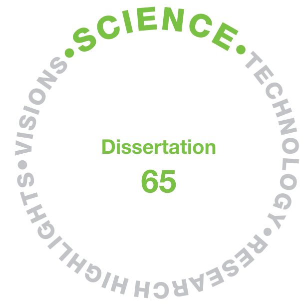
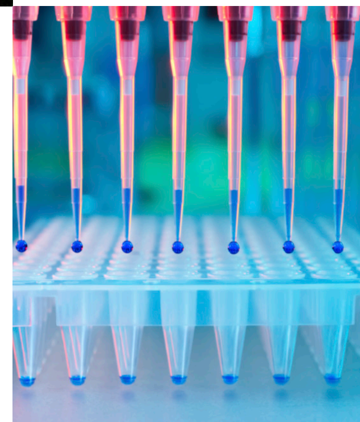


Transcriptional analysis of *Trichoderma reesei* under conditions inducing cellulase and hemicellulase production, and identification of factors influencing protein production

Mari Häkkinen



VTT

VTT SCIENCE 65

Transcriptional analysis of *Trichoderma reesei* under conditions inducing cellulase and hemicellulase production, and identification of factors influencing protein production

Mari Häkkinen

Department of Food and Environmental Sciences
Faculty of Agriculture and Forestry,
University of Helsinki, Finland

Thesis for the degree of Doctor of Science (Agriculture and Forestry) to be presented, with due permission for public examination and criticism in Auditorium 1 of the Viikki Infokeskus Korona (Viikinkaari 11), at the University of Helsinki, on the 29th of October at 12 noon.

ISBN 978-951-38-8161-0 (Soft back ed.)
ISBN 978-951-38-8162-7 (URL: http://www.vtt.fi/publications/index.jsp)

VTT Science 65

ISSN-L 2242-119X
ISSN 2242-119X (Print)
ISSN 2242-1203 (Online)

Copyright © VTT 2014

JULKAISIJA – UTGIVARE – PUBLISHER

VTT
PL 1000 (Tekniikantie 4 A, Espoo)
02044 VTT
Puh. 020 722 111, faksi 020 722 7001

VTT
PB 1000 (Teknikvägen 4 A, Esbo)
FI-02044 VTT
Tfn +358 20 722 111, telefax +358 20 722 7001

VTT Technical Research Centre of Finland
P.O. Box 1000 (Tekniikantie 4 A, Espoo)
FI-02044 VTT, Finland
Tel. +358 20 722 111, fax +358 20 722 7001

Grano Oy, Kuopio 2014

Preface

This study was carried out at the VTT Technical Research Centre of Finland in the Protein production team. The financial support provided by The European Commission within the Sixth Framework Program (NILE, New Improvements for Lignocellulosic Ethanol, Contract No. 019882), Tekes – the Finnish Funding Agency for Innovation (SugarTech, Hydrolysis technology to produce biomass-based sugars for chemical industry raw materials, decision 40282/08), Academy of Finland (The regulatory network of the cellulolytic and hemicellulolytic system of *Trichoderma reesei*, Decision number 133455) and the VTT Biorefinery theme is greatly appreciated. The University of Helsinki is acknowledged for a grant for writing this thesis.

I wish to thank Vice President, Professor Anu Kaukovirta-Norja, former Technology Manager Dr. Tiina Nakari-Setälä and Technology Manager Docent Kirsi-Marja Oksman-Caldentey for the possibility to utilise the excellent working facilities for preparing this thesis. I sincerely thank Team Leader Docent Markku Saloheimo for giving me the opportunity to work in your team and for the guidance you provided throughout the years.

I am indebted to my supervisor Dr. Tiina Pakula for guidance, for sharing your vast knowledge and for teaching me so much about science and how to be a scientist. I truly admire your enthusiasm and devotion to science.

Professor Fred Asiegbu and Dr. Kirk Overmyer are warmly thanked for the thorough pre-examination of the thesis and for their valuable comments. I wish to thank Professor Kaarina Sivonen for actively facilitating the process towards my doctoral defence. Michael Bailey is acknowledged for the careful revision of the English language.

I wish to thank my co-authors, Mikko Arvas, Merja Oja, Nina Aro, Mari Valkonen, Ann-Westerholm Parvinen, Marika Vitikainen, Dhinakaran Sivasiddarthan, Tiina Pakula, Markku Saloheimo and Professor Merja Penttilä, for their contributions to the research work and writing of the manuscripts. Special thanks are addressed to Mikko Arvas for sharing your knowledge on gene annotation and phylogenetic analysis.

I wish to express my deepest gratitude to all my wonderful co-workers for the supporting, friendly and accepting working atmosphere you provided. I feel so privileged to have had the opportunity to work with so many talented, motivated and emphatic people. I want to thank my officemate, Mari for your friendship and

especially for all the mental support you provided throughout the years. Thank you, Aili for sharing your extensive expertise on fungal biotechnology and for being such a friendly and fun co-worker. Thank you Outi for the countless meaningful discussions we had while simultaneously pipetting. I truly appreciate all the friends I was able to make during the years I spent working on my thesis. I want to thank a special group of colleagues and friends including Outi, Maija, Yvonne, Joosu, Mira, Kiira, Pekka, Eero and Dominik for the company during coffee and lunch breaks, for all the support and help and for providing me social life also outside work.

I want to thank all my friends for supporting me and for being there for me for so many years. Thank you for keeping my feet on the ground and for appreciating me as who I am. Finally, my warmest gratitude goes to my family. I want to thank my parents for providing me the good basis for my life and for letting me make my own decisions about my studies and what I wanted to do in life. I thank my sister for her ambitions in life. Your example motivated me to aim high. Thank you, Jyri for making the life outside work so much fun, for all the love and support and for making me realise what are the most important things in life.

Espoo, September 2014

Mari

Academic dissertation

Supervisors	Dr. Tiina Pakula VTT Technical Research Centre of Finland Espoo, Finland
	Docent Markku Saloheimo VTT Technical Research Centre of Finland Espoo, Finland
Reviewers	Professor Fred Asiegbu Department of Forest Sciences University of Helsinki, Finland
	Dr. Kirk Overmyer Department of Biosciences University of Helsinki, Finland
Opponent	Professor David Archer School of Life Sciences University of Nottingham, United Kingdom
Custos	Professor Kaarina Sivonen Department of Food and Environmental Sciences University of Helsinki, Finland

List of publications

This thesis is based on the following original publications which are referred to in the text as I–III. The publications are reproduced with kind permission from the publishers.

I Häkkinen M, Arvas M, Oja M, Aro N, Penttilä M, Saloheimo M, Pakula TM: **Re-annotation of the CAZy genes of *Trichoderma reesei* and transcription in the presence of lignocellulosic substrates**. *Microb Cell Fact* 2012, **11**:134.

II Häkkinen M, Valkonen MJ, Westerholm-Parvinen A, Aro N, Arvas M, Vitikainen M, Penttilä M, Saloheimo M, Pakula TM: **Screening of candidate regulators for cellulase and hemicellulase production in *Trichoderma reesei* and identification of a factor essential for cellulase production**. *Biotechnol Biofuels* 2014, **7**:14.

III Häkkinen M, Sivasiddarthan D, Aro N, Saloheimo M, Pakula TM: **The effects of extracellular pH and of the transcriptional regulator PACI on the transcriptome of *Trichoderma reesei***. Submitted to *Microb Cell Fact* 2014.

Author's contributions

Publication I
Mari Häkkinen carried out fungal cultivations and microarray detection of the expression signals, and participated in the phylogenetic analysis of CAZy genes, annotation of the CAZymes as well as in the analysis and interpretation of the microarray data, and drafted the manuscript.

Publication II
Mari Häkkinen carried out cloning of the genes, participated in the construction and cultivation of the recombinant strains, enzymatic activity measurements and qPCR analysis, carried out fungal cultivations and microarray detection of the expression signals for the second cultivation set and drafted the manuscript.

Publication III
Mari Häkkinen constructed the deletion strain, carried out enzymatic activity measurements and microarray detection of the expression signals, participated in the analysis and interpretation of the microarray data, and drafted the manuscript.

Contents

Preface......3

Academic dissertation......5

List of publications......6

Author's contributions......7

List of abbreviations......10

1. Introduction......13
- 1.1 Breakdown of cellulose and hemicellulose......14
- 1.2 Filamentous fungus *Trichoderma reesei*......17
- 1.3 Carbohydrate active enzyme gene content of the *Trichoderma reesei* genome......19
- 1.4 *Trichoderma reesei* cellulases and hemicellulases......20
  - 1.4.1 Characterized enzymes active against cellulose and hemicellulose......20
  - 1.4.2 Predicted cellulase and hemicellulase genes......22
  - 1.4.3 Secreted cellulases and hemicellulases......23
- 1.5 Regulation mechanisms of *T. reesei* cellulase and hemicellulase gene expression......25
  - 1.5.1 Inducers of cellulase and hemicellulase genes and recognition of the inducing substrate......26
  - 1.5.2 Regulators of cellulase and hemicellulase gene expression......28
    - 1.5.2.1 Activators of cellulase and hemicellulase genes......29
    - 1.5.2.2 Repressors of cellulase and hemicellulase genes......30
    - 1.5.2.3 Enviromental and physiological conditions and novel factors affecting the expression of cellulase and hemicellulase genes......30
- 1.6 Aims of the study......34

2. Materials and methods......36
- 2.1 Strains, media and culture conditions......37
- 2.2 Expression analysis by microarray hybridisation and quantitative PCR......38

2.3 Construction of fungal over-expression and deletion strains 38
2.4 Enzyme activity measurements..39

3. Results and discussion...40
3.1 Identification and re-annotation of the CAZy genes of *T. reesei*..........40
3.2 Phylogenetic analysis of the CAZy genes of *T. reesei*........................43
3.2.1 Functional diversification of cellulases 44
3.2.2 Functional diversification of hemicellulases.............................45
3.3 Genome-wide transcriptional analysis of *T. reesei* genes 46
3.3.1 Impact of ambient pH on cellulase and hemicellulase gene expression 46
3.3.2 Impact of different inducing substrates on the expression of CAZy genes..51
3.4 Screening of candidate regulators for cellulase and hemicellulase genes...59
3.4.1 Preliminary analysis of the effects of the candidate regulators.......60
3.4.2 Genes 80291 and 74765 have an effect on cellulase and hemicellulase gene expression...60
3.4.3 *ace3* gene is essential for cellulase gene expression and for production of cellulase activity 62
3.5 Co-regulated genomic gene clusters 65

4. Conclusions and recommendations...67

References...69

Appendices

Publications I–III

List of abbreviations

AA	auxiliary activity
ABF	arabinofuranosidase
ACE	activator of cellulase expression
AES	acetyl esterase
AGL	alpha-galactosidase
AMP	adenosine monophosphate
AraR	L-arabinose responsive transcriptional activator of *Aspergillus niger*
AreA	global nitrogen metabolism regulator of *Aspergillus nidulans*
AXE	acetyl xylan esterase
BGL	beta-glucosidase
BGLR	beta-glucosidase regulator of *Trichoderma reesei*
BLAST	Basic Local Alignment Search Tool
BLR	blue light regulator
BXL	beta-xylosidase
CAZy	carbohydrate active enzyme
CBH	cellobiohydrolase
CBM	cellulose binding module
cDNA	complementary deoxyribonucleic acid
CE	carbohydrate esterase
ClbR	cellobiose response regulator of *Aspergillus aculeatus*
CLR-1/2	cellulose degradation regulators of *Neurospora crassa*
CREI	carbon catabolite repressor of *Trichoderma reesei*

DNA	deoxyribonucleic acid
DNS	dinitrosalicylic acid
EG	endoglucanase
ENV	envoy protein
FbxA	f-box protein of *Aspergillus nidulans*
Frp	F-box protein required for pathogenicity of *Fusarium oxysporum*
GH	glycoside hydrolase
GLR	glucuronidase
GNA	G-protein alpha
GNB	G-protein beta
GNG	G-protein gamma
GRDI	glucose-ribitol dehydrogenase of *Trichoderma reesei*
GT	glycosyltransferase
HAP	heme activator protein complex
LAE	loss of *aflR* expression
LIMI	LIMPET, E3 ubiquitin ligase of *Trichoderma reesei*
MAN	mannanase
ManR	endo-beta-mannanase regulator of *Aspergillus oryzae*
McmA	MADS box protein of *Aspergillus nidulans*
MFS	major facilitator superfamily
MU	methyl umbelliferone
MUL	4-methyl umbelliferyl-β-D-lactoside
PCP	pentose catabolic pathway
PCR	polymerase chain reaction
PD	potato-dextrose
PHLP	phosducin-like protein
PL	polysaccharide lyase
qPCR	quantitative polymerase chain reaction
RESS	repression under secretion stress
RNA	ribonucleic acid

SPPR	specific protein production rate
SWO	swollenin
W/V	weight/volume
XLN	xylanase
XlnR	xylanase regulator of *Aspergillus niger*
XYN	xylanase
XYRI	xylanase regulator 1 of *Trichoderma reesei*

1. Introduction

Depletion of fossil fuels together with the increasing need for oil based commodities, energy and chemicals has created a demand for alternative energy sources. In addition, the environmental issues related to the use of fossil oil have directed research and industrial applications towards the utilisation of bio-based fuels and chemicals. First generation biofuels are produced primarily from the sugars and starch present in food crops (for example corn and sugar cane) and from vegetable oils (biodiesel production). Thus, the usage of edible plants as a raw material for bio-based products and the dedication of land area for growing these plants have raised questions concerning the environmental and economic sustainability of these first generation biofuels and chemicals.

Lignocellulosic plant material is the most abundant terrestrial renewable resource, and one of its main components, cellulose, is the most abundant polysaccharide in nature. Lignocellulosic material from for example industrial side-streams and by-products of agriculture and forestry can be used for the production of second generation bio-based products which do not compete with the production of food. Lignocellulose is an extremely recalcitrant material and therefore a physical and/or chemical pre-treatment step is needed to break the structure and make it more accessible to enzymes. After the pre-treatment step, the polymers are hydrolysed by enzymes to sugars that are further fermented to valuable products. The pre-treatment steps used in the manufacturing of first generation bio-based products are less harsh and energy-consuming (for example grinding and liquefying), therefore making the production of second generation products more costly as compared to the first generation products. However, the biggest barrier to economically viable commercial production of second generation biofuels is the inefficient conversion of insoluble plant cell wall polysaccharides into fermentable sugars. In first generation processes, enzymes are used for example for saccharification of the glucose polymer, starch, Complex lignocellulosic biomass materials contain different types of polymers and side chains that need to be digested by various different enzymes. Chemical hydrolysis of these polymers is possible but environmentally unsustainable and produces inhibitory by-products. Therefore, in order to optimise enzymatic conversion of the polymers and to decrease the price of enzymes, the enzyme composition needs to be adjusted according to the raw material used and the efficiency of enzyme production needs to be enhanced.

In nature, fungi and bacteria of different species participate in the continuation of the global carbon cycle by degrading plant biomass material for energy and carbon source. Filamentous fungi are especially efficient degraders of plant biomass and hence are the main source of commercial enzymes for lignocellulose degradation. *Trichoderma reesei* is an industrial fungus used widely for the production of homologous and heterologous proteins. Especially the extremely efficient secretion of cellulases and hemicellulases by *T. reesei* is of interest concerning the production of enzymes for various industrial applications including the production of second generation biofuels and other environmentally friendly chemicals from biomass substrates. Although the regulatory network of *T. reesei* starting from substrate recognition and leading to the production of enzymes needed for degradation of the substrate has been widely studied, the precise regulation mechanisms are still under debate. Novel factors important for the regulation of hydrolytic genes of *T. reesei* are believed to exist, but finding these factors can be challenging. Furthermore, it is important to identify all the minor activities produced by the fungus during degradation of complex plant cell wall material in order to optimise the enzyme composition used in commercial degradation of biomass substrates.

1.1 Breakdown of cellulose and hemicellulose

Lignocellulose biomass is composed of the polysaccharides cellulose and hemicellulose together with the polyphenol lignin (Table 1). The breakdown of plant cell wall material in an industrial application begins with a pre-treatment step in order to break down the structure of the material, thereby facilitating the access of enzymes to the cellulose and hemicellulose components of the cell walls. The cellulose chain is a simple linear polymer of β-1,4-linked D-glucose units that are bundled together to form microfibrils (Kolpak & Blackwell 1976) (Figure 1). Two specific structures of cellulose exist; amorphous cellulose is easily accessible by enzymes but in crystalline regions the cellulose chains are tightly packed by hydrogen bonding to prevent the access of water or enzymes.

Hemicelluloses are heterogeneous, branched materials that can be classified as xylans, mannans, xyloglucans, glucomannans and mixed-linkage glucans (for a review, see Scheller and Ulvskov 2010). Classification is based on the main sugar units forming the β-1,4-linked backbone. Xylan, mannan and xyloglucan backbones are built from β-1,4-linked D-xylose, D-mannose and D-glucose units, respectively. The xyloglucan backbone is highly substituted with D-xylose side chains. Mixed-linkage glucans contain glucose units linked by both β-1,3 and β-1,4 linkages and in glucomannan, the backbone consists of both D-glucose and D-mannose units. Hemicelluloses contain different side chains such as D-galactose, D-xylose, L-arabinose, D-glucuronic acid and acetyl groups. The structure of hemicellulose varies greatly between different biomass sources. Arabinoxylans and glucuronoxylans are common in the cell walls of cereals and hardwood, respectively. The most common hemicelluloses present in softwood are mannans (espe-

cially galactoglucomannans). Of the xylans, arabinoglucuronoxylans dominate in softwood. The biological function of hemicellulose is to cross-link the cellulose microfibrils to each other with non-covalent bonds thereby further strengthening the cell wall.

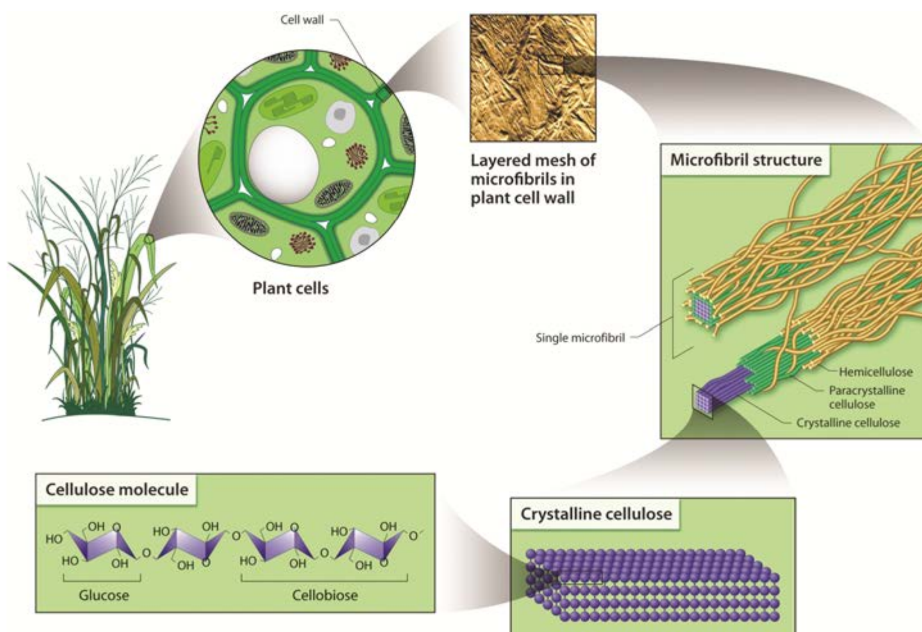


Figure 1. The structure of cellulose in plant cell walls. The picture is reprinted from genomics.energy.gov.

Table 1. Cellulose, hemicellulose and lignin contents of selected biomass materials.

Biomass	Lignin (%)	Cellulose (%)	Hemicellulose (%)
Softwood[a]	27–30	35–40	25–30
Hardwood[a]	20–25	45–50	20–25
Wheat straw[a]	15–20	33–40	20–25
Bagasse[b]	19–24	32–44	27–32

a) (McKendry 2002)
b) (Sánchez 2009)

Complete hydrolysis of cellulose to glucose is achieved by combined and co-ordinated action of several different enzymes (Béguin 1990; Teeri 1997) (Figure 2A). Endoglucanases hydrolyse the amorphous cellulose chain internally, creating more free ends for the cellobiohydrolases. Cellobiohydrolases are also able to

attack the crystalline cellulose and cleave units of two glucose molecules (cellobiose) either from the reducing or non-reducing ends of the cellulose chain. The two *T. reesei* cellobiohydrolases are able to completely degrade ammonia-treated cellulose (Igarashi et al. 2011). The end product of cellobiohydrolase activity, cellobiose, inhibits the activity of cellobiohydrolases (end-product inhibition). Therefore, cellobiose must be hydrolysed efficiently into glucose to ensure continuous cellobiohydrolase activity. In the final step of cellulose hydrolysis, β-glucosidases release glucose molecules from non-reducing β-D-glucosyl residues of disaccharides. Another function of the β-glucosidases is to carry out transglycosylation reactions of cellobiose, resulting in the formation of sophorose (glucosyl-β-1,2-D-glucoside), which is a disaccharide of two β-1,2-linked glucose units (Vaheri et al. 1979). The fungal cellulases are usually made up of two different domains. The N- or C-terminal carbohydrate-binding module (CBM) is connected to the catalytic domain by a linker peptide. The carbohydrate-binding domain enhances the degradation of cellulose by binding to cellulose microfibrils but is not essential for the hydrolysis (Guillén et al. 2010).

A larger repertoire of enzymes is required for the degradation of heterogeneous hemicellulose (for a review, see van den Brink & de Vries 2011) (Figure 2B). Endo-1,4-β-D-xylanases and β-mannanases are needed for hydrolysis of the xylan or mannan backbone, respectively. The degradation products of xylan and mannan backbone polymers are further digested by β-xylosidases and β-1,4-mannosidases, respectively. Different residues forming hemicellulose side chains are cleaved by enzymes such as acetyl xylan esterases, acetyl esterases, α-xylosidases, α-fucosidases, α-galactosidases, α-L-arabinofuranosidases and α-glucuronidases. In addition, a role for β-galactosidases in hemicellulose degradation has been suggested (Ivanova et al. 2013).

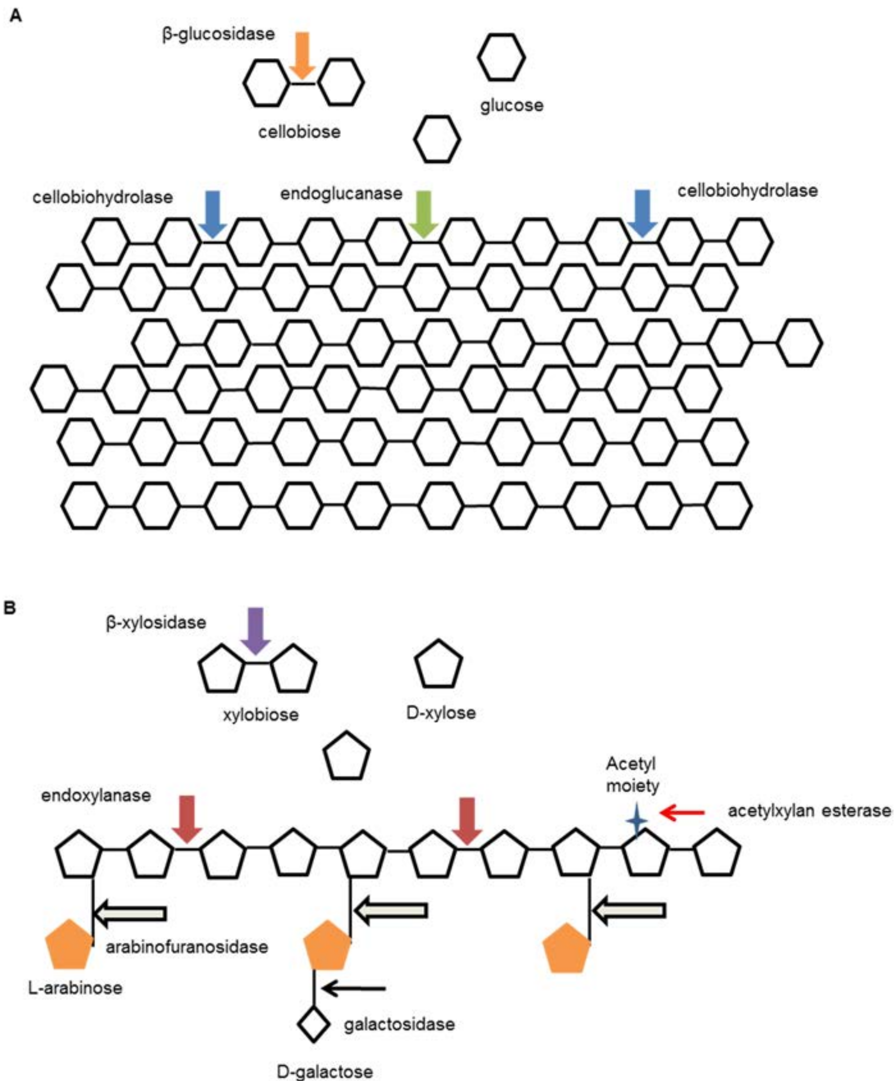


Figure 2. Schematic illustration visualising the breakdown of cellulose (A) and hemicellulose (B) by the synergistic action of several different enzymes.

1.2 Filamentous fungus *Trichoderma reesei*

Fungi of the genus *Trichoderma* are ascomycetes characterized by green spores, repetitively branched conidiophores and adaptation to various ecological environments (Kredics et al. 2014). *Trichoderma reesei* was originally isolated during World War 2 on the Solomon Islands (Reese 1976) and was later identified as an asexual form of *Hypocrea jecorina* (Kuhls et al. 1996). *T. reesei* exhibits a sapro-

trophic lifestyle by degrading lignocellulosic substrates from decaying material. *T. reesei* is an exceptionally efficient producer of especially cellulase and hemicellulase enzymes and is widely used as a host for heterologous protein production (Harkki et al. 1989; Saloheimo & Niku-Paavola 1991; Nyyssönen et al. 1993; Nevalainen et al. 2005). Enzymes produced by *T. reesei* have been traditionally employed in the pulp and paper (Torres et al. 2012), food (Kunamneni et al. 2014), feed (Walsh et al. 1993) and textile (Puranen et al. 2014) industries. Examples of cellulase and hemicellulase applications are presented in Table 2. Currently, the use of enzymes in biorefinery applications is of increasing importance as the research is focusing on designing more efficient enzyme cocktails and on generic improvement of the enzyme production capabilities of the fungus (Kumar et al. 2008). *T. reesei* is an important model organism for the different aspects of cellulase and hemicellulase production. Sequencing of the genome of *T. reesei* enabled gene content comparisons between different fungi and the use of genome-wide methods for studying the protein production properties of *T. reesei* (Martinez et al. 2008). Classical strain improvement methods such as mutagenesis and screening have yielded a large number of mutant strains producing high amounts of cellulases and hemicellulases (Mandels et al. 1971; Montenecourt & Eveleigh 1977a; Montenecourt & Eveleigh 1977b; Durand et al. 1988). The availability of the whole genome sequence and of recombinant DNA techniques has made it possible to use sophisticated molecular biological methods for strain improvement. The industrial strains of *T. reesei* are able to produce more than 100 g/l of extracellular proteins (Cherry & Fidantsef 2003).

