification using the protein database of *Trichoderma reesei* and *Penicillium* downloaded from UniProt repository. Prediction of the presence of secretion signal motifs was achieved using SignalP 5.0 [39].

Identification of CAZymes from secretome data

Annotation of CAZy (carbohydrate active enzymes) family of proteins as per CAZy database [40] was done through dbCAN2 meta server [41].

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13068-020-01830-9.

Additional file 1. Proteins identified from the induced and uninduced secretome of *T. reesei* RUTC30.

Additional file 2. Proteins identified from the induced and uninduced secretome of *P. janthinellum* NCIM 1366.

Acknowledgements

ASR, MC, PKV and MS are thankful to the Council of Scientific and Industrial Research (CSIR), Govt. of India, for the Junior/Senior research fellowships for Ph.D. This study was partially funded by Department of Biotechnology Govt. of India (BT/PR20695/BBE/117/211/2016) and CSIR (MLP 0035: 33/2018/MD-FTT&FTC-ANB). We thank Prof. George Szakacs, Technical University of Budapest for the *Trichoderma reesei* RUT-C30 strain.

Authors' contributions

ASR—performed most of the experimental work and protein studies. MC, PKV and AA did data analyses and categorization. RKP initiated the original studies on cellulase production from Pj and Tr and established protocols. DVG originally isolated the Pj culture, did classical mutations and also corrected manuscript. MS and AA did biomass pretreatment and hydrolysis of biomass. AP was involved in experiment planning and manuscript corrections. RKS conceptualized the work, planned experiments and work distribution, did

data interpretation and graphing, drafted manuscript and performed corrections and coordinated the entire work. All authors read and approved the final manuscript.

Funding
Primary funding of the work is from the contingencies of ASR, MC, PKV and part of the funding from grants for projects BT/PR20695/BBE/117/211/2016 (DBT, Govt. of India) and MLP 0035: 33/2018/MD-FTT&FTC-ANB (CSIR, Govt. of India).

Availability of data and materials
All essential data generated or analyzed during this study are included in this published article and its supplementary information files. More elaborate datasets generated during and/or analyzed during the current study available from the corresponding author on reasonable request.

Ethics approval and consent to participate
This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication
RKS, the corresponding author, declares on behalf of all authors that they do not have data pertaining to individuals.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Biofuels and Biorefineries Section, Microbial Processes and Technology Division, CSIR-National Institute for Interdisciplinary Science and Technology, Industrial Estate, Pappanamcode, Thiruvananthapuram 695019, India. [2] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India. [3] Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560012, India. [4] National Chemical Laboratory, Pune, India. [5] Department of Chemical Engineering, Hanyang University, Seoul, Republic of Korea. [6] Centre for Innovation and Translational Research, CSIR-Indian Institute of Toxicology Research, Lucknow, India.

Received: 5 August 2020 Accepted: 16 November 2020
Published online: 01 December 2020