Table 2. Applications of cellulases and hemicellulases. The table is adapted from (Kuhad et al. 2011).

Food industry	**Animal feed**	**Textile industry**	**Laundry and detergents**	**Pulp and paper**	**Bio-conversion**
Extraction, clarification and stabilisation of juices	Improvement of nutritive quality	Biostoning of denim	Washing powder additives	Biomechanical pulping	Bioethanol
Extraction of olive oil	Improvement of digestibility	Finishing of textiles	Detergent additives	Modification of fiber properties	Organic acids
Malting of barley		Softening		Deinking	Single cell proteins
Maceration, color extraction and must clarification of wines		Biopolishing		Pulp bleaching	Chemicals
Carotenoid extraction				Drainage improvement	
				Biodegradable cardboard	
				Soft paper	

1.3 Carbohydrate active enzyme gene content of the *Trichoderma reesei* genome

The term “CAZyme” stands for “Carbohydrate-active enzyme” and includes different activities involved in the breakdown, modification and synthesis of glycosidic bonds. The CAZy database compiles enzymes belonging to the CAZy classification covering glycoside hydrolases (GH), carbohydrate esterases (CE), polysaccharide lyases (PL), glycosyltransferases (GT), auxiliary activities (AA) and also enzymes containing a carbohydrate binding module (CBM) (Cantarel et al. 2009, http://www.cazy.org/). Classification is based on amino acid sequence similarities of the catalytic modules supplemented by structural information and experimental evidence. Enzymes involved in degradation of the cellulose and hemicellulose portions of the cell walls are abundant in glycoside hydrolase and carbohydrate esterase families, whereas polysaccharide lyases mainly target pectin. Auxiliary activities is the newest classification covering redox enzymes working alongside with CAZymes (Levasseur et al. 2013). Enzymes of AA family 9 (previously classified as GH61) are involved in the enhancement of lignocellulose degradation (Harris et al. 2010; Langston et al. 2011).

The carbohydrate active enzyme gene content of the *T. reesei* genome was first examined during the initial genome annotation (Martinez et al. 2008). Surprisingly, it was noticed that this efficient degrader of plant biomass does not have an expansion of genes encoding activities towards plant cell wall components. On the contrary, the genome of *T. reesei* contains an unexpectedly low number of glycoside hydrolase and carbohydrate esterase genes, and the number of cellulase and hemicellulase genes is also low as compared to other cellulolytic fungi (Foreman et al. 2003; Martinez et al. 2008). Due to the saprotrophic lifestyle of *T. reesei* it is likely that during speciation it has lost some of the genes that are not needed for the degradation of decaying wood (for example genes necessary for a mycotrophic lifestyle) and that the cellulase and hemicellulase pattern of *T. reesei* has evolved to be sufficient for the efficient degradation of predigested lignocellulose material. It has been suggested that the presence of ligning-degrading basidiomycete fungi signals the presence of pre-digested biomass and hence *T. reesei* had found its natural habitat and ecological niche by following basidiomycetes (Rossman et al. 1999; Druzhinina et al. 2012). Due to the unusually low number of cellulase and hemicellulase genes found from the genome of *T. reesei* it can be speculated that the genome encodes some minor unidentified activities that are vital for the total degradation of complex biomass substrates.

A major observation during initial analysis of the *T. reesei* genome was the non-random distribution of CAZy genes (Martinez et al. 2008). 41% of the CAZy genes were found to localise in 25 regions ranging from 14 kb to 275 kb in length. Among these regions, examples of adjacent co-expressed genes were detected. Co-expression indicates possible common regulatory mechanisms (co-regulation) for the genes. Further analysis of the clustering of CAZy genes by taking into account only those genes that are up-regulated on lactose or cellulose, confirmed that 25%

of the lactose-induced CAZy genes are clustered in the genome and the clusters were predominantly located at the scaffold ends (Kubicek 2013). Location at the scaffold ends indicates that clustering of CAZy genes might be a result of rearrangements that have led to evolutionary benefit. Furthermore, the location of the clusters in non-syntenic blocks of the genome further supports the theory that the co-localisation of the genes has given the fungus a competitive advantage in its natural environment. Non-syntenic blocks are regions where gene order and content is not conserved between closely related species.

1.4 *Trichoderma reesei* cellulases and hemicellulases

As noted in Section 1.1, the coordinated action of several different enzymes is needed for the degradation of plant cell wall material to mono- and oligosaccharides that can be assimilated by the fungus. Several *T. reesei* cellulases and hemicellulases have been characterized in detail. In addition, many genes possibly encoding for hydrolytic enzymes have been identified from the genome by for example cDNA sequencing and on the basis of conserved domains and sequence homology with enzymes from other fungi (Foreman et al. 2003; Martinez et al. 2008). The actual function of these genes remains to be elucidated but they are believed to include novel enzymes active against plant cell wall material and possibly able to enhance the process of biomass degradation.

1.4.1 Characterized enzymes active against cellulose and hemicellulose

Characterized cellulases of *T. reesei* are found from glycoside hydrolase families 1, 3, 5, 6, 7, 12 and 45 and hemicellulases from carbohydrate esterase families 5 and 16 and glycoside hydrolase families 3, 5, 10, 11, 27, 30, 36, 54, 67 and 74 (Table 3). The main cellulases secreted by *T. reesei* include two exo-acting cellobiohydrolases from families GH7 and GH6 (CBHI/CEL7A and CBHII/CEL6A) (Teeri et al. 1983; Shoemaker et al. 1983; Teeri et al. 1987; Mong Chen et al. 1987) and two endo-acting cellulases from families GH7 and GH5 (EGI/CEL7B and EGII/CEL5A) (Penttilä et al. 1986; Okada et al. 1998). Two additional endoglucanases from families GH12 and GH45 have been characterized (EGIII/CEL12A and EGV/CEL45A) (Saloheimo et al. 1988; Saloheimo et al. 1994). The xyloglucanase CEL74A (Grishutin et al. 2004) was originally annotated as an endoglucanase (Foreman et al. 2003). Similarly, EGIV/CEL61A belongs to the family GH61 containing genes previously mistaken as endoglucanases (Saloheimo et al. 1997; Foreman et al. 2003) and later annotated as putative copper-dependent polysaccharide monooxygenases (Harris et al. 2010; Langston et al. 2011).

The amount of secreted β-glucosidases is often a rate-limiting factor in the hydrolysis of lignocellulose biomass (Sternberg et al. 1977). Two *T. reesei* GH1 β-glucosidases (BGLII/CEL1A and CEL1B) and a GH3 β-glucosidase (BGLI/CEL3A) have been characterized (Barnett et al. 1991; Fowler & Brown 1992; Takashima et al. 1999; Saloheimo, Kuja-Panula, et al. 2002; Foreman et al. 2003; Zhou et al.

2012). Of these, BGLI is the main secreted β-glucosidase and BGLII and CEL1B are intracellular enzymes.

For the degradation of hemicellulose backbone (mainly xylan or mannan) the genome of *T. reesei* encodes four characterized xylanases from families GH10, GH11 and GH30 (XYNIII, XYNI, XYNII and XYNIV) (Tenkanen et al. 1992; Torronen et al. 1992; Xu et al. 1998; Saloheimo et al. 2003) and one GH5 β-mannanase (MANI) (Stalbrand et al. 1995). In addition, several enzymes are needed for cutting the various side groups from the hemicellulose backbone. These include a CE5 acetyl xylan esterase (AXEI) (Margolles-Clark, Tenkanen, Söderlund, et al. 1996), a GH67 α-glucuronidase (GLRI) (Margolles-Clark, Saloheimo, Siika-aho, et al. 1996), a GH54 α-L-arabinofuranosidase (ABFI) (Margolles-Clark, Tenkanen, Nakari-Setälä, et al. 1996), two GH27 and one GH36 α-galactosidases (AGLI, AGLII and AGLIII) (Zeilinger et al. 1993; Margolles-Clark, Tenkanen, Luonteri, et al. 1996) and an acetyl esterase (AESI) from family CE16 (Li et al. 2008). The only functionally characterized β-xylosidase (BXLI) of *T. reesei* is needed for the digestion of oligosaccharides derived from xylan (Margolles-Clark, Tenkanen, Nakari-Setälä, et al. 1996). In addition to GH61 enzymes, few other accessory enzymes have been identified. Glucuronoyl esterase CIPII cleaves ester linkages between lignin and hemicellulose, facilitating the removal of lignin (Foreman et al. 2003; Li et al. 2007; Pokkuluri et al. 2011). Swollenin (SWOI) resembles plant expansins and disrupts crystalline cellulose structure without hydrolytic activity (Saloheimo, Paloheimo, et al. 2002). The effect of SWOI in assisting cellulose degradation probably results from disruption of the hydrogen bonding between cellulose fibrils, thereby making the fibres more accessible for the cellulases.

Table 3. Characterized cellulases, hemicellulases and accessory enzymes.

Gene ID	Name	CAZy family	Annotation
123283	abf1	GH54	α-L-arabinofuranosidase
121418	aes1	CE16	Acetyl esterase
72632	agl1	GH27	α-galactosidase
124016	agl2	GH36	α-galactosidase
72704	agl3	GH27	α-galactosidase
73632	axe1	CE5	Acetyl xylan esterase
80240	bga1	GH35	β-galactosidase
76672	bgl1/cel3a	GH3	β-glucosidase
120749	bgl2/cel1a	GH1	β-glucosidase
121127	bxl1	GH3	β-xylosidase
123989	cbh1/cel7a	GH7	Cellobiohydrolase
72567	cbh2/cel6a	GH6	Cellobiohydrolase
49081	cel74a	GH74	Xyloglucanase
123940	cip2	CE15	Glucuronoyl esterase
122081	egl1/cel7b	GH7	Endo-β-1,4-glucanase
120312	egl2/cel5a	GH5	Endo-β-1,4-glucanase
123232	egl3/cel12a	GH12	Endo-β-1,4-glucanase
73643	egl4/cel61a	GH61	Candidate copper-dependent polysaccharide mono-oxygenase/endo-β-1,4-glucanase
49976	egl5/cel45a	GH45	Endo-β-1,4-glucanase
72526	glr1	GH67	α-Glucuronidase
56996	man1	GH5	β-Mannanase
123992	swo1	CBM1	Swollenin
74223	xyn1	GH11	Endo-β-1,4-xylanase
123818	xyn2	GH11	Endo-β-1,4-xylanase
120229	xyn3	GH10	Endo-β-1,4-xylanase
111849	xyn4	GH30	Endo-β-1,4-xylanase

Gene IDs are as in *T. reesei* database version 2.0 (http://genome.jgi-psf.org/Trire2/Trire2.home.html)

1.4.2 Predicted cellulase and hemicellulase genes

In order to achieve an economically viable conversion of complex biomass materials it is vital to identify all the possible activities produced by the fungus against the components of plant cell walls. After the initial sequencing and annotation of the *T. reesei* genome, several studies have attempted to identify novel genes encoding for hydrolytic enzymes active against cellulose and hemicellulose.

Foreman et al. (2003) used cDNA sequencing to identify several candidate β-glucosidases (CEL3B, CEL3C, CEL3D and CEL3E), a candidate membrane-bound endoglucanase (CEL5B), a candidate acetyl xylan esterase (AXEII), a candidate arabinofuranosidase (ABFII) and a candidate family GH61 protein (CEL61B, initially annotated as an endoglucanase). A cellulose binding domain and a signal sequence containing protein CIPI was also identified by Foreman et al. (2003), and is a putative accessory enzyme enhancing cellulose degradation.

Another putative GH54 arabinofuranosidase named ABFIII, a candidate CE5 acetyl xylan esterase and a protein from family GH61 were detected from the secretome of *T. reesei* during a proteomics study (Herpoël-Gimbert et al. 2008). Candidate β-glucosidase genes named *bgl3i*, *bgl3j* and *bgl3f* were identified in a genome-wide analysis together with a candidate β-xylosidase gene, *xyl3b* (Ouyang et al. 2006). A candidate GH27 α-galactosidase was found to be expressed during conidiation (Metz et al. 2011). The genome of *T. reesei* also encodes a putative fifth xylanase, XYNV, from family GH11 (Metz et al. 2011; Herold et al. 2013).

A table listing all the candidate cellulolytic and hemicellulolytic genes of *T. reesei* can be found from the results section.

1.4.3 Secreted cellulases and hemicellulases

Proteomic analysis of different *T. reesei* strains cultivated on media promoting cellulase and hemicellulase gene expression has shed light on the protein pattern produced by the fungus. One of the first secretome studies conducted after the sequencing of the *T. reesei* genome was able to identify 22 proteins secreted by the strain CL847 on lactose-xylose medium, most of which are potentially involved in biomass degradation (Herpoël-Gimbert et al. 2008). As expected, the cellobiohydrolases CBHI and CBHII were the most abundantly secreted enzymes under these conditions. Subsequent proteomic studies have identified additional enzymes secreted on cellulase- and hemicellulase-inducing conditions (Adav et al. 2011; Jun et al. 2011; Adav et al. 2012; Saloheimo & Pakula 2012; Jun et al. 2013; Marx et al. 2013; dos Santos Castro et al. 2014). The greatest number of different enzymes was detected when *T. reesei* was cultivated on complex lignocellulosic biomass (Adav et al. 2012; Marx et al. 2013). Table 4 lists the cellulolytic and hemicellulolytic enzymes detected from the secretome of different *T. reesei* strains grown on various inducing substrates. However, the presence of a protein in the cultivation medium does not necessarily indicate that the protein is secreted by the fungus. Especially at later time points of cultivation, cell lysis might result in leaking of intracellular proteins. Nevertheless, proteomic studies are important assets for example when analysing the results of transcriptional profiling. Furthermore, some of the proteins produced might not be detectable with the proteomics methods used due to low production levels or production only under specific growth conditions. In total 747 proteins encoded by the genome of *T. reesei* were predicted to be secreted according to the presence of a signal sequence that

directs the protein to the endoplasmic reticulum for post-translational processing and further for secretion (Druzhinina et al. 2012). Proteins destined for intracellular locations or for plasma membrane were distinguished from the secreted proteins by using computational criteria to predict the subcellular localization of the proteins and the possible transmembrane helices present in the proteins (Druzhinina et al. 2012).

Table 4. Secreted cellulolytic and hemicellulolytic enzymes identified in different proteomic studies.

Gene ID	Name	CAZy family	Annotation
123989	CBHI	GH7	cellobiohydrolase
72567	CBHII	GH6	cellobiohydrolase
76672	BGLI	GH3	β-glucosidase
121735	CEL3B	GH3	candidate β-glucosidase
82227	CEL3C	GH3	candidate β-glucosidase
47268	BGL3I	GH3	candidate β-glucosidase
104797	BGL3J	GH3	candidate β-glucosidase
122081	EGI	GH7	endo-β-1,4-glucanase
120312	EGII	GH5	endo-β-1,4-glucanase
123232	EGIII	GH12	endo-β-1,4-glucanase
49976	EGV	GH45	endo-β-1,4-glucanase
82616	CEL5B	GH5	candidate membrane-bound endoglucanase
49081	CEL74A	GH74	xyloglucanase
80240	BGAI	GH35	β-galactosidase
123283	ABFI	GH54	α-L-arabinofuranosidase
76210	ABFII	GH62	candidate α-L-arabinofuranosidase
55319	ABFIII	GH54	candidate α-L-arabinofuranosidase
121127	BXLI	GH3	β-xylosidase
58450	XYL3B	GH3	candidate β-xylosidase
74223	XYNI	GH11	endo-β-1,4-xylanase
123818	XYNII	GH11	endo-β-1,4-xylanase
120229	XYNIII	GH10	endo-β-1,4-xylanase
111849	XYNIV	GH30	endo-β-1,4-xylanase
112392	XYNV	GH11	candidate endo-β-1,4-xylanase
69276		GH30	candidate endo-β-1,4-xylanase

56996	MANI	GH5	β-mannanase
73632	AXEI	CE5	acetyl xylan esterase
44214	AXEII	CE5	candidate acetyl xylan esterase
54219		CE5	candidate acetyl xylan esterase
72632	AGLI	GH27	α-galactosidase
124016	AGLII	GH36	α-galactosidase
72704	AGLIII	GH27	α-galactosidase
55999		GH27	candidate α-galactosidase
27259		GH27	candidate α-galactosidase
69944		GH31	candidate α-xylosidase/α-glucosidase
72526	GLRI	GH67	α-glucuronidase
71394		GH79	candidate β-glucuronidase
106575		GH79	candidate β-glucuronidase
121418	AESI	CE16	Acetyl esterase
103825		CE16	candidate acetyl esterase
62166		GH2	candidate β-mannosidase
5836		GH2	candidate β-mannosidase
69245		GH2	candidate β-mannosidase
58802		GH95	candidate α-fucosidase
123992	SWOI	CBM1	swollenin
73638	CIPI	CBM	cellulose binding domain protein
123940	CIPII	CE15	glucuronoyl esterase
27554		GH61	candidate polysaccharide mono-oxygenase
31447		GH61	candidate polysaccharide mono-oxygenase
73643	CEL61A	GH61	candidate polysaccharide mono-oxygenase/endoglucanase
120961	CEL61B	GH61	candidate polysaccharide mono-oxygenase

Proteins were identified in the studies of (Herpoël-Gimbert et al. 2008; Adav et al. 2011; Jun et al. 2011; Adav et al. 2012; Saloheimo & Pakula 2012; Jun et al. 2013; Marx et al. 2013; dos Santos Castro et al. 2014).

1.5 Regulation mechanisms of *T. reesei* cellulase and hemicellulase gene expression

Production of extracellular enzymes by the fungus is an energy-consuming process and therefore both inducing and repressing mechanisms have evolved to ensure the economical production of enzymes. The various genes encoding en-

zymes needed for the degradation of plant cell wall material are activated only in the presence of an inducing substrate. In addition, expression of the genes is repressed in the presence of easily metabolized carbon sources (for example glucose), that are preferred over plant biomass. This mechanism is called carbon catabolite repression. The addition of a repressing carbon source to induced cultures overrides the induction, resulting in down-regulation of cellulase gene expression (el-Gogary et al. 1989; Ilmen et al. 1997). Several inducing substrates have been identified for cellulase and hemicellulase genes of *T. reesei*. However, the majority of studies have focused on individual, simple substrates or purified polymers, and the impact of more complex biomass substrates on the gene expression patterns has received less attention. In addition to carbon source, several other environmental and physiological factors affect protein production by the fungus and many transcription factors specifically regulating cellulase and hemicellulase gene expression have been characterized. Various filamentous fungi utilise partially different sets of regulatory factors, indicating that different strategies are used for the regulation of hydrolase genes. Novel regulatory factors identified from *T. reesei* and from other fungi indicate that the regulatory network of enzyme production for plant biomass degradation is complex and also includes hitherto unidentified regulatory mechanisms and factors.

1.5.1 Inducers of cellulase and hemicellulase genes and recognition of the inducing substrate

The inducing substrates for cellulase genes include for example the direct (cellobiose) and indirect (sophorose) degradation products of the natural substrate cellulose (Mandels et al. 1962; Sternberg & Mandels 1979; Fritscher et al. 1990; Ilmen et al. 1997). Sophorose is formed from cellobiose by a transglycosylation reaction performed by β-glucosidases. Sophorose has been considered to be the natural inducer of *T. reesei* cellulases. However the observation that the absence of three β-glucosidase genes of *T. reesei* did not abolish induction in the presence of cellobiose but rather enhanced it, indicates that sophorose might not be the natural inducer of cellulase genes, although on crystalline cellulose, the absence of β-glucosidases delayed cellulase gene expression (Fowler & Brown 1992; Mach et al. 1995; Zhou et al. 2012).

Other substrates inducing cellulase gene expression include xylans, lactose, L-arabitol, L-sorbose and xylobiose (Ilmen et al. 1997; Margolles-Clark et al. 1997; Nogawa et al. 2001; Verbeke et al. 2009). Many of the hemicellulase genes studied are induced with cellulose, xylans, xylobiose, L-arabinose, L-arabitol and sophorose (Zeilinger et al. 1996; Margolles-Clark et al. 1997; Akel et al. 2009; Mach-Aigner et al. 2011; Herold et al. 2013). Low concentrations of the xylan degradation product, D-xylose, induce xylanase gene expression whereas high concentrations have a repressing effect (Mach-Aigner et al. 2010; Herold et al. 2013). Lactose (1,4-*O*-β-D-galactopyranosyl-D-glucose) is a carbohydrate predominantly present in dairy products and therefore is not a natural substrate of

T. reesei. However it is an economically feasible soluble carbon source for the production of cellulases and hemicellulases by the industry. The induction mechanism of lactose is not fully understood but it has been suggested to be mediated via the metabolic pathway for galactose utilisation (Seiboth et al. 2004; Seiboth et al. 2005; Seiboth et al. 2007). *T. reesei* produces at least one extracellular β-galactosidase (BGAI) that is able to hydrolyze lactose into galactose and glucose (Seiboth et al. 2005; Gamauf et al. 2007). The further catabolism of D-galactose could therefore produce the necessary inducer molecule. However, over-expression of the *bga1* gene has a negative effect on cellulase induction by lactose, indicating that the uptake of lactose is an important step in the induction process (Seiboth et al. 2005). Lactose is also a stronger inducer of *T. reesei* cellulases than D-galactose (Karaffa et al. 2006). However, no intracellular β-galactosidase has been detected that could be involved in the catabolism of lactose.

The synergistic action of several different enzymes is needed for the complete degradation of lignocellulose biomass, which has led to coordinated regulation of the main cellulase genes (Fowler & Brown 1992; Ilmen et al. 1997; Foreman et al. 2003; Verbeke et al. 2009). Especially in the case of the hemicellulase genes, specific regulation mechanisms depending on the inducing carbon source are believed to exist (Zeilinger et al. 1996; Margolles-Clark et al. 1997). For example, the xylanase genes are differentially activated by various inducers (Zeilinger et al. 1996; Xu et al. 2000; Herold et al. 2013). A question that still remains is how the fungus is able to initially sense the polymeric, insoluble substrate in order to initiate a signalling cascade leading to induced expression of genes needed for assimilation of the substrate. The mechanisms of substrate recognition are not thoroughly understood. However, three different hypotheses have been suggested.

There are indications that low amounts of enzymes such as CBHI and CBHII are formed constitutively under non-inducing conditions (el-Gogary et al. 1989; Carle-Urioste et al. 1997). Addition of antibodies against the main cellulases and a β-glucosidase inhibited cellulase gene expression (el-Gogary et al. 1989). These constitutively expressed enzymes could initiate the degradation of the substrate, thereby releasing small amounts of inducing components that are able to enter the cell and further induce gene expression. Furthermore, the membrane-bound candidate endoglucanase CEL5B has been suggested to be involved in the substrate recognition due to its low basal expression level in the absence of an inducing substrate (Foreman et al. 2003).

The second hypothesis links substrate recognition to the conidiation of *T. reesei*. Several enzymes active against plant polymers have been detected on the surface of conidia (Kubicek 1987; Messner et al. 1991). Therefore, the conidial-bound cellulases could have a role in releasing the inducer from a polymeric substrate. In the study of (Metz et al. 2011), a whole-genome oligonucleotide array was utilised to identify transcripts that are significantly regulated during conidium formation. Genes encoding carbohydrate active enzymes were shown to be enriched among the genes up-regulated during conidiation. These genes included several cellulase and hemicellulase genes that were up-regulated during the early phase of sporulation. XYRI was shown to control the sporulation-associated cellu-

lase gene transcription in the absence of an inducing substrate. On a cellulosic substrate the conidia-located cellulase genes were vital for rapid germination, indicating a role in substrate recognition.

The third possible mechanism for substrate sensing involves conditions after complete consumption of the easily metabolizable carbon source, during which cellulase genes are transcribed (Ilmen et al. 1997). The mechanisms behind this phenomenom are not completely understood. The transcription of cellulases in the absence of an inducer is not due to the lack of carbon source but might be caused for example by carbohydrates released from the fungal cell wall or by an inducer formed from glucose (Sternberg & Mandels 1979; Ilmen et al. 1997). The low amounts of cellulases produced during starvation could act as scouts searching for carbon sources, and in the presence of plant biomass would produce the inducing molecule subsequently activating the cellulase machinery.

1.5.2 Regulators of cellulase and hemicellulase gene expression

After the presence of an inducing substrate has been detected via an inducer molecule, an intracellular signalling process leads to the activation of transcriptional regulators, transporters and metabolic enzymes needed for activation of the genes encoding hydrolytic enzymes and for assimilation of the carbon source. Five transcription factors regulating the expression of cellulase and hemicellulase genes of *T. reesei* have been studied in detail (Table 5). These include three positively acting factors and two negatively acting factors. In addition, other less studied regulators are also known. After the discovery of the first activator of cellulase and hemicellulase gene transcription, XlnR from *Aspergillus niger* (van Peij et al. 1998), several novel factors involved in the expression of these genes have been identified from *T. reesei* and from other filamentous fungi.

Table 5. Characterized regulators of *T. reesei* cellulase and hemicellulase genes.

Factor	**Function**	***T. reesei* gene ID**	**Orthologue in model fungus**
XYRI	General activator of cellulase and hemicellulase genes	122208	XlnR of *Aspergillus* spp.
ACEI	Repressor of the main cellulase and xylanase genes	75418	-
ACEII	Activator of the main cellulase genes and a xylanase gene	78445	-
HAP2/3/5	Induction of the *cbh2* promoter	124286/121080/ 124301	HAPB/C/E of *Aspergillus* spp.
CREI	Regulator of carbon catabolite repression	120117	CREA of *Aspergillus* spp.

1.5.2.1 Activators of cellulase and hemicellulase genes

Of the activators of *T. reesei* cellulase and hemicellulase genes, xylanase regulator 1 (XYRI), a zinc binuclear cluster protein, has been most extensively characterized and is considered to be the general activator of cellulase and hemicellulase gene expression regardless of the inducer used (Stricker et al. 2006; Stricker et al. 2007). XYRI is an orthologue of *Aspergillus niger* XlnR, which was the first positively acting transcriptional regulator of cellulase and hemicellulase genes isolated from a filamentous fungus (van Peij et al. 1998). XYRI binds an inverted repeat of a GGCTAA-motif and a single $GGC(A/T)_3$ motif on promoters of the genes under its regulation (Rauscher et al. 2006; Furukawa et al. 2009). The importance of the *xyr1* gene for the expression of cellulase and hemicellulase genes was confirmed by constructing a knock-out strain. Functional *xyr1* gene was found to be essential for the transcriptional regulation of several cellulase and hemicellulase genes and for the production of cellulase and xylanase activity (Stricker et al. 2006; Stricker et al. 2007). The expression of the main cellulase genes, *cbh1* and *cbh2*, strictly follows the transcript levels of *xyr1*, in contrast to the expression of xylanase genes, which is not directly dependent on the amount of XYRI but possibly also involves other mechanisms (Derntl et al. 2013). In contrast to *T. reesei*, the XlnR orthologue of *Neurospora crassa* and *Fusarium* species is more specific for genes involved in xylan utilization and is not essential for cellulase gene expression (Brunner et al. 2007; Calero-Nieto et al. 2007; Sun et al. 2012).

Ace2 gene (activator of cellulase expression 2) encodes a zinc binuclear cluster DNA-binding protein that activates expression of the main cellulase genes and a xylanase gene. ACEII binds the same promoter motif as XYRI (GGSTAA) by phosphorylation and dimerization, and therefore might be involved in fine tuning the effect of XYRI (Aro et al. 2001; Wurleitner et al. 2003; Stricker et al. 2008). Deletion of *ace2* gene reduced the expression of *cbh1*, *cbh2*, *egl1*, *egl2* and *xyn2* but did not abolish it, most likely because XYRI is still able to up-regulate these genes. ACEII has been suggested to be involved in maintaining a constitutive transcriptional level and in antagonising early induction of *xyn2* (Stricker et al. 2008). The growth medium also affected the result of *ace2* deletion; expression was reduced when cellulose was used as the sole carbon source but not when sophorose was used as an inducer, indicating that the sophorose signal needed for the induction of hydrolase genes is not mediated by ACEII (Aro et al. 2001).

Binding of the trimeric HAP2/3/5 complex (heme activator protein complex) to the CCAAT box located adjacent to the XYRI and ACEII binding site is vital for induction of the *cbh2* promoter (Zeilinger et al. 1998; Zeilinger et al. 2001).The complex has also been suggested to regulate the *xyn1* and *xyn2* genes (Zeilinger et al. 1996). The mechanism of regulation of gene expression by the complex is believed to involve the formation of an open chromatin structure, but the precise regulation mechanism is not known.

1.5.2.2 Repressors of cellulase and hemicellulase genes

T. reesei ace1 gene encodes a Cys_2-His_2 transcription factor that was initially isolated as an activator of the *cbh1* promoter (Saloheimo et al. 2000). In later studies, ACEI was suggested to work as a repressor due to increased expression of the main cellulase and xylanase genes in an *ace1* defective strain grown on sophorose or cellulose (Aro et al. 2003). ACEI has also been shown to compete with XYRI for the binding site (Rauscher et al. 2006) and to repress *xyr1* gene expression (Mach-Aigner et al. 2008). More evidence on the antagonistic function of ACEI towards XYRI was gained by combining the constitutive expression of *xyr1* under a strong promoter and the down-regulation of *ace1*, which led to improved production of cellulase and xylanase activity by *T. reesei* Rut-C30 grown on cellulose (S. Wang et al. 2013).

As mentioned above, cellulose- and hemicellulose-degrading enzymes are produced only when a more easily metabolizable carbon source, such as glucose, is absent. This carbon source dependent regulation (carbon catabolite repression) is mediated by a negatively acting Cys_2-His_2 type transcription factor CREI, which binds to two adjacent SYGGRG motifs on the promoters of its target genes (Ilmén et al. 1996). The deletion of *cre1* results in derepression of cellulase and hemicellulase genes in the presence of glucose and in enhanced production of these enzymes under inducing conditions (Nakari-Setälä et al. 2009). CREI is an orthologue of CreA from *Aspergillus* spp (Dowzer & Kelly 1989; Dowzer & Kelly 1991). CreA functions via a double-lock mechanism repressing expression of the transcriptional activator XlnR and also expression of the genes under XlnR regulation (Tamayo et al. 2008). Accordingly, *xyr1* of *T. reesei* is also believed to be under CREI regulation (Mach-Aigner et al. 2008). However, on lactose cultures the full induction of *xyr1* appeared to be dependent on a functional CREI (Portnoy et al. 2011). Induction of *ace2* was also suggested to be partially dependent on CREI, whereas *ace1* is carbon catabolite repressed (Portnoy et al. 2011). In *A. nidulans*, additional proteins have been shown to be involved in the carbon catabolite repression. CreB is a de-ubiquitinating enzyme that is stabilised by a WD40-repeat protein CreC (Todd et al. 2000; Lockington & Kelly 2001; Lockington & Kelly 2002). CREII of *T. reesei* is orthologous to CreB and has been shown to affect the production of cellulases (Denton & Kelly 2011). Disruption of the *cre2* gene resulted in elevated cellulase activity on sophorose, lactose and cellulose cultures.

1.5.2.3 Enviromental and physiological conditions and novel factors affecting the expression of cellulase and hemicellulase genes

In addition to the presence of an inducing carbon source, several other environmental signals, intracellular metabolism and the physiological state of the fungal cell participate in the modulation of protein production. An increasing number of factors and conditions affecting cellulase and hemicellulase production have been discovered in recent years. Examples of novel regulatory factors from different fungi affecting the production of cellulases and/or hemicellulases are presented in

Table 6. Cellulose degradation regulators 1 and 2 (CLR-1 and CLR-2) of *Neurospora crassa* are required for the induction of the major cellulase genes and some of the major hemicellulase genes on cellulose medium, whereas on xylan, the XlnR orthologue of *N. crassa* is the dominant activator of hemicellulase genes (Coradetti et al. 2012). On cellobiose culture, a *T. reesei* transcription factor BGLR (beta-glucosidase regulator) up-regulates specific β-glucosidase genes, resulting in formation of glucose and subsequently in carbon catabolite repression (Nitta et al. 2012). Accordingly, deletion of the *bglr* gene results in elevated cellulase levels on cellobiose cultures. The only SRF-MADS box protein (for a review, see Messenguy & Dubois 2003) encoded by the genome of *A. nidulans* (McmA) was suggested to mediate the cellobiose induction of two endoglucanase genes and one cellobiohydrolase gene (Yamakawa et al. 2013). Cellobiose response regulator ClbR of *Aspergillus aculeatus* induces genes that are not under XlnR regulation in response to cellobiose and cellulose and also XlnR-dependent genes in response to cellulose (Kunitake et al. 2013).

F-box proteins are involved in the ubiquitination of proteins that are subsequently degraded in the proteasome (for a review, see Jonkers & Rep 2009a). The involvement of these proteins in the regulation of plant cell wall degrading enzymes has been studied in *Aspergillus* (FbxA) and *Fusarium* (F-box protein required for pathogenicity, Frp1) (Duyvesteijn et al. 2005; Jonkers et al. 2009; Jonkers & Rep 2009b; Colabardini et al. 2012). ManR of *A. oryzae* initially identified as a regulator of mannanolytic genes was later shown to control positively at least three cellobiohydrolase genes, one endoglucanase gene and one β-glucosidase gene (Ogawa et al. 2012; Ogawa et al. 2013). A gene encoding a putative glucose-ribitol dehydrogenase named GRDI was found from a screen aiming at identifying *T. reesei* genes specific for sophorose induction of cellulase genes. GRDI was shown to have a positive influence on cellulase gene expression and on extracellular cellulase activity (Schuster et al. 2011).

In *Aspergillus* spp., chromatin level regulation of secondary metabolism gene clusters is known to be under the global regulator, putative methyltransferase LaeA (Bok & Keller 2004; Reyes-Dominguez et al. 2010). LaeA is believed to counteract histone H3 lysine 9 trimethylation, known to lead to a transcriptionally silent chromatin structure (Reyes-Dominguez et al. 2010). LAEI of *T. reesei* is an orthologue of the *A. nidulans* LaeA (Seiboth et al. 2012). A high-density oligonucleotide microarray method was utilised to reveal the targets and regulation mechanisms of LAEI (Karimi-Aghcheh et al. 2013). A large proportion of the genes down-regulated in *lae1* deletion strain and up-regulated in *lae1* over-expression strain were glycoside hydrolases. However, histone methylation patterns studied were not affected by *lae1* modifications, indicating that in *T. reesei* the effect of LAEI might not be mediated by direct histone methylation. The effects of changes in *lae1* expression on secondary metabolism genes were also low suggesting that regulation of secondary metabolism biosynthesis is not the main function of LAEI.

Table 6. Novel candidate regulators for cellulase and hemicellulase genes of different fungi.

Factor	Putative function in cellulase and/or hemicellulase gene regulation	Organism
CLR-1	Cellobiose-mediated activation of *clr-1* gene leads to activation of the *clr-2* gene together with β-glucosidase genes and transporter genes that are important for the utilisation of cellobiose	*Neurospora crassa*
CLR-2	Activation of the cellulose regulon	*Neurospora crassa*
BGLR	Activation of β-glucosidase genes	*Trichoderma reesei*
McmA	Mediates cellobiose induction by binding to a promoter region different from the XlnR binding site	*Aspergillus nidulans*
ClbR	Induces both XlnR- dependent and -independent genes	*Aspergillus aculeatus*
Frp1	Cooperation with CRE1 to inhibit constitutive carbon catabolite repression	*Fusarium oxysporum*
FbxA	Necessary for the full expression of xylanolytic genes and of the regulator gene *xlnR*	*Aspergillus nidulans*
ManR	Controls positively the expression of cellulolytic genes coordinately with XlnR	*Aspergillus oryzae*
GRDI	Controls positively the expression of cellulase genes	*Trichoderma reesei*
LAEI	Unclear, essential for the formation of cellulases and hemicellulases	*Trichoderma reesei*

The metabolism of carbon and nitrogen sources by a fungus has been demonstrated to be linked to the regulation of cellulase and hemicellulase genes through specific transcription factors. XYRI of *T. reesei* was shown to have a role in D-xylose and lactose metabolism (Stricker et al. 2006; Stricker et al. 2007). In *Aspergilli*, the L-arabinose responsive transcriptional activator, AraR, regulates genes involved in releasing L-arabinose from hemicellulose as well as the metabolism of the sugars by the pentose catabolic pathway (PCP) (Witteveen et al. 1989; Battaglia, Visser, et al. 2011; Battaglia, Hansen, et al. 2011). AraR regulates the PCP together with XlnR (de Groot 2003; de Groot et al. 2007). AreA is a global nitrogen metabolism regulator (Arst & Cove 1973). Constitutive expression of the *areA* gene of *A. nidulans* resulted in elevated production of cellulase activity, whereas a loss-of-function mutant of the *areA* gene caused reduced cellulase production (Lockington et al. 2002). In addition, sulphur metabolism was shown to be linked to cellulase gene expression of *T. reesei* via a candidate sulphur regulator LIMI (E3 ubiquitin ligase) (Gremel et al. 2008).

Recently, more attention has been given to the role of sugar permeases in the induction of cellulase and hemicellulase genes. Soluble inducer molecules created from the polymer need to enter the cell in order to start a signalling cascade leading to the up-regulation of genes necessary for degradation of the polymeric substrate. Involvement of a disaccharide permease in the induction of cellulase genes by cellobiose and sophorose was demonstrated already over two decades ago (Kubicek et al. 1993). At low concentration the uptake of these disaccharides is favoured, leading to induction of cellulase genes, whereas at high cellobiose concentrations down-regulation of cellulase genes was observed due to hydrolysis of cellobiose into glucose by β-glucosidases. However, the gene encoding this permease has not been identified. More recently, two separate studies identified three transporter genes important for lactose uptake and for production of cellulases on lactose cultures (Porciuncula de Oliveira et al. 2013; Ivanova et al. 2013). These results indicate that lactose uptake is an important event for cellulase induction. One of the genes was later further analysed as also being essential for cellulase gene expression on cellulose cultures (Zhang et al. 2013).

Filamentous fungi respond to change of ambient pH of their habitat by an intracellular homeostatic system and by adjusting the expression of the gene products that are directly exposed to the surrounding environment. In *Aspergillus* spp., information on the ambient pH is signalled through a network made up of products of six pal-genes (palA, palB, palC, palF, palH and palI) (Caddick et al. 1986; Arst Jr. et al. 1994; Denison et al. 1995; Maccheroni et al. 1997; Negrete-Urtasun et al. 1997; Denison et al. 1998; Negrete-Urtasun et al. 1999; Herranz et al. 2005; Calcagno-Pizarelli et al. 2007; Hervás-Aguilar et al. 2010). The target of the signalling cascade is the transcription factor PacC, which acts as an activator of alkaline-expressed genes and as a repressor of acidic-expressed genes in alkaline conditions (Tilburn et al. 1995; Espeso et al. 1997). Extracellular hydrolases controlled by PacC include for example *A. nidulans* xylanase genes (*xlnA* and *xlnB*) and an α-L-arabinofuranosidase gene (*abfB*) (MacCabe et al. 1998; Gielkens et al. 1999). One of the first studies on the pH-dependent enzyme production of *T. reesei* indicated that xylanases are preferably produced at higher pH (up to pH7) (Bailey et al. 1993). Cellulase activity increased when pH was decreased from 6 to 4, but the difference was less substantial than the difference in xylanase production between pH4 and pH6. In a recent study, optimal pH for endoglucanase, exoglucanase and β-glucosidase production by *T. reesei* was suggested to be 4.5, 5 and 5.5, respectively (Li et al. 2013). However, both of these studies used Rut-C30 or a mutant derived from it and therefore the results do not necessarily apply to the wild type strain or to other mutants of *T. reesei*, as was shown in a study comparing the protein production of different *T. reesei* mutants at different pH (Adav et al. 2011). The gene encoding the PacC orthologue of *T. reesei* was shown to be up-regulated on media containing cellulose, indicating a possible role in regulation of the production of cellulose-degrading enzymes (dos Santos Castro et al. 2014).

Light has been shown to have an impact on the expression of *T. reesei* cellulase genes, suggesting that the light and nutrient signals received by the fungus

from the surrounding environment are not fully separated from each other but instead exhibit some level of crosstalk. The light signal is mediated via the photoreceptors BLR1, BLR2 (blue light regulator) and the light regulatory protein Envoy (ENV1), which are the central components of the light signalling machinery (Schmoll et al. 2005; Castellanos et al. 2010). In addition, the heterotrimeric G-protein signalling pathway via the G-protein alpha (GNA1 and GNA3), beta (GNB1) and gamma (GNG1) subunits together with a phosducin-like protein PHLP1 and the cyclic AMP pathway are involved in the light-modulated cellulase gene expression (Schmoll et al. 2009; Seibel et al. 2009; Tisch et al. 2011a; Tisch et al. 2011b; Schuster et al. 2012; Tisch et al. 2014). Transcriptome analysis studying the effect of light and darkness and especially gene regulation by ENV1, BLR1 and BLR2 revealed that approximately 75% of *T. reesei* glycoside hydrolases are differentially regulated in darkness and in light (Tisch & Schmoll 2013).

For an organism with the potential to produce very large amounts of proteins, regulation mechanisms must exist in order to ensure that the secretion machinery does not get overwhelmed, resulting in inefficient folding of proteins and in stress responses. *T. reesei* has been shown to secrete proteins most efficiently at low specific growth rates (Pakula et al. 2005; Arvas et al. 2011). However, especially at low specific growth rate the capacity of the fungal cell to fold and secrete proteins sets a limitation and can result in secretion stress. During secretion stress, a feed-back regulation mechanism (repression under secretion stress, RESS) down-regulates the transcription of genes encoding secreted proteins (Pakula et al. 2003). In a study of the effect of growth rate on gene expression and protein production in chemostat cultures, genes were identified of which the expression was either positively or negatively correlated with the specific protein production rate (SPPR) (Arvas et al. 2011). The gene group with a positive correlation of expression with the specific protein production rate was enriched with glycoside hydrolase genes including cellulase genes.

The physiological state of the energy factories of the cell, the mitochondria, has been shown to affect the expression of cellulase genes (Abrahao-Neto et al. 1995). Inhibition of the mitochondrial functions resulted in down-regulation of *cbh1* and *egl1* transcripts. The authors suggested that in the presence of cellulose the glucose released by cellulases and further processed to energy by the tricarboxylic acid cycle would act as a signal for the mitochondria on the availability of energy. When metabolic activity of the cell decreases for example due to oxygen limitation, the cell adjusts by down-regulating the expression of cellulase genes.

1.6 Aims of the study

In this study, genome-wide methods were utilised to investigate the induction of especially cellulase and hemicellulase genes of *T. reesei* in the presence of different substrates and while exposed to different ambient pH conditions. The study started by updating the CAZy gene content of the *T. reesei* genome and by re-annotating the genes in order to assign a putative function for the encoded pro-

teins and to remove discrepancies between different genome versions and published literature. It was believed that re-annotation would reveal novel, previously unidentified CAZy genes and facilitate the identification of necessary activities for biomass degradation. Phylogenetic analysis of the annotated genes was perfomed in order to identify possible functional diversification of the encoded enzymes, thereby explaining the differences observed in the expression patterns of genes presumably encoding similar enzymatic activities. Ambient pH was chosen as the environmental condition to be studied due to the low number of studies performed concerning the effect of extracellular pH on the cellulase and hemicellulase gene expression of *T. reesei*, even though pH is one of the important determinants of protein production efficiency. We were also interested to elucidate whether the *T. reesei* orhologue of ambient pH regulator (PacC) of *Aspergillus* species would be involved in the regulation of cellulase and/or hemicellulase genes. Transcriptional profiling was further utilised for identification of the main activities expressed in the presence of different types of complex and purified substrates and for identifying novel regulators modulating the expression of cellulase and hemicellulase genes. The objective was to identify novel candidate regulators for cellulase and hemicellulase genes by comparing the expression profiles of cellulase and hemicellulase genes to those of candidate regulatory genes. As a summary, this study was based on the hypotheses that there are hitherto unidentified CAZy genes present in the genome of *T. reesei*, the CAZymes of *T. reesei* are functionally diversified, ambient pH has an effect on cellulase and hemicellulase gene expression, the PacC orthologue of *T. reesei* is involved in regulation of cellulase and hemicellulase genes, there are differences in the expression profiles of the *T. reesei* CAZy genes in the presence of different lignocellulose derived materials and there are novel regulatory genes in the genome of *T. reesei* that are involved in regulation of cellulase and hemicellulase production.

2. Materials and methods

The materials and methods used are described in detail in the original papers (I–III). The genomic sequence of *T. reesei* utilised in this work is publicly available in *T. reesei* database 2.0: http://genome.jgi-psf.org/Trire2/Trire2.home.html and in the *T. reesei* database 1.0 (archived genome version): http://genome.jgi-psf.org/trire1/trire1.home.html. Table 7 lists the methods used in the publications.

Table 7. Methods used in Publications I–III.

Method	Publication
Constructing a fungal deletion strain	II, III
Constructing a fungal over-expression strain	II
MUL activity measurement	II, III
Xylanase activity measurement	II, III
Total protein production measurement	II, III
Transcriptional profiling by microarray	I–III
Quantitative PCR	I, II
Bioreactor cultivation	III
Southern hybridisation	II, III
Northern hybridisation	II, III
Shake flask cultivation	II
Induction cultivation	I
Phylogenetic analysis	I
Plate assay for β-glucan hydrolysis	II

2.1 Strains, media and culture conditions

Escherichia coli DH5α (*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) was used for propagation of the plasmids in Publications II and III. *Saccharomyces cerevisiae* FY834 (*MAT*α *his3Δ200 ura3-52 leu2Δ1 lys2Δ202 trp1Δ63*) was used for yeast recombinational cloning in Publication III.The strains used for the transcriptional profiling in Publications I and III were *T. reesei* Rut-C30 (ATCC 56765, VTT-D-86271) and QM9414 (ATCC 26921, VTT-D-74075), respectively. The genomic DNA of *T. reesei* QM6a (ATCC13631, VTT-D-071262T) was used in Publications II and III for PCR amplification of the genes of interest. Strain QM9414 was used in Publications II and III as a host for over-expression of candidate regulatory genes and deletion of the *pac1* gene, respectively. In Publication II the *T. reesei* QM9414Δmus53 strain was used for the construction of a deletion strain. This strain has high targeted integration frequency due to the deletion of gene 58509 (homologue for human DNA ligase IV, Steiger et al. 2011). QM6a is the natural isolate of *T. reesei*. All of the *T. reesei* strains used by industry and for research purposes are originally derived from QM6a. QM9414 strain was derived from the QM9123 strain (first mutant of QM6a with enhanced cellulase production capabilities) (Mandels et al. 1971) by irradiation-induced mutagenesis, and produces two to four times more cellulases than QM6a. Rut-C30 was produced from a separate line of high-producing mutants by three mutagenesis steps (Montenecourt & Eveleigh 1979). Rut-C30 is a carbon catabolite repression deficient mutant. All the fungal strains were obtained from the VTT Culture Collection and were maintained on potato-dextrose (PD) plates. For long term storage, spore suspensions were prepared from PD plates and frozen at -80 °C. For DNA isolation, the fungus was grown in minimal medium supplemented with 0.2% proteose peptone and 2% glucose.

Minimal medium refers to a medium containing $(NH_4)_2SO_4$, KH_2PO_4, $MgSO_4.7H_2O$, $CaCl_2.H_2O$, $CoCl_2$, $FeSO_4.7H_2O$, $ZnSO_4.7H_2O$ and $MnSO_4.7H_2O$. In Publication I the fungus was initially cultivated on minimal medium supplemented with sorbitol. The pre-cultured mycelium was subsequently combined with media containing the inducing substrate suspended in sorbitol-containing minimal medium. Control cultures contained only minimal medium and sorbitol, without an inducing carbon source. The induction experiments were performed in two separate cultivation sets. In the first cultivation set, the inducing substrates used were 0.75% (w/v) Avicel cellulose, 1% (dry matter w/v) pretreated wheat straw, 1% (dry matter w/v) pretreated spruce, or 0.75 mM α-sophorose. In the second cultivation set the inducing substrates were 1% (w/v) Avicel cellulose, 1% (w/v) bagasse ground to homogenous composition, 1% (dry matter w/v) bagasse pretreated using steam explosion, 1% (dry matter w/v) enzymatically hydrolysed pretreated bagasse, 1% (w/v) birch xylan and 1% (w/v) oat spelt xylan. Detailed information on the preparation of the inducing substrates can be found in Publication I.

In Publication II, shake flask cultivations were performed in minimal medium supplemented with 4% lactose and 2% spent grain extract. All of the shake flask

cultivations were performed at 28 ^{o}C in conical flasks with shaking at 250 rpm. In Publication III, bioreactor cultivations were performed in minimal medium supplemented with Avicel cellulose, proteose peptone, Tween80 and an antifoam agent. pH of the bioreactor cultivations was controlled with 15% KOH and 15% H_3PO_4. Bioreactor cultivations were performed in 1.0 litre working volume Sartorius Q plus reactors at 28 °C with dissolved oxygen saturation level of > 30%, agitation 500 rpm – 1200 rpm and constant air flow of 0.6 l/min.

2.2 Expression analysis by microarray hybridisation and quantitative PCR

For microarray analysis, total RNA was first isolated from mycelial samples collected from cultivations and subsequently synthesised into double stranded cDNA. The array designs of the first and second cultivation sets of the induction experiment were based on the genome versions 1.0 and 2.0, respectively. The array design of the ambient pH study was based on the genome version 2.0. Microarray analysis of samples from the first cultivation set was carried out by Roche NimbleGen (Roche-NimbleGen, Inc., Madison, WI, USA) as part of their array service (also including cDNA synthesis). For the microarray analysis of samples from the other experiments, the samples were processed according to the instructions from RocheNimblegen. Double-stranded cDNA of good quality was labelled with Cy3 fluorescent dye, hybridized to microarray slides and scanned using a Roche NimbleGen Microarray scanner. The microarray data was analysed using the R package Oligo for preprocessing of the data and the package Limma for identifying differentially expressed genes between different strains or cultivation conditions (Bolstad et al. 2003; Smyth et al. 2005, http://www.bioconductor.org/). The cut-off used for statistical significance were p-value <0.01 and log2-scale fold change >0.4. Mfuzz clustering was utilised for identifying genes with similar expression profiles (Kumar & Futschik 2007).

Quantitative PCR was used to verify the results of the microarray analysis and for studying the expression of individual genes. Single-stranded cDNA was prepared for the qPCR analysis. The qPCR reactions were performed using a LightCycler 480 SYBR Green I Master kit (Roche) and a Light Cycler 480 II instrument according to the instructions of the manufacturer. The results were analysed with LightCycler 480 Software release 1.5.0. (version 1.5.0.39). Signal from the *gpd1*or *sar1* gene was used for normalisation.

2.3 Construction of fungal over-expression and deletion strains

In Publication II the candidate regulatory genes were amplified by PCR using Gateway compatible primers and the genomic DNA of *T. reesei* QM6a as a template. The amplified genes were subsequently inserted into an expression vector

by the Gateway recombination method. Deletion cassettes were constructed by yeast recombinational cloning or by the Golden Gate method (Colot et al. 2006; Engler et al. 2008). The over-expression/deletion cassettes were transformed to *T. reesei* QM9414 by polyethylene glycol mediated protoplast transformation (Penttilä et al. 1987). Selection of correct transformants was based on hygromycin resistance obtained by integrating the expression/deletion cassette into the genome. After the initial selection, stable transformants were obtained by streaking on plates containing hygromycin B for two successive rounds. Single colonies of the transformants were isolated by plating dilutions of spore suspensions. Integration of the over-expression/deletion cassette was verified by PCR and Southern hybridisation. Over-expression or deletion of genes was verified by Northern hybridisation. In Publication II, a β-glucan plate assay was used for selecting transformants for shake flask cultivation.

2.4 Enzyme activity measurements

Enzymatic activity measured against the substrate 4-methyl umbelliferyl-β-D-lactoside (MUL) is able to detect the total activity produced by cellobiohydrolase 1 (CBHI), endoglucanase 1 (EGI) and β-glucosidase 1 (BGLI). The combined activity of these enzymes was determined by detecting the fluorescent hydrolysis product methyl umbelliferone (MU) released from the MUL substrate (Bailey & Tähtiharju 2003). The combined activity of EGI and CBHI was measured by inhibiting β-glucosidase activity with glucose. EGI activity was measured by adding cellobiose to inhibit CBHI and glucose to inhibit β-glucosidase. CBHI activity was deduced by subtracting EGI activity from the combined CBHI and EGI activity. Endo-β-1,4-xylanase activity was assayed using 1.0% birch glucuronoxylan as a substrate and by detecting the reducing sugars released from the substrate with DNS (Bailey et al. 1992).