References

1. Chen H. Chemical Composition and Structure of Natural Lignocellulose. In: Biotechnology of Lignocellulose. Dordrecht: Springer Netherlands. 2014; pp. 25–71. https://doi.org/10.1007/978-94-007-6898-7_2
2. Van Dyk JS, Pletschke BI. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-Factors affecting enzymes, conversion and synergy. Biotechnol Adv. 2012;30(6):1458–80. https://doi.org/10.1016/j.biotechadv.2012.03.002.
3. Johnson E. Integrated enzyme production lowers the cost of cellulosic ethanol. Biofuels, Bioprod Biorefining. 2016;10(2):164–74. https://doi.org/10.1002/bbb.16344.
4. Glass NL, Schmoll M, Cate JHD, Coradetti S. Plant Cell Wall Deconstruction by Ascomycete Fungi. Annu Rev Microbiol. 2013;67:477–98. https://doi.org/10.1146/annurev-micro-092611-150044.
5. Zhang X-Z, Zhang Y-HP. Cellulases: Characteristics, Sources, Production, and Applications. In: Yang ST, El-Enhasy H, Thongchul N (eds) Bioprocessing Technologies in Biorefinery for Sustainable Production of Fuels, Chemicals, and Polymers, Wiley, Inc., Hoboken: 2013; pp 131–46. https://doi.org/10.1002/9781118642047.ch8
6. Montenecourt BS, Eveleigh DE. Selective Screening Methods for the Isolation of High Yielding Cellulase Mutants of *Trichoderma reesei*. In: Brown RD, Jurasek L (eds) Hydrolysis of Cellulose: Mechanisms of Enzymatic and Acid Catalysis, Advances in Chemistry Vol 181, Chapter 14. 1979; pp 289–301. https://doi.org/10.1021/ba-1979-0181.ch014
7. Foreman PK, Brown D, Dankmeyer L, Dean R, Diener S, Dunn-Coleman NS, et al. Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus T*richoderma reesei*. J Biol Chem. 2003;278(34):31988–97. https://doi.org/10.1074/jbc.m304750200.
8. Ellilä S, Fonseca L, Uchima C, Cota J, Goldman GH, Saloheimo M, et al. Development of a low-cost cellulase production process using *Trichoderma reesei* for Brazilian biorefineries. Biotechnol Biofuels. 2017;10:30. https://doi.org/10.1186/s13068-017-0717-0.
9. Adsul MG, Ghule JE, Singh R, Shaikh H, Bastawde KB, Gokhale DV, et al. Polysaccharides from bagasse: applications in cellulase and xylanase production. Carbohydr Polym. 2004;57:67–72. https://doi.org/10.1016/j.carbpol.2004.04.001.
10. Adsul MG, Ghule JE, Shaikh H, Singh R, Bastawde KB, Gokhale DV, et al. Enzymatic hydrolysis of delignified bagasse polysaccharides. Carbohydr Polym. 2005;62:6–10. https://doi.org/10.1016/j.carbpol.2005.07.010.
11. Adsul MG, Bastawde KB, Varma AJ, Gokhale DV. Strain improvement of *Penicillium janthinellum* NCIM 1171 for increased cellulase production. Bioresour Technol. 2007;98(7):1467–73. https://doi.org/10.1016/j.biortech.2006.02.03612.
12. Adsul MG, Terwadkar AP, Varma AJ, Gokhale DV. Cellulases from penicillium janthinellum mutants: solid-state production and their stability in ionic liquids. BioResources. 2009;4:1670–81. https://doi.org/10.15376/biores.4.4.1670-1681
13. Singhvi MS, Adsul MG, Gokhale DV. Comparative production of cellulases by mutants of *Penicillium janthinellum* NCIM 1171 and its application in hydrolysis of Avicel and cellulose. Bioresour Technol. 2011;102:6569–72. https://doi.org/10.1016/j.biortech.2011.01.014.
14. Singhania RR, Saini JK, Saini R, Adsul M, Mathur A, Gupta R, et al. Bioethanol production from wheat straw via enzymatic route employing penicillium janthinellum cellulases. Bioresour Technol. 2014;169:490–5. https://doi.org/10.1016/j.biortech.2014.07.011.
15. Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, et al. Fungal cellulases. Chem Rev. 2015;115(3):1308–48. https://doi.org/10.1021/cr500351c.
16. Atreya ME, Strobel KL, Clark DS. Alleviating product inhibition in cellulase enzyme Cel7A. Biotechnol Bioeng. 2016;113(2):330–8. https://doi.org/10.1002/bit.25809.
17. Miura N, Ueda M. Evaluation of unconventional protein secretion by *Saccharomyces cerevisiae* and other fungi. Cells. 2018;7(9):128. https://doi.org/10.3390/cells7090128.
18. Su X, Schmitz G, Zhang M, Mackie RI, Cann IKO. Heterologous gene expression in filamentous fungi. In: Gadd GM, Sariaslani E (eds) Advances in applied microbiology, Volume 81, Elsevier, San Diego, USA; 2012, pp 1–61. https://doi.org/10.1016/B978-0-12-394382-8.00001-0
19. Cherry JR, Fidantsef AL. Directed evolution of industrial enzymes: an update. Curr Opin Biotechnol. 2003;14:438–43. https://doi.org/10.1016/s0958-1669(03)00099-5.
20. Bischof RH, Ramoni J, Seiboth B. Cellulases and beyond: the first 70 years of the enzyme producer *Trichoderma reesei*. Microb Cell Fact. 2016;15:1–13. https://doi.org/10.1186/s12934-016-0507-6.
21. Derntl C, MacH RL, MacH-Aigner AR. Fusion transcription factors for strong, constitutive expression of cellulases and xylanases in *Trichoderma reesei*. Biotechnol Biofuels. 2019;12:231. https://doi.org/10.1186/s13068-019-1575-8.
22. Chen Y, Wu C, Fan X, Zhao X, Zhao X, Shen T, et al. Engineering of *Trichoderma reesei* for enhanced degradation of lignocellulosic biomass by truncation of the cellulase activator ACE3. Biotechnol Biofuels. 2020;13:62. https://doi.org/10.1186/s13068-020-01701-3.
23. Zhang YHP, Lynd LR. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. Biotechnol Bioeng. 2004;88(7):97–824. https://doi.org/10.1002/bit.20282.
24. Sukumaran RK, Singhania RR, Pandey A. Microbial cellulases-Production, applications and challenges. J Sci Ind Res. 2005; 832–844. http://nopr.niscair.res.in/handle/123456789/5375
25. Mandels M, Weber J. The production of cellulases, Chap 23. In: Hajny GJ, Reese ET, editors. Cellulases and their applications. advances in chemistry, vol 95. American Chemical Society; 1969. pp 391–414. https://doi.org/10.1021/ba-1969-0095.ch023
26. Saloheimo M, Kuja-Panula J, Ylösmäki E, Ward M, Penttilä M. Enzymatic properties and intracellular localization of the novel *Trichoderma reesei* β-glucosidase BGLII (Cel1A). Appl Environ Microbiol. 2002;68(9):4546–53. https://doi.org/10.1128/aem.68.9.4546-4553.2002.
27. Berlin A, Maximenko V, Gilkes N, Saddler J. Optimization of enzyme complexes for lignocellulose hydrolysis. Biotechnol Bioeng. 2007;97(2):287–96. https://doi.org/10.1002/bit.21238.