3. Results and discussion

In Publication I, the carbohydrate active enzyme gene content of the *T. reesei* genome and the annotations of the genes were updated and the functional diversification of the corresponding enzymes was studied. The expression patterns of the CAZy genes induced by different substrates were studied by transcriptional profiling. In Publication II, the transcriptome data was further analysed to identify candidate regulators for cellulase and hemicellulase genes and a novel gene essential for cellulase gene expression was identified. In Publication III, the response of the *T. reesei* transcriptome and especially of the cellulase and hemicellulase genes to changing ambient pH was studied.

3.1 Identification and re-annotation of the CAZy genes of *T. reesei*

The CAZy gene content of the *T. reesei* genome was updated by combining computational and manual methods. The purpose of the update was to resolve discrepancies between different genome versions (http://genome.jgi-psf.org/Trire2/Trire2.home.html, http://genome.jgi-psf.org/trire1/trire1.home.html) and published literature, and to refine the annotations of the genes. The whole *T. reesei* proteome was mapped to the CAZy database (Cantarel et al. 2009, http://www.cazy.org) by the blastp method (Altschul et al. 1997) (Publication I, Additional file 1). All the non-CAZy genes were subsequently removed, including those that were incorrectly annotated as CAZy genes and genes enoding for other functions than carbohydrate active enzymes.

The annotation process was further facilitated by mapping the *T. reesei* gene products with significant similarity to the CAZy database members to the protein homology clusters including 49 different fungal species (Arvas et al. 2007; Gasparetti et al. 2010, Publication I, Additional file 5). The clusters contain orthologous proteins from different fungi together with paralogues derived from gene duplications and can be utilised to study whether the proteins from other fungi support the given annotation. All the clusters found were subsequently mapped to the CAZy database by blastp. Homology clusters containing CAZymes were identified by filtering the clusters based on the average sequence identity percentage

and length of the blast alignment with CAZymes (at least 97% identity covering 200 amino acids, Publication I, Additional file 2). The remaining clusters were further manually proofed. In total 228 CAZy genes belonging to 61 different families remained after the computational and manual filtering (Publication I, Additional file 3). These included 201 glycoside hydrolase genes, 22 carbohydrate esterase genes and five polysaccharide lyase genes. 13 putatively novel CAZy genes were identified for the first time during this study and for 31 CAZy genes the former annotation was corrected or a new annotation was given. The candidate cellulolytic and hemicellulolytic genes of *T. reesei* are listed in Table 8.

The number of GH61 genes (family AA9 according to the new classification) was updated to include a total of six genes, emphasizing the importance of the enzymes encoded by these genes in assisting cellulose degradation. In addition to the characterized GH67 α-glucuronidase and a CE16 acetyl esterase, a novel candidate acetyl esterase of family CE16 possibly involved in deacetylation of hemicellulose was identified together with a candidate GH115 xylan-α-1,2-glucuronidase/α-(4-O-methyl)-glucuronidase. Updated annotations revealed several genes possibly involved in cellulose and hemicellulose degradation. These genes included a fifth candidate GH2 β-mannosidase gene, a GH2 candidate β-galactosidase/β-glucuronidase gene, a candidate GH5 β-1,3-mannanase/endo-β-1,4-mannosidase gene, a putative second GH12 endoglucanase gene, the first candidate GH39 β-xylosidase gene and four candidate GH79 β-glucuronidase genes.

Updating the annotations and removing discrepancies is critical for the identification of the core set of enzymes and also possible new ezymes and activities necessary for complete biomass degradation. For example, the GH2 candidate β-galactosidase/β-glucuronidase reannotated during this study could be the missing intracellular β-galactosidase responsible for the processing of lactose, as the glycoside hydrolase family 2 is known to include intracellular β-galactosidases.

Table 8. Candidate cellulase, hemicellulase and accessory genes identified from the *T. reesei* genome.

Gene ID	Name	CAZy family	Annotation
73638	cip1	CBM	Candidate cellulose-binding protein
70021		CE3	Candidate acetyl xylan esterase
41248		CE3	Candidate acetyl xylan esterase
44214	axe2	CE5	Candidate acetyl xylan esterase
54219		CE5	Candidate acetyl xylan esterase
103825		CE16	Candidate acetyl esterase
22197	cel1b	GH1	Candidate β-glucosidase
59689		GH2	Candidate β-mannosidase
69245		GH2	Candidate β-mannosidase

5836		GH2	Candidate β-mannosidase
62166		GH2	Candidate β-mannosidase
57857		GH2	Candidate β-mannosidase
76852		GH2	Candidate β-galactosidase/β-glucuronidase
108671	bgl3f	GH3	Candidate β-glucosidase/ glucan 1,4-β-glucosidase
47268	bgl3i	GH3	Candidate β-glucosidase
104797	bgl3j	GH3	Candidate β-glucosidase
121735	cel3b	GH3	Candidate β-glucosidase
82227	cel3c	GH3	Candidate β-glucosidase
46816	cel3d	GH3	Candidate β-glucosidase
76227	cel3e	GH3	Candidate β-glucosidase
58450	xyl3b	GH3	Candidate β-xylosidase
66832		GH3	Candidate β-glucosidase
82616	cel5b	GH5	Candidate membrane-bound endoglucanase
53731		GH5	Candidate endo-β-1,4-glucanase
71554		GH5	Candidate β-1,3-mannanase/ endo-β-1,4-mannosidase
112392	xyn5	GH11	Candidate endo-β-1,4-xylanase
77284		GH12	Candidate endo-β-1,4-glucanase
65986		GH27	Candidate α-galactosidase
55999		GH27	Candidate α-galactosidase
27219		GH27	Candidate α-galactosidase
27259		GH27	Candidate α-galactosidase
59391		GH27	Candidate α-galactosidase
75015		GH27	Candidate α-galactosidase
69276		GH30	Candidate endo-β-1,4-xylanase
69944		GH31	Candidate α-xylosidase/α-glucosidase
73102		GH39	Candidate β-xylosidase
3739		GH43	Candidate β-xylosidase/ α-L-arabinofuranosidase
68064		GH43	Candidate β-xylosidase/ α-L-arabinofuranosidase
55319	abf3	GH54	Candidate α-L-arabinofuranosidase
120961	cel61b	GH61	Candidate copper-dependent polysaccharide mono-oxygenase
22129		GH61	Candidate copper-dependent polysaccharide mono-oxygenase
31447		GH61	Candidate copper-dependent polysaccharide mono-oxygenase
76065		GH61	Candidate copper-dependent polysaccharide mono-oxygenase

27554		GH61	Candidate copper-dependent polysaccharide mono-oxygenase
76210	abf2	GH62	Candidate α-L-arabinofuranosidase
71394		GH79	Candidate β-glucuronidase
73005		GH79	Candidate β-glucuronidase
106575		GH79	Candidate β-glucuronidase
72568		GH79	Candidate β-glucuronidase
72488		GH95	Candidate α-L-fucosidase
5807		GH95	Candidate α-L-fucosidase
111138		GH95	Candidate α-L-fucosidase
58802		GH95	Candidate α-L-fucosidase
79606		GH115	Candidate xylan-α-1,2-glucuronidase or α-(4-O-methyl)-glucuronidase

Cells with blue shading contain genes identified during this study and cells with tan shading contain genes reannotated during this study.

3.2 Phylogenetic analysis of the CAZy genes of *T. reesei*

Phylogenetic methods and analysis of the content of the protein homology clusters were utilised during the annotation process to study the possible functional diversification of the CAZy genes and to compare the number of *T. reesei* proteins to those from other fungi inside the same protein homology cluster (Publication I, Additional files 6 and 7). Phylograms from the protein homology clusters were constructed using 100 bootstraps per tree (Koivistoinen et al. 2012). Phylogenetic analysis revealed several cases of putative horizontal gene transfer from bacteria. Two *T. reesei* proteins (CHI18-15 and 73101) were assigned to protein homology clusters without any other homologues from other fungi (Publication I, Additional files 8 and 9). However, these proteins had homologues in other *Trichoderma* species and in bacteria. The candidate chitinase CHI18-15 was suggested to be a product of horizontal gene transfer in a previous publication (Seidl et al. 2005). In the case of the candidate GH3 β-glucosidase encoded by the gene *bgl3f*, the closest homologues are also from other *Trichoderma* species and from bacteria (Publication I, Additional file 10).

Some of the most striking differences detected in the number of the genes per species inside the protein homology clusters were three expansions and two reductions of *T. reesei* genes (Publication I, Additional file 7). A protein homology cluster including the characterized GH27 α-galactosidase AGLIII together with four candidate α-galactosidases was found to be unique to *T. reesei.* The cluster containing four candidate β-glucuronidases from family GH79 is expanded in *T. reesei* as compared to other fungi. The *T. reesei* genome is also enriched with hemicellulase genes encoding for example GH54 α-arabinofuranosidases, GH67 α-glucuronidases and GH95 α-fucosidases (Druzhinina et al. 2012). All of these enzymes are active against the hemicellulose side chains revealed on the surface

of decaying plant cell wall material, supporting the role of *T. reesei* as a utiliser of pre-digested lignocellulosic substrates. In addition, the cluster containing six candidate GH18 chitinases is expanded. Expansion of GH18 genes of *T. reesei* has been suggested to be involved in functions related to pathogenicity to other fungi (Martinez et al. 2008), although the number of genes is lower than in the two mycotrophic *Trichoderma* species (*T. atroviride* and *T. virens*) (Kubicek et al. 2011; Gruber & Seidl-Seiboth 2012).

The reduction in the number of GH43 and GH61 genes already observed during the initial genome analysis of *T. reesei* (Martinez et al. 2008) was also detected during the phylogenetic analysis. One of the two clusters that contain genes encoding members of family GH43 is hugely reduced in *T. reesei* as compared to other Pezizomycotina species. Reduction is also visible in two protein homology clusters containing members from the family GH61, even though the number of GH61 proteins was updated to include six members.

Phylograms constructed from the individual homology clusters enabled further division of the proteins into different functional subgroups inside the homology clusters (Publication I, Table 1 and Additional file 6). The genome duplication of *Saccharomyces cerevisiae* took place approximately 100 million years ago (Wolfe & Shields 1997; Kellis et al. 2004). The resulting duplicated genes have over time diverged in at least cellular if not in molecular functions (Costenoble et al. 2011). Sordariomycetes diverged from other fungi approximately 400 million years ago, giving more than enough time for the duplicated genes to diverge functionally (Taylor & Berbee 2006). Based on this assumption of functional differentiation, phylograms of each protein homology cluster with multiple *T. reesei* CAZymes were searched for signs of gene duplications that predated the common ancestor of Sordariomycetes.

Several characterized and candidate lignocellulose-degrading enzymes of *T. reesei* belonging to the same CAZy family displayed functional diversification within the protein homology clusters, even when the annotation of the genes indicated similar activity. From the genes encoding activities against cellulose or hemicellulose, functional diversification was abundant among the β-glucosidases from family GH3 and α-galactosidases of family GH27. In addition, the GH18 chitinases were extremely diverse. The protein homology clusters of cellulases and hemicellulases are described in more detail below. Some of the characterized cellulases and hemicellulases were the only members of their CAZy family in *T. reesei* and therefore no functional diversification could be identified for these enzymes. These included CBHII, XYNIII, EGV, ABFII, CEL74A and CIPII. As a conclusion, functional diversification was found to be rather common for the CAZymes of *T. reesei*.

3.2.1 Functional diversification of cellulases

The main cellulases CBHI, CBHII and EGI are divided into two protein homology clusters, the glycoside hydrolase family 6 and 7 members being in different clusters and CBHI and EGI functionally diversified further according to their known

enzymatic activities. From the endoglucanases the characterized GH5 endoglucanase EGII is in the same protein homology cluster and functional subgroup as the candidate membrane-bound endoglucanase CEL5D, emphasizing the possible functional similarity between these enzymes. However, the candidate GH5 endoglucanase (53731) is in a separate cluster, indicating functional diversification from EGII and CEL5D. Similarly, the characterized GH12 endoglucanase EGIII is in a different protein homology cluster than the candidate GH12 endoglucanase 77284 that was reannotated during this study.

According to the updated annotation, *T. reesei* genome harbours six GH61 family genes of which *cel61a* and *cel61b* were previously annotated as endoglucanase genes (Saloheimo et al. 1997; Foreman et al. 2003). The encoded proteins are divided into as many as three protein homology clusters and four functional subgroups inside the clusters. CEL61A is in the same cluster as CEL61B but in a different subgroup. Three candidate GH61 proteins, including two identified during this study, are in the same cluster but in two different subgroups (the novel proteins are in the same subgroup). The fifth candidate protein is assigned to a separate protein homology cluster.

The β-glucosidases of family GH3 were functionally especially diverse.The nine β-glucosidases are divided into two protein homology clusters, and further to nine functional subgroups inside the clusters (Publication I, Figure 1). The candidate β-glucosidases BGL3I, 66832 and BGL3J are assigned to the same cluster as BGLI, CEL3B and CEL3E. The second protein homology cluster includes CEL3D, CEL3C and BGL3F. The two intracellular GH1 β-glucosidases are in the same protein homology cluster but in different functional subgroups.

3.2.2 Functional diversification of hemicellulases

All the three GH11 endo-β-1,4-xylanases of *T. reesei* are in the same protein homology cluster. The candidate xylanase XYNV is predicted to be functionally similar to XYNI but diversified from XYNII. The candidate GH30 xylanase 69276 is not functionally diversified from XYNIV. The only characterized GH5 mannanase of *T. reesei* is in a different protein homology cluster than the candidate β-1,3-mannanase/endo-β-1,4-mannosidase 71554 the annotation of which was updated during this study.

The characterized GH3 β-xylosidase BXLI and a candidate β-xylosidase XYL3B are in the same protein homology cluster, emphasizing the possibly common enzymatic activity of the enzymes. Separate functional subgroups indicate however some type of functional diversification between these enzymes. The candidate β-xylosidase of family GH39 (73102) reannotated during this study is the only member of the family and therefore in a separate homology cluster. Two proteins of the family GH43 predicted to have either β-xylosidase or α-L-arabinofuranosidase activity (68064, 3739) are both assigned to separate protein homology clusters. As mentioned above, the cluster with 3739 as the only *T. reesei* protein is hugely reduced in *T. reesei* compared to especially *Fusarium*

spp., which are the closest relatives of *Trichoderma* in the data set. For example *Fusarium oxysporum* has 12 genes in this cluster.

The two candidate CE5 acetyl xylan esterases 54219 and AXEII are in the same protein homology cluster as the characterized enzyme AXEI. AXEI and 54219 are functionally diversified from AXEII inside the cluster. The arabinofuranosidases of family GH54 (ABFI and ABFIII) are not functionally diversified from each other.

The number of candidate GH2 β-mannosidases was updated to five proteins (5836, 69245, 59689, 57857 and 62166) that are divided into three different functional subgroups within the same protein homology cluster.

The GH27 α-galactosidases are assigned to two protein homology clusters. Proteins encoded by genes 27219, 27259, 59391 and 75015 are in the same cluster and functional subgroup as AGLIII. This cluster is unique to *T. reesei*. The remaining candidate α-galactosidases (55999, 65986 and 72632) are in the same cluster as AGLI and are divided into two functional subgroups within the cluster.

The four novel candidate GH79 β-glucuronidases are all in the same protein homology cluster, and only one of the proteins is functionally diversified from the others. The same is true for the five candidate GH95 α-L-fucosidases. Finally, the new member of the CE16 family acetyl esterases is in the same functional subgroup as the characterised acetyl esterase AESI, providing further support for the annotation of this gene.

3.3 Genome-wide transcriptional analysis of *T. reesei* genes

After the whole genome sequence of *T. reesei* became available, an increasing number of genome-wide studies using methods such as microarray hybridization and RNA sequencing have been conducted. These holistic approaches enable identification of the global responses instead of studying individual genes or pathways.

3.3.1 Impact of ambient pH on cellulase and hemicellulase gene expression

The environmental pH changes as a result of the growth of fungi. Some species of fungi increase the environmental pH whereas others decrease it. Therefore, the presence of other fungal species in the ecosystem affects the ambient pH encountered by *T. reesei*. For example, some of the wood-degrading basidiomycetes have the tendency to decrease the pH of the wood material (Humar et al. 2001). As a saprotroph utilising pre-digested wood as a substrate, *T. reesei* must have developed sufficient regulation mechanisms in order to adapt to the change of extracellular pH.

T. reesei was grown in a bioreactor at different extracellular pH in a medium containing Avicel cellulose in order to study the global response of genes to the change of pH and the expression of particularly cellulase and hemicellulase genes in different pH conditions. The global effects of the orthologue for the character-

ized ambient pH regulator PacC (designated as PACI) were studied by constructing a deletion strain. The *pac1* deletion strain was grown at pH6 in parallel with the parental strain QM9414. In addition, the parental strain was grown at pH3, 4.5 and 6. Transcriptional analysis by the microarray method was applied to the samples collected from the cultivations.

Statistical methods (LIMMA analysis with fold change log2 >0.4 and p-value <0.01 as a threshold) were used to identify the genes responding significantly to the change of pH. The expression of approximately 940 genes changed significantly when different pH conditions were compared (pH6 vs. pH4.5, pH6 vs. pH3 and pH4.5 vs. pH3, Publication III, Figure 1A, Additional file 1). According to the Eukaryotic orthologous (KOG) groups classification, genes with functions related to energy production and conversion; posttranslational modification, protein turnover and chaperones; signal transduction mechanisms; carbohydrate, inorganic ion, lipid and amino acid transport and metabolism and secondary metabolite biosynthesis, transport and catabolism, were abundantly represented among the pH-responsive genes (Publication III, Figure 1B). Closer examination of the gene classes revealed that especially different transporter genes, protease genes, signalling and regulation-related genes and genes possibly having a role in different metabolic reactions were abundant among the pH-responsive genes. The genes responding significantly to the presence of the *pac1* gene were identified by comparing the expression of genes between the *pac1* deletion strain and the parental strain at pH6. In total, 189 genes were found to be differentially expressed between the parental strain and the Δpac1 strain at pH6 (Publication III, Figure 1A, Additional file 2). Approximately 9% of the transcripts from the microarray analysis responded significantly to the change of pH, indicating that ambient pH is an important determinant of *T. reesei* gene expression. The ~2% of transcripts affected by PACI transcription factor include genes that are under direct PACI regulation and genes that are indirectly affected by the deletion of the *pac1* gene. However, the group of pH-responsive genes most probably also includes genes that are not directly affected by pH but instead respond to other factors caused for example by altered growth of the fungus and by stress reactions.

Of the pH-responsive genes as much as 60 were classified as glycoside hydrolases or carbohydrate esterases and one encoded a polysaccharide lyase (Publication III, Figure 1B). From these genes 23 were up-regulated and 38 down-regulated in a pair-wise comparison between pH6 and pH3 and/or pH6 and pH4.5 and/or pH4.5 and pH3. Seven glycoside hydrolase genes and one carbohydrate esterase gene were down-regulated in the *pac1* deletion strain and one glycoside hydrolase gene was up-regulated in the deletion strain (Publication III, Figure 2A). The majority of pH-responsive glycoside hydrolases belong to families GH16, GH18, GH27 and GH55 (Publication III, Figure 4A). All of these families contain activities against the cell wall of fungi, indicating a function in the cell wall rearrangement during growth and/or recycling of cell wall components during autolysis. Rearrangements of the cell wall in order to decrease its permeability could be a response to pH stress.

The pH-responsive genes encoding activities against cellulose and hemicellulose included nearly all the characterized and candidate endo-β-1,4-xylanase genes (except for *xyn4*), one characterized and one candidate endoglucanase gene (*egl3* and *cel5b*), a glucuronoyl esterase gene (*cip2*), two candidate β-glucosidase genes (*bgl3i*, *bgl3j*), a candidate β-xylosidase gene (*xyl3b*), two candidate copper-dependent polysaccharide mono-oxygenase genes (22129 and 31447), candidate acetyl xylan esterase genes (*axe2*, 70021 and 54219), characterized and candidate α-galactosidase genes (*agl1*, *agl3*, 27259 and 27219), a candidate CE16 acetyl esterase gene (103825), a candidate β-1,3-mannanase/endo-β-1,4-mannosidase gene (71554), a candidate β-glucuronidase gene (71394) and two candidate α-fucosidase genes (72488 and 111138) (Figure 3). In addition, a heat map representation from the fold change data assigned two α-galactosidase genes (*agl2*, 59391), a candidate β-glucosidase gene (*bgl3f*), an α-L-arabinofuranosidase gene *(abf1*), a candidate GH5 endoglucanase gene (53731) and a candidate GH39 β-xylosidase gene (73102) to the same branches with genes more highly expressed at low pH (Publication III, Figure 5). Similarly, a candidate GH3 β-glucosidase gene (*cel3d*), an acetyl xylan esterase gene (*axe1*), a candidate copper-dependent polysaccharide mono-oxygenase gene (*cel61b*), a xyloglucanase gene (*cel74a*) and two candidate α-L-arabinofuranosidase genes (*abf2* and *abf3*) were assigned to the same branch with genes more highly expressed at higher pH.

The pH-responsive glycoside hydrolase genes of *T. reesei* included several examples of genes encoding for the same enzymatic activity but responding differently to changes in ambient pH. This phenomenon has been suggested to be due to the need of the fungus to use the same enzyme activities in changing pH conditions (Alkan et al. 2013). A good example is the xylanase genes of *T. reesei*, from which *xyn1* and *xyn5* were preferably expressed at low pH whereas *xyn2* and *xyn3* favoured higher pH. Interestingly, the response of the GH11 xylanase genes to the ambient pH reflects the division of these genes into different functional subgroups as described in section 3.2. Differential pH-regulation of xylanase genes is also known from other fungi. For example, the *A. nidulans* xylanase gene *xlnA* is preferably expressed in alkaline (pH 7.5) conditions and *xlnB* in acidic (pH 4.5) conditions when the fungus is cultivated on medium containing D-xylose (MacCabe et al. 1998).

Additional examples of gene groups with presumably the same enzymatic activity but including both high pH up-regulated and low pH up-regulated genes were detected especially among β-glucosidase and acetyl xylan esterase genes of *T. reesei*. By contrast, all the pH-responsive α-L-galactosidase genes were more highly expressed in acidic conditions, whereas the pH responsive GH61 genes appeared to prefer a higher pH. Similar to the GH11 xylanase genes, acetyl xylan esterase genes of the family CE5 were also divided into different functional subgroups according to the direction of change in expression in different pH conditions (Publication I, Table 1).

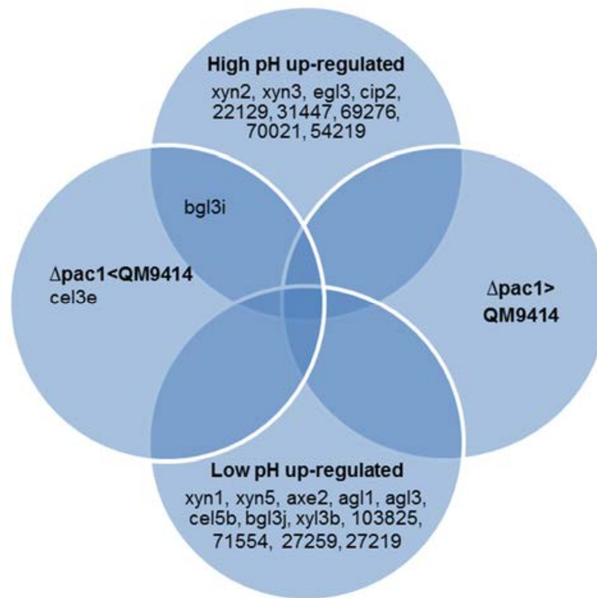


Figure 3. Venn diagram representing the differentially expressed pH-responsive genes encoding activities against cellulose and hemicellulose.

Among the genes down-regulated in the *pac1* deletion strain are two candidate GH3 β-glucosidase genes *bgl3i* and *cel3e* (Figure 3). Genes possibly under direct PACI-mediated repression were identified by applying two criteria. Because the PacC transcription factor of *Aspergillus* spp. is known to be active at alkaline pH, genes under its negative regulation should be expressed at a lower level at pH6 as compared to pH3. The *pacC* deletion strain mimics acidic conditions and therefore the expression of genes normally active only in acidic conditions should be higher at pH6 as compared to the parental strain. Similarly, the genes under direct PACI-mediated induction are expressed at a higher level at pH6 as compared to pH3. Accordingly, these genes should be expressed at a lower level in the deletion strain as compared to the parental strain at pH6.

Suprisingly, the expression patterns of only a few *T. reesei* cellulase and hemicellulase genes indicated PACI-mediated regulation. The strongest and statistically most significant regulation was observed for the candidate β-glucosidase gene *bgl3i* that was therefore suggested to be up-regulated by PACI. In addition, *xyn2* was assigned to the same Mfuzz cluster as the majority of the genes putatively induced by PACI, indicating possible partial PACI-mediated regulation also for this gene. Similarly, the majority of the pH-responsive genes putatively repressed by PACI were assigned to three different Mfuzz clusters (Publication III, Additional file 2). These genes included for example three α-galactosidase genes (*agl1*, 27219 and 27259), two candidate β-xylosidase genes (*xyl3b* and 73102), a candidate α-L-

fucosidase gene (72488), a β-galactosidase gene (*bga1*), a candidate β-galactosidase/β-glucuronidase gene (76852), a candidate acetyl xylan esterase gene (41248), a candidate β-glucuronidase gene (71394) and a candidate acetyl esterase gene (103825). It is worth mentioning that the expression pattern of *xyn1* might also indicate slight repression by PACI (Publication III, Additional file 3). However, more studies are needed to prove the suggested PACI-mediated regulation of these genes.

The fact that the statistical test with the used parameters could not detect PACI-mediated regulation for most of the cellulase and hemicellulase genes could be due to other regulation mechanisms functioning simultaneously and masking the effect of PACI. Avicel cellulose was used as a carbon source and therefore induction by cellulose could partially override the pH regulation of these genes. For example, the promoter of an *Aspergillus tubingensis* xylanase gene has been shown to contain overlapping binding sites for XlnR and PacC, indicating competition between these two regulators (Graaff et al. 1994).

The majority of genes encoding the characterized regulators (*xyr1*, *ace1*, *ace2* and *cre1*) of cellulase and hemicellulase genes were expressed independently of the changing ambient pH, indicating that the pH-dependent expression detected for some of the cellulase and hemicellulase genes is mediated via other regulation mechanisms independently or working together with PACI and/or with the main regulators of genes encoding hydrolytic enzymes.

Supernatant samples were collected throughout the bioreactor cultivations and used for analysis of cellulase and xylanase activity produced during the cultivations. The results indicated that for the strain QM9414 the production of especially xylanase activity was most efficient at pH6 and declined clearly at pH3 (Figure 4). The pH optima of the GH11 xylanases XYNI and XYNII are 4.0-4.5 and 4.0-6.0, respectively (Tenkanen et al. 1992; Torronen et al. 1992). For the GH10 xylanase, XYNIII, the pH optimum has been determined to be 6-6.5 (Xu et al. 1998; J. Wang et al. 2013). Accordingly, when the strain Rut-C30 was cultivated on lactose medium, low pH (pH4) appeared to favour the production of XYNI whereas high pH (pH6) was more optimal for the production of XYNIII (Xiong et al. 2004). XYNII was produced both at high and low pH. Therefore, the high pH up-regulated genes *xyn2* and *xyn3* are most likely responsible for the high xylanase activity produced at pH6. These genes also had higher expression levels as compared to the low pH up-regulated xylanase genes (*xyn1*, *xyn5*), partially explaining increased enzyme activity at pH6.

The difference between cellulase activities produced in different pH conditions was less pronounced. The results of a proteomic study indicate that in strain QM9414 the production of CBHI increases with pH, reaching a maximum at pH6 and subsequently declining (Adav et al. 2011). For the strain Rut-C30 the optimal pH for production of CBHI was 4, which is in accordance with the results of an earlier study showing that the cellulase activity produced by Rut-C30 is improved at pH4 compared to pH6 (Bailey et al. 1993). Similarly, on lactose culture *T. reesei* Rut-C30 produced the highest cellulase and xylanase activities at pH 4.0-4.5 and pH6, respectively (Xiong et al. 2004). However, in our study the growth of the

fungus was slowest at pH3 (Publication III, Figure 8). Therefore, during a longer cultivation the cellulase activity produced at pH3 could possibly reach or even exceed that produced at higher pH.

The cellulase and xylanase activities produced by the *pac1* deletion strain in the bioreactor cultivation at pH6 were clearly decreased in comparison with the parental strain (Figure 4). The effect was especially pronounced for the production of xylanase activity. The transcriptional analysis did not detect clear indications of PACI-mediated regulation of the main cellulase genes. Thus the effect of *pac1* deletion on the enzyme production could be mediated via indirect mechanisms. At least one of the xylanase genes (*xyn2*) was proposed to be activated by PACI according to the Mfuzz clustering of the data. This gene was also found to be the most highly expressed of the high pH up-regulated xylanase genes, indicating an important role under the conditions studied. Therefore, deletion of the *pac1* gene could lower the expression level of *xyn2*, resulting in decreased xylanase activity produced.

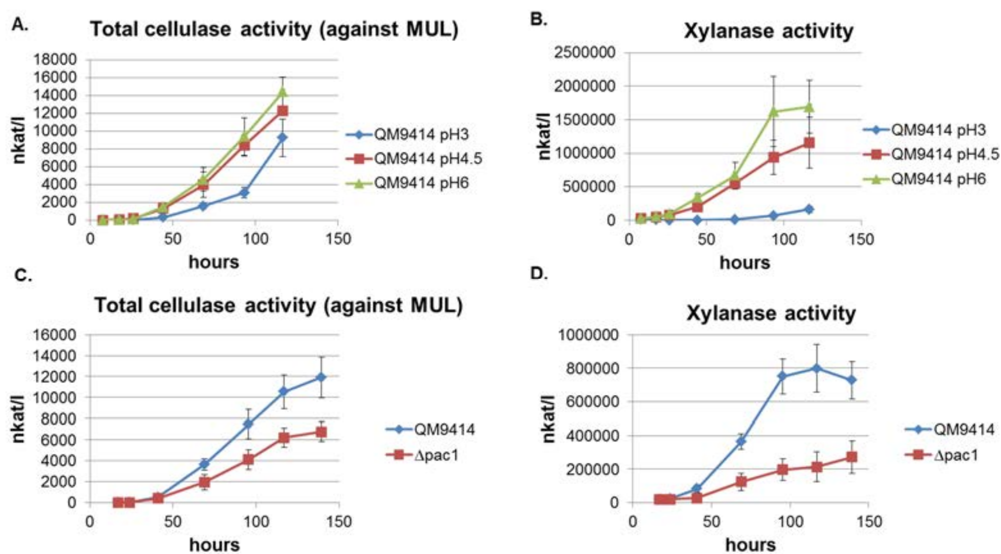


Figure 4. Enzymatic activities produced during bioreactor cultivations of the strains QM9414 and Δpac1. Error bars show the standard error of the mean of three biological replicates.

3.3.2 Impact of different inducing substrates on the expression of CAZy genes

The induction of *T. reesei* CAZy genes was studied by cultivating the fungus in the presence of different inducing substrates including complex polymeric materials (wheat straw, spruce, bagasse), purified polymers (Avicel cellulose, oat spelt xylan

and birch xylan) and a simple disaccharide (α-sophorose). Bagasse was either only ground into smaller texture or steam exploded or further enzymatically pre-treated after steam explosion. Wheat straw and spruce were also pre-treated by steam explosion. The goal of using different types of carbohydrate substrates was to reveal co-expressed gene groups and also genes under different regulation mechanisms. In addition, the biomass substrates selected are of interest concerning biorefinery applications (Talebnia et al. 2010; Cardona et al. 2010). For example, wheat straw is an especially abundant agricultural waste in Europe and bagasse in Brazil (Pessoa-Jr et al. 2005; Talebnia et al. 2010). It is also expected that cultivation on a particular complex material would induce production of the enzymes needed for degradation of the material, and thus the induction patterns of CAZy genes could give information concerning enzyme activities likely to be needed for hydrolysis of the substrate. Of the complex biomass substrates, the composition of the pre-treated spruce was most simple, containing mostly cellulose (Publication I, Table 2). Bagasse and wheat straw also contained arabinoxylan together with galactose and mannose units. Arabinose was the most abundant sugar after glucose and xylose. The birch xylan used was deacetylated glucuronoxylan, and the oat spelt xylan contained mostly xylose together with arabinose.

The fungus was first cultivated on sorbitol in order to get an equal starting material for the induction experiment and to avoid for example growth-dependent differences as a result of the different substrates used. Sorbitol is considered to be a neutral carbon source with respect to induction of cellulase and hemicellulase genes (el-Gogary et al. 1989; Ilmen et al. 1997). The mycelia were subsequently combined with the inducing substrates. The first sample was collected immediately after the inducing substrate was added and other samples after 6 and 17 hours of cultivation (also after 41 hours of cultivation in the case of sophorose). The transcriptional responses were analysed at different time points of induction using oligonucleotide microarrays. The differentially expressed genes were identified by comparing the transcript signals in induced cultures to those in uninduced control cultures at the same time point (Publication I, Additional file 11).

Strong expression in the presence of a specific substrate might indicate that the gene is somehow important for utilisation of the substrate. The main cellulase genes *cbh1*, *cbh2* and *egl2* were highly expressed in the presence of all the tested substrates (signal intensity ≥ 14, calculated as a mean signal intensity of the 6 h, 17 h and 41 h time points). The second main endoglucanase gene, *egl1*, was also highly expressed on most of the substrates, whereas *cel61a/egl4* was most active on wheat straw and spruce. Interestingly, *swo1*, *cip1* and a candidate GH72 β-1,3-glucanosyltransferase gene (82633) were also highly expressed in the presence of all the substrates. The gene encoding for the intracellular GH1 β-glucosidase, *cel1b*, was the most strongly expressed β-glucosidase gene especially on the bagasse materials, birch xylan, Avicel cellulose and sophorose. Of the chitinase genes, *chi18-18* was strongly expressed on the majority of the substrates.

Of the hemicellulase genes, *xyn1*, *xyn4* and *bxl1* were highly expressed on most of the substrates. In addition, *xyn2* was most active on the two purified xylan materials, oat spelt xylan and birch xylan, and *xyn5* on steam exploded bagasse

and oat spelt xylan. Of the genes encoding enzymes cleaving hemicellulose side chains, *axe1* was strongly expressed on the majority of the substrates whereas *aes1* preferred the bagasse materials and wheat straw, *glr1* steam exploded bagasse and the two xylans, *abf1* bagasse, enzymatically treated bagasse, oat spelt xylan and sophorose, and *abf2* untreated bagasse.

Of the 228 CAZy genes (GH, CE and PL), 179 were induced by at least one of the substrates used. The best inducers of CAZy genes were bagasse, xylans and wheat straw. Common to these substrates is that they contain or are composed of hemicellulose. These substrates induced 68–124 genes from 39–47 different CAZy families (Publication I, Figure 2). However, the cellulosic materials including Avicel cellulose and pretreated spruce (contains mostly cellulose) together with the disaccharide sophorose induced a clearly smaller number of genes (43–58 genes in 28–36 families). In accordance with the results of our study, when the transcriptomes of *T. reesei* grown on wheat straw and on lactose were compared, the complex substrate wheat straw was shown to cause stronger and more versatile induction of CAZy genes than the simple disaccharide lactose, suggesting that the differences between the two transcriptomes is due to the xylan component of wheat straw, which represents an additional inducer for several genes (Bischof et al. 2013).

The transcriptomics data enabled identification of the common core of genes induced in the presence of cellulose and hemicellulose substrates (induction in the presence of at least 70% of the substrates used and on both cellulose and xylan). The enzymes encoded by these genes may represent the activities needed for the complete degradation of different plant cell wall materials. Quantitative PCR was used to study induction of the main cellulase genes, of which the signals were saturated in the microarray analysis (Publication I, Figure 5). Two of the main cellulase genes, the GH6 cellobiohydrolase gene *cbh2* and the GH5 endoglucanase gene *egl2*, belong to the core set. The GH7 cellobiohydrolase gene *cbh1,* was induced by all the other substrates except for the xylans, and the GH7 endoglucanase gene *egl1*, was induced only by ground and steam exploded bagasse, Avicel, spruce and sophorose. Other activities encoded by the genes forming the core set are shown in Table 9. It is of interest that *xyn1*, *xyn4*, *xyn5*, *cip1*, *egl4*, *bxl1*, *axe1*, *glr1*, a candidate GH31 α-glucosidase/α-xylosidase gene (69944) and a candidate CE3 acetyl xylan esterase gene (41248) were induced by all the substrates used.

Table 9. Common core of genes induced in the presence of at least 70% of the substrates and on both xylan and cellulose.

Gene ID	Name	CAZy family	Annotation	pH-respon-sive	Highly ex-pressed
62166		GH2	Candidate β-mannosidase		
5836		GH2	Candidate β-mannosidase		
121127	bxl1	GH3	β-xylosidase		x
104797	bgl3j	GH3	Candidate β-glucosidase	x	
76227	cel3e	GH3	Candidate β-glucosidase		
121735	cel3b	GH3	Candidate β-glucosidase		
108671	bgl3f	GH3	Candidate β-glucosidase/ glucan 1,4-β-glucosidase		
58450	xyl3b	GH3	Candidate β-xylosidase	x	
120312	egl2/cel5a	GH5	Endo-β-1,4-glucanase		x
56996	man1	GH5	β-Mannanase		x
72567	cbh2/cel6a	GH6	cellobiohydrolase		x
120229	xyn3	GH10	Endo-β-1,4-xylanase	x	
74223	xyn1	GH11	Endo-β-1,4-xylanase	x	x
112392	xyn5	GH11	Candidate endo-β-1,4-xylanase	x	x
123818	xyn2	GH11	Endo-β-1,4-xylanase	x	x
66041	chi18-18	GH18	Candidate chitinase		x
81598	chi18-7	GH18	Candidate chitinase		
23346	nag2	GH20	Candidate exochitinase		
59391		GH27	Candidate α-galactosidase		
27259		GH27	Candidate α-galactosidase	x	
111849	xyn4	GH30	Endo-β-1,4-xylanase		x
69944		GH31	Candidate α-xylosidase/α-glucosidase		
3739		GH43	Candidate β-xylosidase/ α-L-arabinofuranosidase		
123283	abf1	GH54	α-L-arabinofuranosidase		x
56418		GH55	Candidate β-1,3-glucanase		
54242		GH55	Candidate β-1,3-glucanase		
73643	egl4/cel61a	GH61	Candidate copper-dependent polysaccharide mono-oxygenase/ endo-β-1,4-glucanase		x
22129		GH61	Candidate copper-dependent polysaccharide mono-oxygenase	x	
120961	cel61b	GH61	Candidate copper-dependent polysaccharide mono-oxygenase		

72526	glr1	GH67	α-Glucuronidase		x
71394		GH79	Candidate β-glucuronidase	x	
58117		GH89	Candidate α-N-acetylglucosaminidase		
74198		GH92	Candidate α-1,2-mannosidase		
111138		GH95	Candidate α-L-fucosidase	x	
57179		GH105	Candidate rhamnogalacturonyl hydrolase		
105288		GH	-		
41248		CE3	Candidate acetyl xylan esterase		
73632	axe1	CE5	Acetyl xylan esterase		x
54219		CE5	Candidate acetyl xylan esterase	x	
79671		CE9	Candidate N-acetyl-glucosamine-6-phosphate deacetylase		
123940	cip2	CE15	Glucuronoyl esterase	x	
121418	aes1	CE16	Acetyl esterase		x
73638	cip1	CBM	Candidate cellulose-binding protein		x

Column named "pH-responsive" shows the genes responding significantly to the change of ambient pH. "Highly expressed" indicates a mean signal intensity calculated from time points 6 h and 17 h (and 41 h for sophorose cultures) that is ≥ 14.

As a conclusion, the genes representing the common core of genes induced in the presence of cellulosic and hemicellulosic substrates encode all the main activities against the backbone of the polymers, the necessary side chain cleavage activities and also accessory enzymes. These genes also include uncharacterized candidate genes possibly encoding important activities needed for the total degradation of biomass material. For example, two candidate GH2 β-mannosidases (5836 and 62166), a candidate CE5 acetyl xylan esterase (54219), a candidate GH3 β-xylosidase (XYL3B), a candidate GH27 α-galactosidase (27259), a candidate GH31 α-glucosidase/α-xylosidase (69944), a candidate GH79 β-glucuronidase (71394), two candidate GH3 β-glucosidases (BGL3J and CEL3B) and a candidate GH11 xylanase (XYNV) have all been detected from the secretome of *T. reesei* (Table 4) providing further support for the possible importance of these enzymes in plant cell wall degradation. In addition, several pH-responsive genes were included in the core set (Table 9). Unlike for example in *N. crassa*, the cellulolytic genes of *T. reesei* are induced by hemicellulose and vice versa. This is logical because *T. reesei* does not encounter pure cellulose or xylan in its natural environment and hence the availability of either polymer usually means that the other is also present.

A recent study revealed that in the strain QM9414 the relative transcriptional level (as compared with glucose cultures) of the *xyn5* gene increased in parallel with the levels of *cbh1*, *cbh2* and *egl1* on cellulose and lactose cultures (Chen et al. 2014). In the same study, the induction patterns of for example *xyn5*, *bgl3j* and a candidate GH43 β-xylosidase/α-L-arabinofuranosidase gene (3739) were shown to

be similar to those of *cbh1*, *cbh2* and *egl1* induction on lactose and cellulose cultures (as compared with glucose cultures) and also between the two strains QM9414 and Rut-C30. Notably, the mRNA level of *xyn5* was also higher in Rut-C30 compared to QM9414 in the presence of all the three different carbon sources, indicating a possible role in the enhanced cellulolytic abilities of this mutant strain.

The amount of β-glucosidase activity is often a rate-limiting step in biomass hydrolysis. Of the GH3 β-glucosidase genes, *cel3b*, *cel3e*, *bgl3j* and *bgl3f* were included in the core set (Table 9). BGL3J and CEL3B have been detected from the secretome of *T. reesei* (Table 4) and CEL3E and BGL3F are also expected to be secreted according to the signal sequence prediction (SignalP 4.0, Petersen et al. 2011). Therefore, these genes represent interesting candidates for further studies of improving the β-glucosidase activity produced by *T. reesei*.

The activities not directly involved in cellulose or hemicellulose degradation, such as the chitinases and the GH55 β-1,3-glucanases, are most probably involved in cell wall remodelling during growth. It is possible that some of the uncharacterized chitinases also have other functions than chitin degradation or that the saprotrophic lifestyle and pathogenicity towards other fungi shares common regulation mechanisms. Fungal cell walls are composed of β-1,4-N-acetylglucosamine (chitin) and β-1,3-glucan together with for example α-glucans and galactomannans (Latgé 2007). Although *T. reesei* has evolved away from the mycotrophic lifestyle, its genome content and especially several chitinase and GH16 β-1,3-glucanase genes still reflects its past (Kubicek et al. 2011). Up-regulation of autophagy-related genes was detected when *T. reesei* was cultivated on wheat straw but not on glucose or lactose (Bischof et al. 2013). The authors therefore suggested that the induction of chitinases by wheat straw could be due to enhanced autophagy. However, the chitinase gene chi18-18 that was induced by almost all the substrates used, and was expressed at a high level, has been shown to be more abundantly expressed in the cellulase-overproducing strain Rut-C30 compared to QM9414 on lactose, cellulose and glucose cultures, indicating that it could also have a role in lignocellulose degradation (Chen et al. 2014).

According to the results of RNA sequencing, Ries et al. (2013) suggested that the main enzymes needed by *T. reesei* QM6a for the degradation of wheat straw include GH3 β-glucosidases, GH7 cellobiohydrolase, GH11 and GH30 xylanases, GH61 copper oxidoreductases and CE5 acetyl xylan esterases. Comparison with the wheat straw-induced transcriptome of *A. niger* revealed that both species use approximately similar sets of enzymes for wheat straw degradation, including proteins from glycoside hydrolase families 3, 5, 6, 7,11, 30, 31, 61 (AA9) and 67, although the carbohydrate esterase genes active in the presence of wheat straw differ between the two fungi (Delmas et al. 2012). Overall, the study of Ries et al. (2013) is well in accordance with ours, giving further support to the results presented here. However, in contrast to our study Ries et al. (2013) did not detect induction of genes from GH families 16, 18, 27, 55, 95 and 105. This could be due to different strains used, different pre-treatment methods of wheat straw, different cultivation conditions and different analysis methods applied. However, Bischof et al. (2013) could identify members of these families from the thermochemically pre-

treated wheat straw-induced transcriptome of *T. reesei* QM9414. In contrast with these two studies, our study did not detect induction of *cel1b*, *egl1*, *agl2* and *cel74a* on wheat straw. Of these genes, *cel1b* and *egl1* are expressed at a high level which hinders the ability of a microarray analysis to detect induction. Thus, RNA sequencing is a more applicable method for the detection of induction of highly expressed genes.

Similar sets of cellulase and hemicellulase genes of *T. reesei* and *A. niger* were also activated in the presence of steam exploded bagasse, including enzymes from glycoside hydrolase families 2, 3, 5–7, 10–12, 27, 54, 61, 62, 67 and 95 (de Souza et al. 2011). Some of the major differences between these two fungi were that *T. reesei* GH1 β-glucosidase, GH74 xyloglucanase and GH43 β-xylosidase/α-L-arabinofuranosida genes and *A. niger* GH5 β-mannanase gene were not induced by steam exploded bagasse. Furthermore, a larger number of *T. reesei* GH2 β-mannosidase, GH27 α-galactosidase and GH61 polysaccharide monooxygenase genes was induced as compared to *A. niger*. The set of cellulase and hemicellulase genes identified by Zhang et al. (2013) as part of the Avicel regulon of *T. reesei* is essentially similar to the gene set identified during our study, although our study did not detect the induction of GH74 xyloglucanase, GH62 α-L-arabinofuranosidase, GH35 β-galactosidase or GH36 α-galactosidase genes on Avicel cellulose and Zhang et al. (2013) did not include *swo1*, *cip1*, GH5 β-mannanase, CE16 acetyl esterase, GH31 α-xylosidase/α-glucosidase, GH43 β-xylosidase/α-L-arabinofuranosidase, GH79 β-glucuronidase or GH95 α-L-fucosidase genes in the regulon.

Initial recognition of the substrate and release of the inducing monomers has been suggested to involve inducer-independent constitutive expression of enzymes such as CBHI, CBHII and CEL5B (el-Gogary et al. 1989; Carle-Urioste et al. 1997; Foreman et al. 2003). In accordance with the hypothesis that the candidate membrane-bound endoglucanase, CEL5B, would be involved in the initial recognition of the polymeric substrate (Foreman et al. 2003), this gene was hardly induced in the presence of the substrates used, indicating a constitutive expression level. Visualisation of the expression data revealed highly variable CAZy gene expression patterns in the presence of different substrates and also in the time-course of expression (Publication I, Figure 3 and Additional file 11). We speculated that induction at an early time point of cultivation immediately after adding the substrate, followed by lower expression, could indicate a role in recognition of the substrate or in initialising hydrolysis. Genes displaying such expression profiles included several hemicellulase genes releasing side chains from hemicellulose and digesting oligosaccharides derived from hemicellulose. These genes included for example two candidate β-mannosidase genes (5836 and 69245), a candidate β-xylosidase gene (58450), two α-L-arabinofuranosidase genes (*abf1* and *abf3*) and two α-galactosidase genes (*agl1* and *agl2*). Therefore, the residues released from hemicellulose during the initial recognition of the substrate could function as inducer molecules necessary for activation of the hydrolytic enzyme machinery. Interestingly, endoglucanase gene (*egl3*) was also induced at an early time point of cultivation by several of the substrates. Several CAZy genes of *A. niger* have

been shown to be induced both during early stages of cultivation with wheat straw as a substrate and during early stages of carbon starvation (Delmas et al. 2012; van Munster et al. 2014). The authors suggested that this kind of expression pattern could indicate a role in the early response to wheat straw. Due to the low amount of soluble substrate present at the early time points of cultivation the role of the early induced enzymes could be scouting for a carbon source during starvation. These genes included for example several β-glucosidase genes, an α-galactosidase gene and α-L-arabinofuranosidase genes (van Munster et al. 2014). Thus, our study together with that of van Munster et al. (2014) supports the hypothesis that at least hemicellulose side chain cleavage activities are induced during early stages of substrate recognition and might therefore be involved in creating an inducing substrate needed for the activation of other cellulase and hemicellulase genes necessary for degradation of the substrate.

When the results of the phylogenetic analysis were combined with the results from transcriptional analysis, two main conclusions could be reached. First, for the majority of the functionally diversified genes the expression profiles of the genes also differed (based on clustering of the expression profiles, Publication I, Table 1 and Figure 3). For example the family GH3 β-glucosidases were all divided into separate functional subgroups and also differed in their expression patterns (Publication 1, Figures 1 and 4). The second observation was that tight co-regulation was rare among the genes belonging to the same functional subgroup, indicating that these genes are also differentially regulated. Overall, the observation that functional diversification is rather common for the CAZymes of *T. reesei*, and that the diversification can be seen in differential expression, suggests that the diversified enzymes might be involved in substrate specific processes, have different locations (intracellular/extracellular) and/or have different biochemical properties.

The variable expression patterns and temporal differences of expression detected for the various CAZy genes indicate that several regulation mechanisms act on the promoters simultaneously and possibly also in an additive manner. Activation of different genes depends on the structure and complexity of the substrate. Cellulose fibrils are embedded in hemicellulose matrix that needs to be removed before the cellulases can access the cellulose component of the cell wall. As the degradation of a complex biomass substrate proceeds, more inducing residues are revealed, resulting in up-regulation of additional genes involved in the degradation of the substrate. For example, in *N. crassa* several different regulatory groups controlling the expression of xylanase genes are believed to exist (Sun et al. 2012). The xylanase regulator XLR-1 is believed to work alone or in combination with other regulators. An XLR-1 independent group of genes was also suggested to exist. In *T. reesei*, additional regulators in addition to ACEI that competes with XYRI from the binding site on the promoters of target genes, and ACEII that has been shown to bind the same concensus sequence as XYRI (Aro et al. 2001; Rauscher et al. 2006), most probably also work together with XYRI or independently to fine tune the expression of genes involved in plant cell wall degradation.

3.4 Screening of candidate regulators for cellulase and hemicellulase genes

The data obtained from a transcriptome analysis with Avicel cellulose, pretreated wheat straw, pretreated spruce or sophorose as substrates was further analysed in order to identify novel regulators for cellulase and hemicellulase genes. A double-lock regulation mechanism leading to similar expression patterns of regulatory genes and their target genes was assumed. A double-lock regulation mechanism stands for a regulation cascade in which a master transcription factor regulates the expression of an additional regulatory gene together with its target genes. Therefore, Mfuzz clustering was utilised to identify regulatory genes that have similar expression patterns in the presence of the substrates as known or candidate cellulase and/or hemicellulase genes. Interestingly, the Mfuzz clustering divided the cellulase and hemicellulase genes into two main clusters. Cluster 10 contained mainly cellulase and β-glucosidase genes and cluster 35 contained predominantly hemicellulase genes (Publication II, Figure 1). Hence, these two clusters were the main targets for searching regulatory genes. The clusters were found to be enriched with genes encoding for putative fungal C6 zinc finger type transcription factors. These transcription factors have been suggested to be good candidates for the regulation of non-syntenic block genes, such as CAZy genes, due to the enrichment of the genes in *Pezizomycotina* genomes as compared to *Saccharomycotina* and the location of the genes in non-syntenic blocks of the *T. reesei* genome (Arvas et al. 2011). Genes of cluster 10 were especially interesting due to the presence of all the main cellulase genes and of the *xyr1* gene.

In order also to identify regulatory genes outside the cellulase and hemicellulase clusters but induced by the majority of the substrates, a statistical test was applied to detect significant changes of gene expression between the control cultivation and the cultivations including the inducing substrates (Publication II, Additional file 1). The non-random localisation of the CAZy genes in the genome was also utilised in the selection of the candidate genes for further studies. Regions where regulatory genes are in the close vicinity of CAZy genes were identified. In some cases co-expression of these regulatory genes with CAZy genes was also detected, which further supports the possible role of the encoded regulator in the regulation of the neighbouring CAZy gene. The co-regulated regions are analysed in more detail in section 3.5. Genes encoding for putative transcription factors were the primary targets for further studies. In addition, genes with InterPro domains indicating different regulatory or signal transduction functions were selected. For example, the expression of GCN5-related acetyltransferase genes had positive correlation with the specific extracellular protein production rate (Arvas et al. 2011). The authors suggested that these acetyltransferases might be involved in creating an open chromatin structure enabling the transcription of other specific regulatory factors directly regulating the expression of secreted enzymes. Therefore, candidate GCN5 acetyltransferase genes were also selected from the transcriptome data. Altogether 28 genes were selected for further studies (Publication II, Table 2).