28. Fernández RG, Novo JVJ. Proteomic protocols for the study of filamentous fungi. In: Gupta V, Tuohy M, Ayyachamy M, Turner K, O'Donovan A, editors. Laboratory protocols in fungal biology. fungal biology. Springer, New York, NY; 2013. pp. 299–308. https://doi.org/10.1007/978-1-4614-2356-0_24
29. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, et al. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat Biotechnol. 2008;26:553–60. https://doi.org/10.1038/nbt1403.
30. Herpoël-Gimbert I, Margeot A, Dolla A, Jan G, Mollé D, Lignon S, et al. Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. Biotechnol Biofuels. 2008;1:18. https://doi.org/10.1186/1754-6834-1-18.
31. Adav SS, Chao LT, Sze SK. Quantitative secretomic analysis of *Trichoderma reesei* strains reveals enzymatic composition for lignocellulosic biomass degradation. Mol Cell Proteomics. 2012;11(7):M111.012419. https://doi.org/10.1074/mcp.M111.012419.
32. Peterson R, Nevalainen H. *Trichoderma reesei* RUT-C30 - Thirty years of strain improvement. Microbiology. 2012;158(Pt 1):58–68. https://doi.org/10.1099/mic.0.054031-0.
33. Ghose TK. Measurement of cellulase activities. Pure Appl Chem. 1987;59(2):257–68. https://doi.org/10.1351/pac198759020257.
34. Rajasree KP, Mathew GM, Pandey A, Sukumaran RK. Highly glucose tolerant β-glucosidase from *Aspergillus unguis*: NII 08123 for enhanced hydrolysis of biomass. J Ind Microbiol Biotechnol. 2013;40:967–75. https://doi.org/10.1007/s10295-013-1291-5.
35. Christopher M, Mathew AK, Kiran Kumar M, Pandey A, Sukumaran RK. A biorefinery-based approach for the production of ethanol from enzymatically hydrolysed cotton stalks. Bioresour Technol. 2017;242:178–83. https://doi.org/10.1016/j.biortech.2017.03.190.
36. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976;72(1–2):248–54. https://doi.org/10.1016/0003-2697(76)90527-3.
37. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;227:680–5. https://doi.org/10.1038/227680a0.
38. Trauger SA, Webb W, Siuzdak G. Peptide and protein analysis with mass spectrometry. J Spectroscopy. 2002;16:15–28. https://doi.org/10.1155/2002/320152.
39. Armenteros AJJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 50 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37:420–3. https://doi.org/10.1038/s41587-019-0036-z.
40. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:D490–5. https://doi.org/10.1093/nar/gkt1178.
41. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. DbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012;40:W445–51. https://doi.org/10.1093/nar/gks479.

Publisher's Note