3.4.1 Preliminary analysis of the effects of the candidate regulators

In order to investigate the effects of the putative regulatory genes chosen from the data, *T. reesei* strains over-expressing the genes were constructed from strain QM9414. The genes were cloned to an expression vector under the strong constitutive *gpdA* promoter of *A. nidulans* and the expression plasmids were transformed to QM9414. Correct transformants were screened for enhanced enzyme production on a β-glucan plate and one transformant for each over-expressed gene was selected for further analysis. The selected transformants were cultivated in shake flasks on lactose medium and samples were collected throughout the cultivation. Produced cellulase and xylanase activities were measured from the supernatant samples (Publication II, Figure 4).

When the enzyme activity produced during the cultivation of the recombinant strains was compared to the activity produced in the cultures of the parental strain, seven strains were shown to produce at least 1.5 times more enzyme activity (constructs pMH15, pMH18, pMH20, pMH25, pMH29, pMH35 and pMH36, Publication II, Figure 4). Over-expression of the *A. nidulans creC* (Todd et al. 2000) homologue (construct pMH36) resulted in approximately 1.5 times increased xylanase activity but clearly decreased cellulase activity. In *A. nidulans*, CreC is involved in stabilizing CreB by preventing its proteolysis (Lockington & Kelly 2001; Lockington & Kelly 2002). Disruption of the *creB* orthologue of *T. reesei* (*cre2*) increased cellulase activity (Denton & Kelly 2011). Therefore, if the *creC* homologue also has a similar function in *T. reesei*, over-expression of the gene might result in higher levels of CREII, resulting in lower production of cellulase activity.

Based on the results of the preliminary screening, three recombinant strains were studied further to confirm the observed effects of the candidate regulatory genes on enzyme production (Table 10).

Table 10. Recombinant strains chosen for further studies.

Gene ID	Annotation	Construct	Total MUL activity	Xylanase activity
77513	Fungal specific transcription factor	pMH15	1.9±0.1	2.6±0.1
80291	Fungal transcriptional regulatory protein	pMH20	2.4±0.3	1.3±0.08
74765	Bromodomain-containing protein	pMH25	3.3±1.6	1.7±0.3

Values are shown as a fold change of maximum volumetric activity produced by the recombinant strains as compared to the maximum volumetric activity produced by the parental strain.

3.4.2 Genes 80291 and 74765 have an effect on cellulase and hemicellulase gene expression

Over-expression of gene 80291 (construct pMH20) resulted in the production of approximately 2.5 times more cellulase activity but only less than 1.5 times more xylanase activity as compared to the parental strain, indicating that this candidate

regulator is more specific to cellulase genes than to xylanase genes (Table 10, Publication II, Figure 4). The specific activities detected for cellobiohydrolase 1 (CBHI) and endoglucanase 1 (EGI) supported the results.

A recombinant strain over-expressing the gene 74765 (construct pMH25) produced the highest amount of cellulase activity volumetrically as compared to the other recombinant strains and to the parental strain and xylanase activity was also increased (Table 10, Publication II, Figure 4). Of the specific enzyme activities especially the EGI activity was improved in this strain.

Quantitative PCR analysis confirmed that the over-expression of both genes improved the expression of the main cellulase genes *cbh1*, *cbh2* and *egl1* although the effect of gene 74765 was much more pronounced (Figure 5). In accordance with the enzymatic activity measurements, gene 80291 had a positive effect only on the expression of *xyn1* from the three xylanase genes studied. 74765 had a major effect on the expression of *xyn1* gene and also a clear positive effect on the other two xylanase genes studied. Improvement of expression of the β-glucosidase gene was specific for the strain over-expressing gene 74765. In addition, both strains exhibited improved expression of the β-xylosidase and acetyl xylan esterase genes. Expression of the *xyr1* gene encoding for the regulator of cellulase and hemicellulase genes was not improved in either of the strains, indicating that the positive effects are not mediated via XYRI transcription factor.

Over-expression of the genes and integration of a single copy of the expression cassette were confirmed by Northern and Southern hybridisation (Publication II, Additional files 3 and 4). As a conclusion, both the candidate fungal transcriptional regulatory protein 80291 and the candidate bromodomain-containing protein 74765 were shown to affect the expression of cellulase and hemicellulase genes, although the effect of gene 74765 was much more pronounced. Further studies are required to confirm the role of these genes in cellulase and hemicellulase gene regulation and to elucidate the actual regulatory mechanisms.

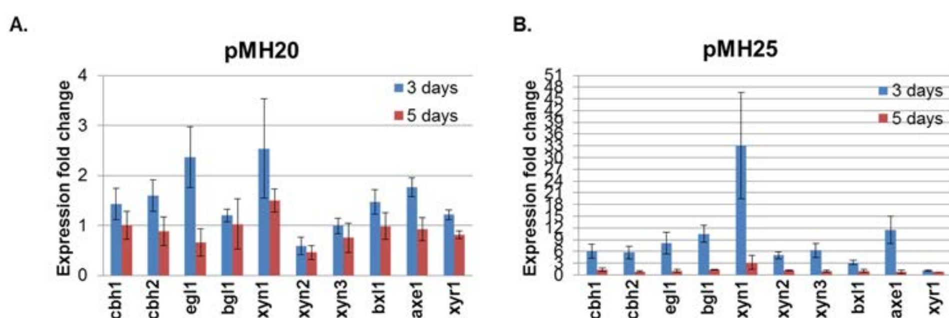


Figure 5. Results of a quantitative PCR analysis for strains over-expressing constructs pMH20 and pMH25. Expression levels normalised against the signal of *sar1* are shown as a fold change as compared to the parental strain. Error bars show the standard error of the mean between three biological replicates.

3.4.3 *ace3* gene is essential for cellulase gene expression and for production of cellulase activity

According to the preliminary analysis, over-expression of gene 77513 (construct pMH15) resulted in a consistently positive effect on both cellulase and xylanase production by *T. reesei* (Table 10, Publication II, Figure 4). The strain produced 2–3 fold more cellulase activity and 2.5–3.5 times more xylanase activity as compared to the parental strain (Publication II, Figure 4). CBHI and EGI activities supported the result and showed that especially CBHI activity was high in this strain. Northern and Southern analyses confirmed that the recombinant strain over-expressed the gene and that two copies of the over-expression cassette were integrated into the genome (Publication II, Additional files 3 and 4).

The over-expression cassette of gene 77513 was integrated to a random position of the genome. In order to ensure that the increase in enzyme production observed was not due to a positional effect, another transformant with a single copy of the over-expression cassette in the genome was analysed together with the double-copy transformant. In order to further study the effect of this candidate regulator, gene 77513 was deleted from the genome of *T. reesei* by replacing the open reading frame with a hygromycin resistance cassette. All three strains were cultivated in shake flasks in the same conditions as previously. Enzymatic activity produced by the strains was measured throughout the cultivation and mycelial samples were collected for quantitative PCR analysis.

The production of cellulase and xylanase activity was improved significantly in both over-expression strains, confirming that the results of the first cultivation were repeatable (Figure 6, Publication II, Figures 5 and 6). In accordance, the improvement in cellulase and xylanase production was higher in the double-copy strain than in the single-copy strain. The results also confirm that the positional effect of the expression cassette is not likely to be significant. In the absence of the gene 77513, the production of cellulase activity by the fungus was abolished completely (Figure 6, Publication II, Figure 8). However, the production of xylanase activity was only decreased to approximately half of the xylanase production of the parental strain, indicating that the gene 77513 is not essential for the production of xylanase activity but is involved in its modulation.

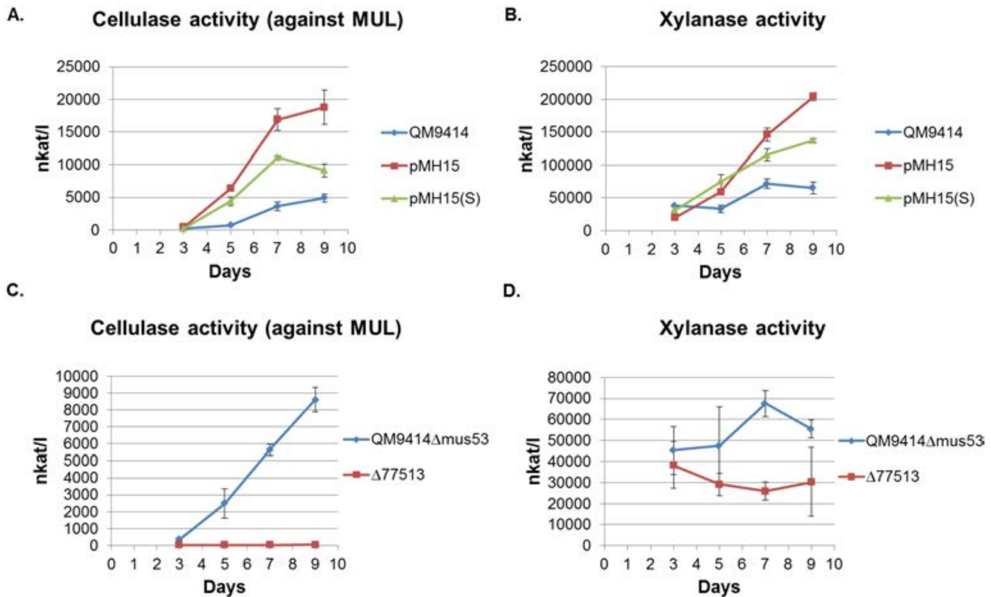


Figure 6. Production of cellulase and xylanase activities by single-copy (pMH15(S)) and double-copy (pMH15) over-expression strains and the deletion strain of gene 77513. Error bars show the standard error of the mean between three biological replicates.

A quantitative PCR analysis was carried out for samples collected from the cultivation (Publication II, Figures 9 and 10). The expression of genes *cbh1*, *cbh2*, *egl1*, *bgl1* and *xyr1* was higher in both of the over-expression strains as compared to the parental strain. Of the xylanase genes, especially *xyn3* had a high expression level in the over-expression strains. In accordance with the enzymatic activity results, the increase in the gene expression was higher in the double-copy strain than in the single-copy strain. These results indicate that it is possible to further enhance the effects of the regulator by increasing the copy number.

Quantitative PCR was also utilised to detect the effect of gene deletion on the expression of individual genes. Deletion of the gene 77513 had the most severe effect on the expression of *cbh1*, *cbh2*, *egl1*, *axe1* and *xyn3*, the expression of which was almost undetectable in the deletion strain as compared to the parental strain (Publication II, Figures 9 and 10). The deletion also had a clear negative effect on the expression of *bxl1*, *xyn1*, *xyn2*, *bgl1* and *xyr1* genes. Hence, the results on the effects of gene 77513 on the production of cellulase and xylanase activity and on the expression of cellulase and hemicellulase genes indicated that this gene encodes for a novel activator of cellulase expression and a modulator of hemicellulase expression. Therefore, this gene was named *ace3* for activator of cellulase expression.

XYRI/XlnR is known to have a major role in the regulation of both cellulase and hemicellulase genes, although in some fungi this regulator is specific for genes involved only in xylan utilization (Brunner et al. 2007; Calero-Nieto et al. 2007; Sun et al. 2012). Thus, the role of *ace3* as primarily affecting cellulase genes is different from that of XYRI/XlnR. However, the possibility of *ace3* affecting cellulase gene expression indirectly cannot be ruled out. As the quantitative PCR-analysis showed, expression of the gene *xyr1* was partially dependent on *ace3*. The expression profiles of *ace3* and *xyr1* are also similar according to the Mfuzz clustering, indicating some level of co-regulation. However, the deletion of *ace3* did not abolish *xyr1* expression, indicating that the absence of a functional XYRI protein is not an explanation for the lack of cellulase activity and gene expression exhibited by the deletion strain. However, it is possible that *ace3* does not bind directly to the promoters of cellulase genes but that instead the effect is mediated via an indirect route, for example by regulating a permease gene. This hypothesis will be further discussed in section 3.5. Furthermore, the effects of different inducing carbon sources on regulation by ACEIII need to be studied. For example, an interesting question is whether, analogously with CLR-1 and CLR-2 of *N. crassa*, the deletion of *ace3* would have a milder effect on hemicellulase genes in the presence of xylan.

Overall, *ace3* is suggested to code for a novel regulator vital for the expression of especially cellulase genes. Some level of crosstalk or cooperation between ACEIII and XYRI is very likely. The positively acting regulator ACEII of *T. reesei* is known to bind the same conserved sequence on the promoters of its target genes as XYRI (Aro et al. 2001; Wurleitner et al. 2003; Stricker et al. 2008). In addition, several novel regulators independent of XlnR or cooperating with it have been identified from different fungi. For example, ManR of *A. oryzea* is involved in the regulation of several genes that are also under XlnR regulation, and ClbR of *A. aculeatus* induces both XlnR-independent and –dependent genes (Marui, Tanaka, et al. 2002; Marui, Kitamoto, et al. 2002; Noguchi et al. 2009; Kunitake et al. 2013). In *A. nidulans*, the F-box protein FbxA has an effect on *xlnR* gene expression (Jonkers et al. 2009; Jonkers & Rep 2009b; Colabardini et al. 2012) and a SRF-MADS box protein mediates induction of two cellulase genes by binding to a promoter region different from the XlnR binding site (Yamakawa et al. 2013). Hence, the differential expression patterns detected with the transcriptional analysis, indicating complex regulation mechanisms of CAZy genes are most likely the result of several different regulators working independently or coordinately, one of these regulators being ACEIII.

Finally, it can be concluded that the approach of utilising transcriptomics data to identify novel regulators was demonstrated to be effective. Clustering of expression profiles enabled identification of similarities in the expression patterns of regulatory genes and their target genes. For example the *ace3* gene was assigned to the same Mfuzz cluster as the genes that were most affected by its modifications (*egl1*, *cbh1*, *cbh2*, *bgl1* and *xyn3*), whereas the genes *axe1*, *bxl1*, *xyn1* and *xyn2* were in different clusters. Interestingly, *ace3* clustered together with the genes *cel1b*, *xyn3*, *cip2* and *egl3* both in the ambient pH data and in the data from

the induction experiment, indicating that these genes have similar expression profiles with *ace3* and with each other both in different ambient pH conditions and in the presence of different inducing substrates.

3.5 Co-regulated genomic gene clusters

In prokaryotes, clustering of functionally related genes is a common feature. In eukaryotes, however, examples of genes involved in the same metabolic or developmental pathway being co-located in genomic clusters are a more recent finding (Keller & Hohn 1997). Especially secondary metabolism genes are often clustered (for a review, see Brakhage 2013). These genes are involved in producing metabolites that are not essential for the growth of the organism but might instead be important for surviving in nutrient-limited conditions and when competing with other organisms. Sequencing of the *T. reesei* genome revealed that CAZy genes are also non-randomly positioned in the genome and often reside close to secondary metabolism gene clusters (Martinez et al. 2008). In later studies, CAZy genes involved in conidiation, and genes of which the expression level correlate with the specific production rate of extracellular proteins, were shown to be located in genomic clusters (Arvas et al. 2011; Metz et al. 2011).

Data from a transcriptional analysis can be utilised for the identification of co-regulated genomic gene clusters. During this study the co-localisation and in some cases also co-regulation of regulatory genes and CAZy genes was revealed (Publication II, Figure 2). This information was utilised in selection of the candidate regulatory genes for further studies, as co-localisation combined with co-expression could indicate that the regulator is involved in the regulation of the other genes forming the cluster. This phenomenon is known especially for secondary metabolism gene clusters, that often contain a regulator specifically regulating the genes of the corresponding pathway (Brakhage 2013). Mfuzz clustering of genes based on the similarity of expression patterns was utilised for the identification of the genomic clusters.

In scaffold 1, the gene 102499 encoding a candidate fungal transcriptional regulatory protein was found to be located between two very tightly co-regulated regions including three CAZy genes forming a cluster below a putative secondary metabolism gene cluster. However, the over-expression of this gene resulted in severe growth deficiency and the strain was therefore omitted from further study. The gene has low homology to the citrinin biosynthesis transcriptional activator CtnA from *Monascus purpureus* and therefore might be involved in regulation of the secondary metabolism cluster. Furthermore, this gene has been shown to be expressed at a higher level on glucose as compared to cellulose and sophorose, indicating that it is not involved in the regulation of cellulolytic genes (dos Santos Castro et al. 2014).

Several regions were found where a β-glucosidase and/or putative sugar transporter gene is located next to and co-expressed with a candidate regulatory gene. Genes *ace3*, 105263 (candidate fungal transcriptional regulatory protein, construct pMH16) and 121121 (candidate fungal transcriptional regulatory protein, construct pMH10) are located next to the candidate β-glucosidase genes *cel1b*, *cel3e* and

cel3d, respectively. Gene *cel3e* encodes a predicted extracellular β-glucosidase according to the signal sequence prediction. Gene *cel3d* is predicted to encode an intracellular enzyme and *cel1b* is the second characterized intracellular β-glucosidase of *T. reesei* (Zhou et al. 2012). However, over-expression of the genes 105263 (pMH16) or 121121 (pMH10) did not have a significant effect on protein production under the conditions studied.

The regions including genes *ace3*, 121121, and 26163 (construct pMH9) contain a putative sugar transporter gene. Gene 26163 is the closest homologue for the *N. crassa clr-2* gene. The transporter gene (3405) next to it has been suggested to be important for cellulase production in lactose cultures (Ivanova et al. 2013). This transporter gene is very highly expressed on different substrates and has a similar expression profile with several cellulase genes according to Mfuzz clustering. In accordance, a recent transcriptional profiling study identified the same transporter gene to be highly expressed on cellulose (Chen et al. 2014). Interestingly, also the transporter gene (77517) close to *ace3* has been shown to be involved in cellulase production on lactose medium (Porciuncula de Oliveira et al. 2013). This gene is also well expressed on the substrates used, although the expression level is clearly lower than for the gene 3405 and it is assigned to the same cellulase gene-enriched cluster according to the expression profile (cluster 10). No published research can be found from the transporter gene next to candidate regulator 121121. Interestingly, this transporter gene is also assigned to cluster 10.

β-Glucosidases release glucose from cellobiose and modify cellobiose into sophorose. Transporters in turn transport the sugars into the cell, where intracellular β-glucosidases might be involved in forming an inducing component such as sophorose. For the novel *N. crassa* cellulase regulators CLR-1 and CLR-2 a regulation mechanism involving activation of β-glucosidases and transporters has been suggested (Coradetti et al. 2012). However, over-expression of the *clr-2* homologue of *T. reesei* did not significantly enhance enzymatic activity produced by the fungus, indicating that this gene might have a different regulation mechanism in *T. reesei*. The *T. reesei* genome does not contain a good homologue for the *clr-1* gene, further underlining the differences between the regulation mechanisms of these two fungi. However, it is still possible that such a mechanism also exists in *T. reesei* and involves ACEIII that activates CEL1B together with the transporter, resulting in accumulation of inducer inside the cell and subsequent induction of cellulase and hemicellulase genes. More studies are needed to reveal the possible co-operation of ACEIII, CEL1B and the transporter in the induction of cellulase genes.

Similar examples of co-localisation of a regulatory gene with a β-glucosidase gene and a transporter gene can be found from other fungi. For example, the homologues of *ace3* of two *Aspergillus* species (*A. fumigatus* and *A. clavatus*) are co-located with a candidate β-glucosidase gene and a candidate hexose transporter gene. Similarly, the homologues of the gene 121121 (pMH10) of *A. fumigatus* and *A.nidulans* are located next to a candidate hexose transporter gene, a candidate MFS multidrug transporter gene, and a β-glucosidase gene. This indicates that there might be an evolutionary benefit behind the co-localisation of regulatory genes, β-glucosidase genes and transporter genes.

4. Conclusions and recommendations

In this study, the whole process from extensive gene annotation to genome-wide transcriptional analysis and to identification of a new important regulator for cellulase genes was described. Furthermore, the power of genome-wide methods in studying the regulatory system of *T. reesei* cellulase and hemicellulase genes was demonstrated. Thorough annotation was shown to be essential for finding new genes possibly involded in the complete degradation of biomass. However, annotations as such might in some cases lead to generalizations concerning the functions encoded by the genes. Phylogenetic analysis gives further evidence for the functional diversification of enzymes but does not identify genes under partially different regulatory mechanisms. Thus, genome-wide gene expression data is needed for identification of differentially regulated genes and of the processes in which the encoded proteins could be involved. A detailed biochemical characterization will be necessary to reveal the actual substrate specificities and functions of the enzymes. Therefore, gene expression data is best utilised for example for developing predictions on the importance of specific genes in particular processes. Based on these predictions, interesting genes can be chosen for further studies which may lead for example to identification of new members of a regulatory cascade.

During this study, several CAZy genes, regulatory genes and transporter genes were identified as good candidates for future studies. A novel transcription factor ACEIII was shown to be vital especially for the production of cellulase activity and for the expression of cellulase genes on lactose cultures. The role of ACEIII in the presence of other carbon sources remains to be elucidated. The detailed characterization of this novel regulator will in the future reveal for example the binding sites and concensus sequence on the promoters of its target genes, the exact regulatory mechanism including possible cooperation with XYRI, and the possible role of β-glucosidases and transporters in the regulation cascade. The overexpression of *ace3* could be utilised in the biotechnology industry for enhancement of cellulase and xylanase production.

Additional information gained from the transcriptional analysis, including recognition of co-regulated genomic clusters, will be useful for studying the evolutionary benefits which this kind of genome organization might have conferred on the fungus. These co-regulated clusters are also good targets for strain improvement if for example one transcriptional regulator regulates the activity of several genes

involved in the efficiency of enzyme production. New information was gained on the pH-dependent expression of *T. reesei* genes. Ambient pH was shown to be an important determinant of gene expression and to represent an additional level of regulation for enzymes degrading plant biomass. Some indications of PACI-mediated regulation of hydrolytic genes were identified. Further studies are needed to distinguish the putative PACI-mediated regulation of cellulase and hemicellulase genes from the other regulators possibly active on the promoters simultaneously, and to determine whether enzyme production could be enhanced by modifying the pH signalling pathway or the *pac1* gene itself.

Based on this study, the most important aspects for further studies in the future are the characterization of ACEIII together with other candidate regulators affecting cellulase and/or hemicellulase production and the candidate CAZy genes found to be activated by several different inducing substrates. The importance of the genomic neighbours of *ace3* could be studied by constructing deletion and over-expression strains for these genes. The function of other genomic co-regulated regions identified could also be investigated. Overall, regulatory mechanisms and especially the regulatory factors of *T. reesei* controlling enzymes degrading plant biomass have been studied extensively. In the future, more attention should be given to other components of the regulatory cascade, such as transporters and intracellular enzymes possibly involved in the formation of an inducer molecule. In addition, novel uncharacterized enzymes active during plant biomass degradation should be studied in order to identify all the activities important for the total hydrolysis of the substrate. Identifying the activities needed for the total degradation of a specific biomass substrate and understanding of the regulatory mechanisms behind the production of these enzymes is vital for designing optimal enzyme cocktails for biomass degradation and for enhancing the production of especially those enzymes limiting the hydrolysis rate. Only then can economical production of biobased second generation fuels and chemicals be possible.

References

Abrahao-Neto, J. et al., 1995. Mitochondrial functions mediate cellulase gene expression in *Trichoderma reesei*. *Biochemistry*, 34(33), pp. 10456–10462.

Adav, S.S. et al., 2011. Proteomic analysis of pH and strains dependent protein secretion of *Trichoderma reesei*. *Journal of Proteome Research*, 10(10), pp. 4579–4596.

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 & cellular proteomics*, 11(7), p.M111.012419.

Akel, E. et al., 2009. Molecular regulation of arabinan and l-arabinose metabolism in Hypocrea jecorina (*Trichoderma reesei*). *Eukaryotic Cell*, 8(12), pp. 1837–1844.

Alkan, N. et al., 2013. Global aspects of pacC regulation of pathogenicity genes in Colletotrichum gloeosporioides as revealed by transcriptome analysis. *Molecular plant-microbe interactions*, 26(11), pp.1345–1358.

Altschul, S.F. et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), pp. 3389–3402.

Aro, N. et al., 2003. ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression. *Applied and Environmental Microbiology*, 69(1), pp. 56–65.

Aro, N. et al., 2001. ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of *Trichoderma reesei*. *Journal of Biological Chemistry*, 276(26), pp. 24309–24314.

Arst, H.N. & Cove, D.J., 1973. Nitrogen metabolite repression in Aspergillus nidulans. *Molecular & General Genetics*, 126(2), pp. 111–141.

Arst Jr., H.N., Bignell, E. & Tilburn, J., 1994. Two new genes involved in signalling ambient pH in Aspergillus nidulans. *Molecular and General Genetics*, 245(6), pp. 787–790.

Arvas, M. et al., 2007. Comparison of protein coding gene contents of the fungal phyla Pezizomycotina and Saccharomycotina. *BMC Genomics*, 8(1), p. 325.

Arvas, M. et al., 2011. Correlation of gene expression and protein production rate - a system wide study. *BMC Genomics*, 12(616).

Bailey, M., Buchert, J. & Viikari, L., 1993. Effect of pH on production of xylanase by *Trichoderma reesei* on xylan- and cellulose-based media. *Applied Microbiology and Biotechnology*, 40(2–3), pp. 224–229.

Bailey, M.J., Biely, P. & Poutanen, K., 1992. Interlaboratory testing of methods for assay of xylanase activity. *Journal of Biotechnology*, 23(3), pp. 257–270.

Bailey, M.J. & Tähtiharju, J., 2003. Efficient cellulase production by *Trichoderma reesei* in continuous cultivation on lactose medium with a computer-controlled feeding strategy. *Applied Microbiology and Biotechnology*, 62(2), pp. 156–162.

Barnett, C.C., Berka, R.M. & Fowler, T., 1991. Cloning and amplification of the gene encoding an extracellular beta-glucosidase from *Trichoderma reesei*: Evidence for improved rates of saccharification of cellulosic substrates. *Nature Biotechnology*, 9(6), pp. 562–567.

Battaglia, E., Visser, L., et al., 2011. Analysis of regulation of pentose utilisation in Aspergillus niger reveals evolutionary adaptations in Eurotiales. *Studies in mycology*, 69(1), pp. 31–38.

Battaglia, E., Hansen, S.F., et al., 2011. Regulation of pentose utilisation by AraR, but not XlnR, differs in Aspergillus nidulans and Aspergillus niger. *Applied microbiology and biotechnology*, 91(2), pp. 387–397.

Béguin, P., 1990. Molecular biology of cellulose degradation. *Annual review of microbiology*, 44, pp. 219–248.

Bischof, R. et al., 2013. Comparative analysis of the *Trichoderma reesei* transcriptome during growth on the cellulase inducing substrates wheat straw and lactose. *Biotechnology for biofuels*, 6(1), p. 127.

Bok, J.W. & Keller, N.P., 2004. LaeA, a regulator of secondary metabolism in Aspergillus spp. *Eukaryotic Cell*, 3(2), pp. 527–535.

Bolstad, B.M. et al., 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2), pp. 185–193.

Brakhage, A.A., 2013. Regulation of fungal secondary metabolism. *Nature Reviews Microbiology*, 11(1), pp. 21–32.

Brunner, K. et al., 2007. Xyr1 regulates xylanase but not cellulase formation in the head blight fungus Fusarium graminearum. *Current genetics*, 52(5-6), pp. 213–220.

Caddick, M.X., Brownlee, A.G. & Arst Jr., H.N., 1986. Regulation of gene expression by pH of the growth medium in Aspergillus nidulans. *Molecular and General Genetics*, 203(2), pp. 346–353.

Calcagno-Pizarelli, A.M. et al., 2007. Establishment of the ambient pH signaling complex in Aspergillus nidulans: PalI assists plasma membrane localization of PalH. *Eukaryotic Cell*, 6(12), pp. 2365–2375.

Calero-Nieto, F. et al., 2007. Role of the transcriptional activator XlnR of Fusarium oxysporum in regulation of xylanase genes and virulence. *Molecular plant-microbe interactions*, 20(8), pp. 977–985.

Cantarel, B.L. et al., 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research*, 37(suppl 1), pp. D233–D238.

Cardona, C.A., Quintero, J.A. & Paz, I.C., 2010. Production of bioethanol from sugarcane bagasse: Status and perspectives. *Bioresource technology*, 101(13), pp. 4754–4766.

Carle-Urioste, J.C. et al., 1997. Cellulase induction in *Trichoderma reesei* by cellulose requires its own basal expression. *Journal of Biological Chemistry*, 272(15), pp. 10169–10174.

Castellanos, F. et al., 2010. Crucial factors of the light perception machinery and their impact on growth and cellulase gene transcription in *Trichoderma reesei*. *Fungal Genetics and Biology*, 47(5), pp. 468–476.

Chen, X. et al., 2014. Transcriptional profiling of biomass degradation-related genes during *Trichoderma reesei* growth on different carbon sources. *Journal of Biotechnology*, 173, pp. 59–64.

Cherry, J.R. & Fidantsef, A.L., 2003. Directed evolution of industrial enzymes: an update. *Current Opinion in Biotechnology*, 14(4), pp. 438–443.

Colabardini, A.C. et al., 2012. Molecular characterization of the Aspergillus nidulans fbxA encoding an F-box protein involved in xylanase induction. *Fungal Genetics and Biology*, 49(2), pp. 130–140.

Colot, H. V. et al., 2006. A high-throughput gene knockout procedure for Neurospora reveals functions for multiple transcription factors. *Proceedings of the National Academy of Sciences*, 103(27), pp. 10352–10357.

Coradetti, S.T. et al., 2012. Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *Proceedings of the National Academy of Sciences*, 109(19), pp. 7397–7402.

Costenoble, R. et al., 2011. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in Saccharomyces cerevisiae under multiple conditions by targeted proteomics. *Molecular Systems Biology*, 7(1), p. 464.

De Groot, M.J.L., 2003. Isolation and characterization of two specific regulatory Aspergillus niger mutants shows antagonistic regulation of arabinan and xylan metabolism. *Microbiology*, 149(5), pp. 1183–1191.

De Groot, M.J.L. et al., 2007. Regulation of pentose catabolic pathway genes of Aspergillus niger. *Food Technology and Biotechnology*, 45(2), pp. 134–138.

Delmas, S. et al., 2012. Uncovering the genome-wide transcriptional responses of the filamentous fungus Aspergillus niger to lignocellulose using RNA sequencing. *PLoS genetics*, 8(8), p. e1002875.

Denison, S.H. et al., 1998. Putative membrane components of signal transduction pathways for ambient pH regulation in Aspergillus and meiosis in Saccharomyces are homologous. *Molecular Microbiology*, 30(2), pp. 259–264.

Denison, S.H., Orejas, M. & Arst, Herbert N., J., 1995. Signaling of ambient pH in Aspergillus involves a cysteine protease. *The Journal of biological chemistry*, 270(48), pp. 28519–28522.

Denton, J.A. & Kelly, J.M., 2011. Disruption of *Trichoderma reesei* cre2, encoding an ubiquitin C-terminal hydrolase, results in increased cellulase activity. *BMC biotechnology*, 11(1), p. 103.

Derntl, C. et al., 2013. Mutation of the Xylanase regulator 1 causes a glucose blind hydrolase expressing phenotype in industrially used Trichoderma strains. *Biotechnology for Biofuels*, 6(1), p. 62.

De Souza, W.R. et al., 2011. Transcriptome analysis of Aspergillus niger grown on sugarcane bagasse. *Biotechnology for biofuels*, 4(1), p. 40.

Dos Santos Castro, L. et al., 2014. Comparative metabolism of cellulose, sophorose and glucose in *Trichoderma reesei* using high-throughput genomic and proteomic analyses. *Biotechnology for Biofuels*, 7(1), p. 41.

Dowzer, C.E. & Kelly, J.M., 1991. Analysis of the creA gene, a regulator of carbon catabolite repression in Aspergillus nidulans. *Molecular and Cellular Biology*, 11(11), pp. 5701–5709.

Dowzer, C.E.A. & Kelly, J.M., 1989. Cloning of the creA gene from Aspergillus nidulans: a gene involved in carbon catabolite repression. *Current Genetics*, 15(6), pp. 457–459.

Druzhinina, I.S., Shelest, E. & Kubicek, C.P., 2012. Novel traits of Trichoderma predicted through the analysis of its secretome. *FEMS microbiology letters*, 337(1), pp. 1–9.

Durand, H., Clanet, M. & Tiraby, G., 1988. Genetic improvement of *Trichoderma reesei* for large scale cellulase production. *Enzyme and Microbial Technology*, 10(6), pp. 341–346.

Duyvesteijn, R.G.E. et al., 2005. Frp1 is a Fusarium oxysporum F-box protein required for pathogenicity on tomato. *Molecular microbiology*, 57(4), pp. 1051–1063.

el-Gogary, S. et al., 1989. Mechanism by which cellulose triggers cellobiohydrolase I gene expression in *Trichoderma reesei*. *Proceedings of the National Academy of Sciences*, 86(16), pp. 6138–6141.

Engler, C., Kandzia, R. & Marillonnet, S., 2008. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE*, 3(11), p. e3647.

Espeso, E.A. et al., 1997. Specific DNA recognition by the Aspergillus nidulans three zinc finger transcription factor PacC. *Journal of molecular biology*, 274(4), pp. 466–480.

Foreman, P.K. et al., 2003. Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *Journal of Biological Chemistry*, 278(34), pp. 31988–31997.

Fowler, T. & Brown, R.D., 1992. The bgl1 gene encoding extracellular beta-glucosidase from *Trichoderma reesei* is required for rapid induction of the cellulase complex. *Molecular microbiology*, 6(21), pp. 3225–3235.

Fritscher, C., Messner, R. & Kubicek, C.P., 1990. Cellobiose metabolism and cellobiohydrolase I biosynthesis by *Trichoderma reesei*. *Experimental Mycology*, 14(4), pp. 405–415.

Furukawa, T. et al., 2009. Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in *Trichoderma reesei*. *Fungal genetics and biology*, 46(8), pp. 564–574.

Gamauf, C. et al., 2007. Characterization of the bga1-encoded glycoside hydrolase family 35 beta-galactosidase of Hypocrea jecorina with galacto-beta-D-galactanase activity. *The FEBS journal*, 274(7), pp. 1691–1700.

Gasparetti, C. et al., 2010. Discovery of a new tyrosinase-like enzyme family lacking a C-terminally processed domain: production and characterization of an Aspergillus oryzae catechol oxidase. *Applied Microbiology and Biotechnology*, 86(1), pp. 213–226.

Gielkens, M. et al., 1999. The abfB gene encoding the major a-L-arabinofuranosidase of Aspergillus nidulans: nucleotide sequence, regulation and construction of a disrupted strain. *Microbiology*, 145(3), pp. 735–741.

Graaff, L.K. et al., 1994. Regulation of the xylanase-encoding xlnA gene of Aspergilius tubigensis. *Molecular Microbiology*, 12(3), pp. 479–490.

Gremel, G., Dorrer, M. & Schmoll, M., 2008. Sulphur metabolism and cellulase gene expression are connected processes in the filamentous fungus Hypocrea jecorina (anamorph *Trichoderma reesei*). *BMC microbiology*, 8(1), p. 174.

Grishutin, S.G. et al., 2004. Specific xyloglucanases as a new class of polysaccharide-degrading enzymes. *Biochimica et Biophysica Acta (BBA) – General Subjects*, 1674(3), pp. 268–281.

Gruber, S. & Seidl-Seiboth, V., 2012. Self versus non-self: fungal cell wall degradation in Trichoderma. *Microbiology*, 158(1), pp. 26–34.

Guillén, D., Sánchez, S. & Rodríguez-Sanoja, R., 2010. Carbohydrate-binding domains: multiplicity of biological roles. *Applied microbiology and biotechnology*, 85(5), pp. 1241–1249.

Harkki, A. et al., 1989. A novel fungal expression system: Secretion of active calf chymosin from the filamentous fungus *Trichoderma reesei*. *Bio/Technology*, 7(6), pp. 596–603.

Harris, P.V. et al., 2010. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry*, 49(15), pp. 3305–3316.

Herold, S. et al., 2013. Xylanase gene transcription in *Trichoderma reesei* is triggered by different inducers representing different hemicellulosic pentose polymers. *Eukaryotic cell*, 12(3), pp. 390–398.

Herpoël-Gimbert, I. et al., 2008. Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. *Biotechnology for Biofuels*, 1(18).

Herranz, S. et al., 2005. Arrestin-related proteins mediate pH signaling in fungi. *Proceedings of the National Academy of Sciences of the United States of America*, 102(34), pp. 12141–12146.

Hervás-Aguilar, A., Galindo, A. & Peñalva, M.A., 2010. Receptor-independent ambient pH signaling by ubiquitin attachment to fungal arrestin-like PalF. *Journal of Biological Chemistry*, 285(23), pp. 18095–18102.

Humar, M., Petrič, M. & Pohleven, F., 2001. Changes of the pH value of impregnated wood during exposure to wood-rotting fungi. *Holz als Roh- und Werkstoff*, 59(4), pp. 288–293.

Igarashi, K. et al., 2011. Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science*, 333(6047), pp. 1279–1282.

Ilmen, M. et al., 1997. Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. *Applied and Environmental Microbiology*, 63(4), pp. 1298–1306.

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(4), pp. 451–460.

Ivanova, C. et al., 2013. Systems analysis of lactose metabolism in *Trichoderma reesei* identifies a lactose permease that is essential for cellulase induction. *PLoS ONE*, 8(5), p. e62631.

Jonkers, W. & Rep, M., 2009a. Lessons from fungal F-box proteins. *Eukaryotic cell*, 8(5), pp. 677–695.

Jonkers, W. & Rep, M., 2009b. Mutation of CRE1 in Fusarium oxysporum reverts the pathogenicity defects of the FRP1 deletion mutant. *Molecular microbiology*, 74(5), pp. 1100–1113.

Jonkers, W., Rodrigues, C.D.A. & Rep, M., 2009. Impaired colonization and infection of tomato roots by the Δfrp1 mutant of Fusarium oxysporum correlates with reduced CWDE gene expression. *Molecular plant-microbe interactions*, 22(5), pp. 507–518.

Jun, H., Guangye, H. & Daiwen, C., 2013. Insights into enzyme secretion by filamentous fungi: comparative proteome analysis of *Trichoderma reesei* grown on different carbon sources. *Journal of proteomics*, 89, pp. 191–201.

Jun, H., Kieselbach, T. & Jönsson, L.J., 2011. Enzyme production by filamentous fungi: analysis of the secretome of *Trichoderma reesei* grown on unconventional carbon source. *Microbial cell factories*, 10(1), p. 68.

Karaffa, L. et al., 2006. D-Galactose induces cellulase gene expression in Hypocrea jecorina at low growth rates. *Microbiology*, 152(5), pp. 1507–1514.

Karimi-Aghcheh, R. et al., 2013. Functional analyses of *Trichoderma reesei* LAE1 reveal conserved and contrasting roles of this regulator. *G3: Genes|Genomes|Genetics*, 3(2), pp. 369–378.

Keller, N.P. & Hohn, T.M., 1997. Metabolic pathway gene clusters in filamentous fungi. *Fungal Genetics and Biology*, 21(1), pp. 17–29.

Kellis, M., Birren, B.W. & Lander, E.S., 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature*, 428(6983), pp. 617–624.

Koivistoinen, O.M. et al., 2012. Characterisation of the gene cluster for l-rhamnose catabolism in the yeast Scheffersomyces (Pichia) stipitis. *Gene*, 492(1), pp. 177–185.

Kolpak, F.J. & Blackwell, J., 1976. Determination of the structure of cellulose II. *Macromolecules*, 9(2), pp. 273–278.

Kredics, L. et al., 2014. Biodiversity of the genus Hypocrea/Trichoderma in different habitats. In V. K. Gupta et al., eds. *Biology and biotechnology of Trichoderma*. Elsevier, pp. 3–18.

Kubicek, C. et al., 1993. Triggering of cellulase biosynthesis by cellulose in *Trichoderma reesei*. Involvement of a constitutive, sophorose-inducible, glucose- inhibited beta-diglucoside permease. *The Journal of biological chemistry*, 268(26), pp. 19364–19368.

Kubicek, C.P. et al., 2011. Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of Trichoderma. *Genome biology*, 12(4), p. R40.

Kubicek, C.P., 1987. Involvement of a conidial endoglucanase and a plasma-membrane-bound β-glucosidase in the induction of endoglucanase synthesis by cellulose in *Trichoderma reesei*. *Microbiology*, 133(6), pp. 1481–1487.

Kubicek, C.P., 2013. Systems biological approaches towards understanding cellulase production by *Trichoderma reesei*. *Journal of biotechnology*, 163(2), pp. 133–142.

Kuhad, R.C., Gupta, R. & Singh, A., 2011. Microbial cellulases and their industrial applications. *Enzyme research*, 2011(280696).

Kuhls, K. et al., 1996. Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete Hypocrea jecorina. *Proceedings of the National Academy of Sciences of the United States of America*, 93(15), pp. 7755–7760.

Kumar, L. & Futschik, M.E., 2007. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation*, 2(1), pp. 5–7.

Kumar, R., Singh, S. & Singh, O.V., 2008. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *Journal of industrial microbiology & biotechnology*, 35(5), pp. 377–391.

Kunamneni, A. et al., 2014. Trichoderma enzymes for food industries. In V. K. Gupta et al., eds. *Biology and biotechnology of Trichoderma*. Elsevier, pp. 339–343.

Kunitake, E. et al., 2013. A novel transcriptional regulator, ClbR, controls the cellobiose- and cellulose-responsive induction of cellulase and xylanase genes regulated by two distinct signaling pathways in Aspergillus aculeatus. *Applied microbiology and biotechnology*, 97(5), pp. 2017–2028.

Langston, J.A. et al., 2011. Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Applied and Environmental Microbiology*, 77(19), pp. 7007–7015.

Latgé, J.-P., 2007. The cell wall: a carbohydrate armour for the fungal cell. *Molecular microbiology*, 66(2), pp. 279–290.

Levasseur, A. et al., 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for biofuels*, 6(1), p. 41.

Li, C. et al., 2013. Effect of pH on cellulase production and morphology of *Trichoderma reesei* and the application in cellulosic material hydrolysis. *Journal of Biotechnology*, 168(4), pp. 470–477.

Li, X.-L. et al., 2007. Identification of genes encoding microbial glucuronoyl esterases. *FEBS letters*, 581(21), pp. 4029–4035.

Li, X.-L. et al., 2008. Novel family of carbohydrate esterases, based on identification of the Hypocrea jecorina acetyl esterase gene. *Applied and Environmental Microbiology*, 74(24), pp. 7482–7489.

Lockington, R.A. et al., 2002. Regulation by carbon and nitrogen sources of a family of cellulases in Aspergillus nidulans. *Fungal Genetics and Biology*, 37(2), pp. 190–196.

Lockington, R.A. & Kelly, J.M., 2001. Carbon catabolite repression in Aspergillus nidulans involves deubiquitination. *Molecular Microbiology*, 40(6), pp. 1311–1321.

Lockington, R.A. & Kelly, J.M., 2002. The WD40-repeat protein CreC interacts with and stabilizes the deubiquitinating enzyme CreB in vivo in Aspergillus nidulans. *Molecular Microbiology*, 43(5), pp. 1173–1182.

MacCabe, A.P. et al., 1998. Opposite patterns of expression of two Aspergillus nidulans xylanase genes with respect to ambient pH. *Journal of bacteriology*, 180(5), pp. 1331–1333.

Maccheroni, W. et al., 1997. The sequence of palF, an environmental pH response gene in Aspergillus nidulans. *Gene*, 194(2), pp. 163–167.

Mach, R.L. et al., 1995. The bgl1 gene of *Trichoderma reesei* QM9414 encodes an extracellular, cellulose-inducible β-glucosidase involved in cellulase induction by sophorose. *Molecular Microbiology*, 16(4), pp. 687–697.

Mach-Aigner, A.R. et al., 2008. Transcriptional regulation of xyr1, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in Hypocrea jecorina. *Applied and Environmental Microbiology*, 74(21), pp. 6554–6562.

Mach-Aigner, A.R., Gudynaite-Savitch, L. & Mach, R.L., 2011. L-Arabitol is the actual inducer of xylanase expression in Hypocrea jecorina (*Trichoderma reesei*). *Applied and Environmental Microbiology*, 77(17), pp. 5988–5994.

Mach-Aigner, A.R., Pucher, M.E. & Mach, R.L., 2010. D-Xylose as a repressor or inducer of xylanase expression in Hypocrea jecorina (*Trichoderma reesei*). *Applied and environmental microbiology*, 76(6), pp. 1770–1776.

Mandels, M., Parrish, F.W. & Reese, E.T., 1962. Sophorose as an inducer of cellulase in Trichoderma viride. *The Journal of Bacteriology*, 83(2), pp. 400–408.

Mandels, M., Weber, J. & Parizek, R., 1971. Enhanced cellulase production by a mutant of Trichoderma viride. *Applied microbiology*, 21(1), pp. 152–154.

Margolles-Clark, E., Tenkanen, M., Söderlund, H., et al., 1996. Acetyl xylan esterase from *Trichoderma reesei* contains an active-site serine residue and a cellulose-binding domain. *European Journal of Biochemistry*, 237(3), pp. 553–560.

Margolles-Clark, E., Tenkanen, M., Nakari-Setälä, T., et al., 1996. Cloning of genes encoding alpha-L-arabinofuranosidase and beta-xylosidase from *Trichoderma reesei* by expression in Saccharomyces cerevisiae. *Applied and Environmental Microbiology*, 62(10), pp. 3840–3846.

Margolles-Clark, E., Saloheimo, M., Siika-aho, M., et al., 1996. The α-glucuronidase-encoding gene of *Trichoderma reesei*. *Gene*, 172(1), pp. 171–172.

Margolles-Clark, E., Tenkanen, M., Luonteri, E., et al., 1996. Three α-galactosidase genes of *Trichoderma reesei* cloned by expression in yeast. *European Journal of Biochemistry*, 240(1), pp. 104–111.

Margolles-Clark, E., Ilmen, M. & Penttilä, M., 1997. Expression patterns of ten hemicellulase genes of the filamentous fungus *Trichoderma reesei* on various carbon sources. *Biochemistry and genetics of cellulases and hemicellulases and their application*, 57(1–3), pp. 167–179.

Martinez, D. et al., 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. Hypocrea jecorina). *Nature Biotechnology*, 26(5), pp. 553–560.

Marui, J., Tanaka, A., et al., 2002. A transcriptional activator, AoXlnR, controls the expression of genes encoding xylanolytic enzymes in Aspergillus oryzae. *Fungal genetics and biology*, 35(2), pp. 157–169.

Marui, J., Kitamoto, N., et al., 2002. Transcriptional activator, AoXlnR, mediates cellulose-inductive expression of the xylanolytic and cellulolytic genes in Aspergillus oryzae. *FEBS Letters*, 528(1–3), pp. 279–282.

Marx, I.J. et al., 2013. Comparative secretome analysis of Trichoderma asperellum S4F8 and *Trichoderma reesei* Rut C30 during solid-state fermentation on sugarcane bagasse. *Biotechnology for biofuels*, 6(1), p. 172.

McKendry, P., 2002. Energy production from biomass (part 1): overview of biomass. *Bioresource technology*, 83(1), pp. 37–46.

Messenguy, F. & Dubois, E., 2003. Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene*, 316, pp. 1–21.

Messner, R. et al., 1991. Cellobiohydrolase II is the main conidial-bound cellulase in *Trichoderma reesei* and other Trichoderma strains. *Archives of Microbiology*, 155(6), pp. 601–606.

Metz, B. et al., 2011. Expression of biomass-degrading enzymes is a major event during conidium development in *Trichoderma reesei*. *Eukaryotic Cell*, 10(11), pp. 1527–1535.

Mong Chen, C., Gritzali, M. & Stafford, D.W., 1987. Nucleotide sequence and deduced primary structure of cellobiohydrolase II from *Trichoderma reesei*. *Bio/Technology*, 5(3), pp. 274–278.

Montenecourt, B.S. & Eveleigh, D.E., 1979. Selective screening methods for the isolation of high yield cellulase mutans of *Trichoderma reesei*. In R. D. Brown & L. Jurasek, eds. *Hydrolysis of Cellulose: Mechanisms of Enzymatic and Acid Catalysis*. Advances in Chemistry. Washington, D. C.: American Chemical Society, pp. 289–301.

Montenecourt, B.S. & Eveleigh, D.E., 1977a. Preparation of mutants of *Trichoderma reesei* with enhanced cellulase production. *Applied and*

Environmental Microbiology, 34(6), pp. 777–782.Montenecourt, B.S. & Eveleigh, D.E., 1977b. Semiquantitative plate assay for determination of cellulase production by Trichoderma viride. *Applied and Environmental Microbiology*, 33(1), pp. 178–183.

Nakari-Setälä, T. et al., 2009. Genetic modification of carbon catabolite repression in *Trichoderma reesei* for improved protein production. *Applied and environmental microbiology*, 75(14), pp. 4853–4860.

Negrete-Urtasun, S. et al., 1999. Ambient pH signal transduction in Aspergillus: completion of gene characterization. *Molecular Microbiology*, 33(5), pp. 994–1003.

Negrete-Urtasun, S., Denison, S. & Arst Jr., H.N., 1997. Characterization of the pH signal transduction pathway gene palA of Aspergillus nidulans and identification of possible homologs. *Journal of bacteriology*, 179(5), pp. 1832–1835.

Nevalainen, K.M.H., Te'o, V.S.J. & Bergquist, P.L., 2005. Heterologous protein expression in filamentous fungi. *Trends in biotechnology*, 23(9), pp. 468–474.

Nitta, M. et al., 2012. A new Zn(II)2Cys6-type transcription factor BglR regulates β-glucosidase expression in *Trichoderma reesei*. *Fungal Genetics and Biology*, 49(5), pp. 388–397.

Nogawa, M. et al., 2001. l -Sorbose induces cellulase gene transcription in the cellulolytic fungus *Trichoderma reesei*. *Current genetics*, 38(6), pp. 329–334.

Noguchi, Y. et al., 2009. Genes regulated by AoXlnR, the xylanolytic and cellulolytic transcriptional regulator, in Aspergillus oryzae. *Applied microbiology and biotechnology*, 85(1), pp. 141–154.

Nyyssönen, E. et al., 1993. Efficient production of antibody fragments by the filamentous fungus *Trichoderma reesei*. *Bio/Technology*, 11(5), pp. 591–595.

Ogawa, M., Kobayashi, T. & Koyama, Y., 2012. ManR, a novel Zn(II)2Cys6 transcriptional activator, controls the β-mannan utilization system in Aspergillus oryzae. *Fungal genetics and biology*, 49(12), pp. 987–995.

Ogawa, M., Kobayashi, T. & Koyama, Y., 2013. ManR, a transcriptional regulator of the β-mannan utilization system, controls the cellulose utilization system in Aspergillus oryzae. *Bioscience, biotechnology, and biochemistry*, 77(2), pp. 426–429.

Okada, H. et al., 1998. Molecular characterization and heterologous expression of the gene encoding a low-molecular-mass endoglucanase from *Trichoderma reesei* QM9414. *Applied and Environmental Microbiology*, 64(2), pp. 555–563.

Ouyang, J. et al., 2006. A complete protein pattern of cellulase and hemicellulase genes in the filamentous fungus *Trichoderma reesei*. *Biotechnology Journal*, 1(11), pp. 1266–1274.

Pakula, T.M. et al., 2005. The effect of specific growth rate on protein synthesis and secretion in the filamentous fungus *Trichoderma reesei*. *Microbiology*, 151(1), pp. 135–143.

Pakula, T.M. et al., 2003. The effects of drugs inhibiting protein secretion in the filamentous fungus *Trichoderma reesei*. *Journal of Biological Chemistry*, 278(45), pp. 45011–45020.

Penttilä, M. et al., 1987. A versatile transformation system for the cellulolytic filamentous fungus *Trichoderma reesei*. *Gene*, 61(2), pp. 155–164.

Penttilä, M. et al., 1986. Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence of the endoglucanase I gene. *Gene*, 45(3), pp. 253–263.

Pessoa-Jr, A. et al., 2005. Perspectives on bioenergy and biotechnology in Brazil. *Applied Biochemistry and Biotechnology*, 121(1), pp. 59–70.

Petersen, T.N. et al., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10), pp. 785–786.

Pokkuluri, P.R. et al., 2011. Structure of the catalytic domain of glucuronoyl esterase Cip2 from Hypocrea jecorina. *Proteins: Structure, Function, and Bioinformatics*, 79(8), pp. 2588–2592.

Porciuncula de Oliveira, J. et al., 2013. Identification of major facilitator transporters involved in cellulase production during lactose culture of *Trichoderma reesei* PC-3-7. *Bioscience, biotechnology, and biochemistry*, 77(5), pp. 1014–1022.

Portnoy, T. 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(2), pp. 262–271.

Puranen, T., Alapuranen, M. & Vehmaanperä, J., 2014. Trichoderma enzymes for textile industries. In V. K. Gupta et al., eds. *Biology and biotechnology of Trichoderma*. Elsevier, pp. 351–359.

Rauscher, R. et al., 2006. Transcriptional regulation of xyn1, encoding xylanase I, in Hypocrea jecorina. *Eukaryotic cell*, 5(3), pp. 447–456.

Reese, E.T., 1976. History of the cellulase program at the U.S. army Natick Development Center. *Biotechnology and bioengineering symposium*, 6, pp. 9–20.

Reyes-Dominguez, Y. et al., 2010. Heterochromatic marks are associated with the repression of secondary metabolism clusters in Aspergillus nidulans. *Molecular microbiology*, 76(6), pp. 1376–1386.

Ries, L. et al., 2013. Genome-wide transcriptional response of *Trichoderma reesei* to lignocellulose using RNA sequencing and comparison with Aspergillus niger. *BMC genomics*, 14(1), p. 541.

Rossman, A.Y. et al., 1999. Genera of Bionectriaceae, Hypocreaceae and Nectriaceae (Hypocreales, Ascomycetes). *Studies in Mycology*, (42), pp. 1–83.

Saloheimo, A. et al., 1994. A novel, small endoglucanase gene, egl5, from *Trichoderma reesei* isolated by expression in yeast. *Molecular microbiology*, 13(2), pp.219–228.

Saloheimo, A. et al., 2000. Isolation of the ace1 gene encoding a Cys2-His2 transcription factor involved in regulation of activity of the cellulase promoter cbh1 of *Trichoderma reesei*. *Journal of Biological Chemistry*, 275(8), pp. 5817–5825.

Saloheimo, M. et al., 1997. cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. *European Journal of Biochemistry*, 249(2), pp. 584–591.

Saloheimo, M. et al., 1988. EGIII, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme. *Gene*, 63(1), pp. 11–21.

Saloheimo, M., Kuja-Panula, J., et al., 2002. Enzymatic properties and intracellular localization of the novel *Trichoderma reesei* β-glucosidase BGLII (Cel1A). *Applied and Environmental Microbiology*, 68(9), pp. 4546–4553.

Saloheimo, M., Paloheimo, M., et al., 2002. Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. *European Journal of Biochemistry*, 269(17), pp. 4202–4211.

Saloheimo, M. et al., 2003. Xylanase from *Trichoderma reesei*, method for production thereof, and methods employing this enzyme I. Genencor International, ed. , 09/658,772(6555335).

Saloheimo, M. & Niku-Paavola, M.-L., 1991. Heterologous production of a ligninolytic enzyme: expression of the Phlebia radiata laccase gene in *Trichoderma reesei*. *Bio/Technology*, 9(10), pp. 987–990.

Saloheimo, M. & Pakula, T.M., 2012. The cargo and the transport system: secreted proteins and protein secretion in *Trichoderma reesei* (Hypocrea jecorina). *Microbiology*, 158(1), pp. 46–57.

Sánchez, C., 2009. Lignocellulosic residues: Biodegradation and bioconversion by fungi. *Biotechnology Advances*, 27(2), pp.185–194.

Scheller, H.V. & Ulvskov, P., 2010. Hemicelluloses. *Annual review of plant biology*, 61, pp. 263–289.

Schmoll, M. et al., 2009. The G-alpha protein GNA3 of Hypocrea jecorina (anamorph *Trichoderma reesei*) regulates cellulase gene expression in the presence of light. *Eukaryotic Cell*, 8(3), pp. 410–420.

Schmoll, M., Franchi, L. & Kubicek, C.P., 2005. Envoy, a PAS/LOV domain protein of Hypocrea jecorina (anamorph *Trichoderma reesei*), modulates cellulase gene transcription in response to light. *Eukaryotic Cell*, 4(12), pp. 1998–2007.

Schuster, A. et al., 2012. Roles of protein kinase a and adenylate cyclase in light-modulated cellulase regulation in *Trichoderma reesei*. *Applied and Environmental Microbiology*, 78(7), pp. 2168–2178.

Schuster, A., Kubicek, C.P. & Schmoll, M., 2011. Dehydrogenase GRD1 represents a novel component of the cellulase regulon in *Trichoderma reesei* (Hypocrea jecorina). *Applied and environmental microbiology*, 77(13), pp. 4553–4563.

Seibel, C. et al., 2009. Light-dependent roles of the G-protein alpha subunit GNA1 of Hypocrea jecorina (anamorph *Trichoderma reesei*). *BMC Biology*, 7(1), p. 58.

Seiboth, B. et al., 2005. Role of the bga1-encoded extracellular beta-galactosidase of Hypocrea jecorina in cellulase induction by lactose. *Applied and Environmental Microbiology*, 71(2), pp. 851–857.

Seiboth, B. et al., 2007. The D-xylose reductase of Hypocrea jecorina is the major aldose reductase in pentose and D-galactose catabolism and necessary for beta-galactosidase and cellulase induction by lactose. *Molecular microbiology*, 66(4), pp. 890–900.

Seiboth, B. et al., 2004. The galactokinase of Hypocrea jecorina is essential for cellulase induction by lactose but dispensable for growth on d-galactose. *Molecular Microbiology*, 51(4), pp. 1015–1025.

Seiboth, B. et al., 2012. The putative protein methyltransferase LAE1 controls cellulase gene expression in *Trichoderma reesei*. *Molecular microbiology*, 84(6), pp. 1150–1164.

Seidl, V. et al., 2005. A complete survey of Trichoderma chitinases reveals three distinct subgroups of family 18 chitinases. *FEBS Journal*, 272(22), pp. 5923–5939.

Shoemaker, S. et al., 1983. Molecular cloning of exo-cellobiohydrolase I derived from *Trichoderma reesei* strain L27. *Nature Biotechnology*, 1(8), pp. 691–696.

Smyth, G.K., Michaud, J. & Scott, H.S., 2005. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*, 21(9), pp. 2067–2075.

Stalbrand, H. et al., 1995. Cloning and expression in Saccharomyces cerevisiae of a *Trichoderma reesei* beta-mannanase gene containing a cellulose binding domain. *Applied and Environmental Microbiology*, 61(3), pp. 1090–1097.

Steiger, M.G. et al., 2011. Transformation system for Hypocrea jecorina (*Trichoderma reesei*) that favors homologous integration and employs reusable bidirectionally selectable markers. *Applied and environmental microbiology*, 77(1), pp.114–121.

Sternberg, D. & Mandels, G.R., 1979. Induction of cellulolytic enzymes in *Trichoderma reesei* by sophorose. *Journal of bacteriology*, 139(3), pp. 761–769.

Sternberg, D., Vuayakumar, P. & Reese, E.T., 1977. β-Glucosidase: microbial production and effect on enzymatic hydrolysis of cellulose. *Canadian Journal of Microbiology*, 23(2), pp. 139–147.

Stricker, A.R. et al., 2008. Role of Ace2 (Activator of Cellulases 2) within the xyn2 transcriptosome of Hypocrea jecorina. *Fungal Genetics and Biology*, 45(4), pp. 436–445.

Stricker, A.R. et al., 2006. Xyr1 (Xylanase Regulator 1) regulates both the hydrolytic enzyme system and d-xylose metabolism in Hypocrea jecorina. *Eukaryotic Cell*, 5(12), pp. 2128–2137.

Stricker, A.R., Steiger, M.G. & Mach, R.L., 2007. Xyr1 receives the lactose induction signal and regulates lactose metabolism in Hypocrea jecorina. *FEBS letters*, 581(21), pp. 3915–3920.

Sun, J. et al., 2012. Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in Neurospora crassa. *Eukaryotic cell*, 11(4), pp. 482–493.

Takashima, S. et al., 1999. Molecular cloning and expression of the novel fungal β-glucosidase genes from Humicola grisea and *Trichoderma reesei*. *Journal of Biochemistry*, 125(4), pp. 728–736.

Talebnia, F., Karakashev, D. & Angelidaki, I., 2010. Production of bioethanol from wheat straw: An overview on pretreatment, hydrolysis and fermentation. *Special Issue on Lignocellulosic Bioethanol: Current Status and Perspectives*, 101(13), pp. 4744–4753.

Tamayo, E.N. et al., 2008. CreA mediates repression of the regulatory gene xlnR which controls the production of xylanolytic enzymes in Aspergillus nidulans. *Fungal Genetics and Biology*, 45(6), pp. 984–993.

Taylor, J.W. & Berbee, M.L., 2006. Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia*, 98(6), pp. 838–849.

Teeri, T., Salovuori, I. & Knowles, J., 1983. The molecular cloning of the major cellulase gene from *Trichoderma reesei*. *Nat Biotech*, 1(8), pp. 696–699.

Teeri, T.T., 1997. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends in Biotechnology*, 15(5), pp. 160–167.

Teeri, T.T. et al., 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: Gene sequence and expression of cellobiohydrolase II. *Gene*, 51(1), pp. 43–52.

Tenkanen, M., Puls, J. & Poutanen, K., 1992. Two major xylanases of *Trichoderma reesei*. *Enzyme and microbial technology*, 14(7), pp. 566–574.

Tilburn, J. et al., 1995. The Aspergillus PacC zinc finger transcription factor mediates regulation of both acid- and alkaline-expressed genes by ambient pH. *The EMBO journal*, 14(4), pp. 779–790.

Tisch, D., Kubicek, C.P. & Schmoll, M., 2011a. New insights into the mechanism of light modulated signaling by heterotrimeric G-proteins: ENVOY acts on gna1 and gna3 and adjusts cAMP levels in *Trichoderma reesei* (Hypocrea jecorina). *Fungal genetics and biology*, 48(6), pp. 631–640.

Tisch, D., Kubicek, C.P. & Schmoll, M., 2011b. The phosducin-like protein PhLP1 impacts regulation of glycoside hydrolases and light response in *Trichoderma reesei*. *BMC Genomics*, 12(1), p. 613.

Tisch, D. & Schmoll, M., 2013. Targets of light signalling in *Trichoderma reesei*. *BMC genomics*, 14(1), p. 657.

Tisch, D., Schuster, A. & Schmoll, M., 2014. Crossroads between light response and nutrient signalling: ENV1 and PhLP1 act as mutual regulatory pair in *Trichoderma reesei*. *BMC genomics*, 15(1), p. 425.

Todd, R.B., Lockington, R.A. & Kelly, J.M., 2000. The Aspergillus nidulans creC gene involved in carbon catabolite repression encodes a WD40 repeat protein. *Molecular and General Genetics*, 263(4), pp. 561–570.

Torres, C.E. et al., 2012. Enzymatic approaches in paper industry for pulp refining and biofilm control. *Applied microbiology and biotechnology*, 96(2), pp. 327–344.

Torronen, A. et al., 1992. The two major xylanases from *Trichoderma reesei*: Characterization of both enzymes and genes. *Nature Biotechnology*, 10(11), pp. 1461–1465.

Vaheri, M., Leisola, M. & Kauppinen, V., 1979. Transglycosylation products of cellulase system of *Trichoderma reesei*. *Biotechnology Letters*, 1(1), pp. 41–46.

Van den Brink, J. & de Vries, R.P., 2011. Fungal enzyme sets for plant polysaccharide degradation. *Applied microbiology and biotechnology*, 91(6), pp. 1477–1492.

Van Munster, J.M. et al., 2014. The role of carbon starvation in the induction of enzymes that degrade plant-derived carbohydrates in Aspergillus niger. *Fungal genetics and biology,* DOI: 10.1016/j.fgb.2014.04.006.

Van Peij, N.N.M.E., Visser, J. & de Graaff, L.H., 1998. Isolation and analysis of xln R, encoding a transcriptional activator co-ordinating xylanolytic expression in Aspergillus niger. *Molecular Microbiology*, 27(1), pp. 131–142.

Verbeke, J. et al., 2009. Transcriptional profiling of cellulase and expansin-related genes in a hypercellulolytic *Trichoderma reesei*. *Biotechnology Letters*, 31(9), pp. 1399–1405.

Walsh, G.A., Power, R.F. & Headon, D.R., 1993. Enzymes in the animal-feed industry. *Trends in biotechnology*, 11(10), pp. 424–430.

Wang, J. et al., 2013. Homologous constitutive expression of Xyn III in *Trichoderma reesei* QM9414 and its characterization. *Folia microbiologica*, 59(3), pp. 229–233.

Wang, S. et al., 2013. Enhancing cellulase production in *Trichoderma reesei* RUT C30 through combined manipulation of activating and repressing genes. *Journal of Industrial Microbiology & Biotechnology*, 40(6), pp. 633–641.

Witteveen, C.F.B. et al., 1989. L-Arabinose and D-Xylose Catabolism in Aspergillus niger. *Microbiology*, 135(8), pp. 2163–2171.

Wolfe, K.H. & Shields, D.C., 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, 387(6634), pp. 708–713.

Wurleitner, E. et al., 2003. Transcriptional regulation of xyn2 in Hypocrea jecorina. *Eukaryotic Cell*, 2(1), pp. 150–158.

Xiong, H. et al., 2004. Influence of pH on the production of xylanases by *Trichoderma reesei* Rut C-30. *Process Biochemistry*, 39(6), pp. 731–736.

Xu, J. et al., 1998. A third xylanase from *Trichoderma reesei* PC-3-7. *Applied Microbiology and Biotechnology*, 49(6), pp. 718–724.

Xu, J. et al., 2000. Regulation of xyn3 gene expression in *Trichoderma reesei* PC-3–7. *Applied Microbiology and Biotechnology*, 54(3), pp. 370–375.

Yamakawa, Y. et al., 2013. Regulation of cellulolytic genes by McmA, the SRF-MADS box protein in Aspergillus nidulans. *Biochemical and biophysical research communications*, 431(4), pp. 777–782.

Zeilinger, S. et al., 1993. Conditions of formation, purification, and characterization of an alpha-galactosidase of *Trichoderma reesei* RUT C-30. *Applied and Environmental Microbiology*, 59(5), pp. 1347–1353.

Zeilinger, S. et al., 1996. Different inducibility of expression of the two xylanase genes xyn1 and xyn2 in *Trichoderma reesei*. *Journal of Biological Chemistry*, 271(41), pp. 25624–25629.

Zeilinger, S., Mach, R.L. & Kubicek, C.P., 1998. Two adjacent protein binding motifs in the cbh2 (cellobiohydrolase II-encoding) promoter of the fungus Hypocrea jecorina (*Trichoderma reesei*) cooperate in the induction by cellulose. *Journal of Biological Chemistry*, 273(51), pp. 34463–34471.

Zeilinger, S.Z. et al., 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(1), pp. 56–63.

Zhang, W. et al., 2013. Two major facilitator superfamily sugar transporters from *Trichoderma reesei* and their roles in induction of cellulase biosynthesis. *The Journal of biological chemistry*, 288(46), pp. 32861–32872.

Zhou, Q. et al., 2012. Differential involvement of β-glucosidases from Hypocrea jecorina in rapid induction of cellulase genes by cellulose and cellobiose. *Eukaryotic cell*, 11(11), pp. 1371–1381.

PUBLICATION I

Re-annotation of the CAZy genes of *Trichoderma reesei* and transcription in the presence of lignocellulosic substrates

In: Microbial Cell Factories 11:134.
Copyright 2012 Authors.

Häkkinen *et al. Microbial Cell Factories* 2012, **11**:134
http://www.microbialcellfactories.com/content/11/1/134

MICROBIAL CELL FACTORIES

RESEARCH Open Access

Re-annotation of the CAZy genes of *Trichoderma reesei* and transcription in the presence of lignocellulosic substrates

Mari Häkkinen*, Mikko Arvas, Merja Oja, Nina Aro, Merja Penttilä, Markku Saloheimo and Tiina M Pakula

Abstract

Background: *Trichoderma reesei* is a soft rot Ascomycota fungus utilised for industrial production of secreted enzymes, especially lignocellulose degrading enzymes. About 30 carbohydrate active enzymes (CAZymes) of *T. reesei* have been biochemically characterised. Genome sequencing has revealed a large number of novel candidates for CAZymes, thus increasing the potential for identification of enzymes with novel activities and properties. Plenty of data exists on the carbon source dependent regulation of the characterised hydrolytic genes. However, information on the expression of the novel CAZyme genes, especially on complex biomass material, is very limited.

Results: In this study, the CAZyme gene content of the *T. reesei* genome was updated and the annotations of the genes refined using both computational and manual approaches. Phylogenetic analysis was done to assist the annotation and to identify functionally diversified CAZymes. The analyses identified 201 glycoside hydrolase genes, 22 carbohydrate esterase genes and five polysaccharide lyase genes. Updated or novel functional predictions were assigned to 44 genes, and the phylogenetic analysis indicated further functional diversification within enzyme families or groups of enzymes. GH3 β-glucosidases, GH27 α-galactosidases and GH18 chitinases were especially functionally diverse. The expression of the lignocellulose degrading enzyme system of *T. reesei* was studied by cultivating the fungus in the presence of different inducing substrates and by subjecting the cultures to transcriptional profiling. The substrates included both defined and complex lignocellulose related materials, such as pretreated bagasse, wheat straw, spruce, xylan, Avicel cellulose and sophorose. The analysis revealed co-regulated groups of CAZyme genes, such as genes induced in all the conditions studied and also genes induced preferentially by a certain set of substrates.

Conclusions: In this study, the CAZyme content of the *T. reesei* genome was updated, the discrepancies between the different genome versions and published literature were removed and the annotation of many of the genes was refined. Expression analysis of the genes gave information on the enzyme activities potentially induced by the presence of the different substrates. Comparison of the expression profiles of the CAZyme genes under the different conditions identified co-regulated groups of genes, suggesting common regulatory mechanisms for the gene groups.

Keywords: Carbohydrate active enzymes, Cellulase, Hemicellulase, Lignocellulose, Transcriptome, Transcriptional profiling, Gene regulation, Wheat, Spruce, Bagasse, Biorefinery

* Correspondence: mari.hakkinen@vtt.fi
VTT Technical Research Centre of Finland, P.O.Box 1000Tietotie 2, Espoo FI-02044VTT, Finland

BioMed Central

© 2012 Häkkinen et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background

Natural resources are diminishing while the demand for commodities, energy and food increases. This sets a requirement to find solutions for efficient utilisation of renewable biological material in production of energy and chemicals. The most abundant terrestrial renewable organic resource is lignocellulose that can be derived from industrial side-streams, municipal waste, or by-products of agriculture and forestry, and can be used as a raw material in biorefinery applications (for reviews, see [1] and [2]). It consists of cellulose, hemicellulose and lignin. Cellulose is a polymer of β-1,4-linked D-glucose units. Hemicelluloses are more heterogeneous materials that can be classified as xylans, mannans, xyloglucans and mixed-linkage glucans according to the main sugar units forming the backbone. In addition, hemicelluloses are often branched and contain side chains such as galactose, arabinan, glucuronic acid and acetyl groups. Lignin is a very resistant material that lacks a precise structure and consists of aromatic building-blocks [3]. Typically, conversion of biomass raw material includes physical and chemical pre-treatment followed by enzymatic hydrolysis of the polymers into monosaccharides and fermentation of the sugars by micro-organisms to produce higher value products, such as transport fuels and chemical feedstocks [1]. Due to the complex and heterogeneous nature of the material, both the pre-treatment method and the enzyme composition need to be adjusted according to the type of the raw material. Since the cost of enzymes is still a major limitation in the utilisation of biomass, improvement of the enzyme production systems and use of optimal mixtures of synergistic enzymes, as well as the choice of raw material and pretreatment method, are of importance in setting up a cost effective biorefinery process [4].

The CAZy database [5,6] contains information on carbohydrate active enzymes involved in breakdown, modification and synthesis of glycosidic bonds. Enzyme classes covered by CAZy classification include glycoside hydrolases (GH), carbohydrate esterases (CE), polysaccharide lyases (PL) and glycosyltransferases (GT). In addition, enzymes containing a carbohydrate binding module (CBM) are also covered. Enzymes are further classified into different families, originally based on hydrophobic cluster analysis and later on utilising sequence similarity analyses supplemented by structural information together with experimental evidence available in scientific literature. Enzymes degrading lignocellulosic plant cell wall material are known to be especially abundant in glycoside hydrolase and carbohydrate esterase families.

Trichoderma reesei (anamorph of *Hypocrea jecorina*), a mesophilic soft-rot of the phylum Ascomycota, is known for its ability to produce high amounts of cellulases and hemicellulases. It is widely employed to produce enzymes for applications in pulp and paper, food, feed and textile industries and, currently, with increasing importance in biorefining [7]. *T. reesei* produces two exo-acting cellobiohydrolases (CBHI/CEL7A and CBHII/CEL6A) [8-10]. Five endo-acting cellulases (EGI/CEL7B, EGII/CEL5A, EGIII/CEL12A, EGIV/CEL61A, EGV/CEL45A) [11-15] have been characterized, and three putative endoglucanases (CEL74A, CEL61B and CEL5B) have been found using cDNA sequencing [16]. From these enzymes CEL74A has been later characterized as a putative xyloglucanase [17], and enzymes of the glycoside hydrolase family GH61 have been shown to enhance lignocellulose degradation by an oxidative mechanism [18]. The *T. reesei* genome also encodes two characterized β-glucosidases (BGLI/CEL3A and BGLII/CEL1A) [19-22] and five predicted β-glucosidases (CEL3B, CEL3D, CEL1B, CEL3C, CEL3E) [16]. β-glucosidases hydrolyse non-reducing β-D-glucosyl residues in oligomeric cellulose degradation products and carry out transglycosylation reactions of them. Swollenin (SWOI) participates in the degradation of biomass by disrupting crystalline cellulose structure without apparent release of sugars [23]. Hemicellulases produced by *T. reesei* include four xylanases (XYNI, XYNII, XYNIII and XYNIV) [24-26], a mannanase (MANI) [27], one characterized (AXEI) [28] and one predicted (AXEII) acetyl xylan esterase [16], α-glucuronidase (GLRI) [29], one characterized (ABFI) [30] and two predicted (ABFII, ABFIII) arabinofuranosidases [16,31], three α-galactosidases (AGLI, AGLII and AGLIII) [32,33] as well as a β-xylosidase (BXLI) [30] that digests oligosaccharides derived from xylan. An acetyl esterase gene (AESI) that removes acetyl groups from hemicellulose has also been identified [34]. Glucuronoyl esterase CIPII is believed to participate in the degradation of lignocellulose biomass by cleaving ester linkages between lignin and hemicellulose and so facilitating the removal of lignin [16,35,36]. In addition to the characterized genes, several novel candidate lignocellulose degrading enzymes have been identified from the genome of *T. reesei* based on conserved domains and homology to enzymes from other fungi [37,38].

The majority of the characterised cellulase and hemicellulase genes of *T. reesei* are regulated by the type of carbon source available, in order to ensure production of hydrolytic enzymes required for degradation of the substrate and, on the other hand, to avoid energy consuming enzyme production under conditions where easily metabolisable carbon source is available. In most cases, the genes encoding the hydrolytic enzymes are repressed by glucose, and induced by various compounds derived from plant cell wall material (for reviews, see [39-41]) or their metabolic derivatives. Cellulase and hemicellulase

genes have been shown to be induced e.g. in the presence of cellulose, β-glucan, xylans and a variety of mono- and disaccharides, such as lactose, cellobiose, sophorose, L-sorbose, L-arabitol, xylobiose, cellobiose and galactose, different sets of genes being induced by the different compounds [16,42-46]. However, information on the co-induced gene groups on different substrates is still rather limited, especially regarding the novel candidate glycoside hydrolase genes identified from the genome sequence.

In this study, we have used computational methods and phylogenetic analysis to update the annotation and functional prediction of the CAZymes of *T. reesei.* Furthermore, we have analysed the expression of the CAZyme genes of *T. reesei* in the presence of different lignocellulose materials as inducing substrates in order to identify co-regulated gene groups, and to get information on the enzyme activities induced by the presence of the substrates. The selected substrates included both purified compounds and polymeric carbohydrates as well as complex lignocellulosic raw materials. The information was used to identify co-regulated gene groups and sets of genes induced by the substrates.

Results

Identification of *T. reesei* CAZyme genes

We have updated the existing annotations of CAZyme genes of *T. reesei* using computational methods and manual proofreading in order to remove discrepancies between annotations in the different genome versions [38,47] and published literature. We have also used comparative genomics and phylogenetic information to update and refine functional prediction of the encoded enzymes. The initial set of candidates for *T. reesei* CAZymes was identified by mapping the *T. reesei* proteome to CAZy database [5,6] using blastp [48], and a preliminary function prediction was assigned for the candidates based on the homologues (Additional file 1). After removal of candidates with blast hits that had apparently incorrect annotation and genes with other function predictions not related to CAZy, a total of 387 candidate genes were retrieved. Glycosyltransferase genes (99 GT genes) and the carbohydrate esterase family 10 genes (31 CE genes) were excluded from the further study since these genes mostly encode activities not involved in degradation of plant cell wall material. Furthermore, the CE10 family is no longer updated in the CAZy database. In order to verify whether the selection of candidate genes for CAZymes was supported by protein sequences from other fungi, the protein homology clusters described in [49] and updated to include 49 fungal species [50], were mapped to the CAZy database. The homology clusters were then filtered based on the average sequence identity percentage and length of the blast alignment with CAZymes (Additional file 2B and C). The clusters containing CAZymes were manually reviewed for consistency of protein domain content and quality of gene models to get reliable candidates. In addition, a few genes not fulfilling all the computational criteria were included in the further study. The proteins encoded by *T. reesei* genes 59791 and 73101 were found in protein homology clusters with no other members. The phylogeny of these two genes is discussed later. The gene 22129 was included in the study due to its previous annotation as distantly related to GH61 family [38]. In total 228 CAZyme genes remained after the computational and manual filtering.

Functional diversification of *T. reesei* CAZyme genes

CAZy family membership was assigned to the *T. reesei* gene products based on the CAZy family members of other species in the same protein homology cluster (clustering of the 49 fungal species) by majority vote. The homology clusters typically corresponded to groups of orthologous gene products supplemented with paralogues derived from gene duplications that have occurred in some sub lineage of the 49 species. *T. reesei* CAZymes were predicted to belong to 61 CAZy families (excluding CE10). The members of 27 CAZy families were divided into more than one protein homology clusters (Table 1). The fact that a family was divided into several clusters was interpreted as a sign of functional diversification within the family. In *Saccharomyces cerevisiae* most duplicated genes are derived from the genome duplication event that took place approximately 100 million years ago [51,52]. Recently, it has been experimentally shown that these duplicates have diverged in cellular, if not molecular, functions [53]. Sordariomycetes diverged from other fungi roughly 400 million years ago hence it is safe to assume that duplicate genes that were likely to exist already in the common ancestor of Sordariomycetes have had ample time to diverge functionally [54]. Thus, each protein homology cluster with multiple *T. reesei* CAZymes was searched for signs of further functional diversification within the cluster. If the phylogenetic analysis of a homology cluster suggested that the gene duplication predated the common ancestor of Sordariomycetes, it was interpreted as a sign of functional diversification between the *T. reesei* gene duplicates (Table 1).

Annotation of *T. reesei* CAZyme genes

The annotations of the CAZyme genes of *T. reesei* were specified/updated (the information is summarized in the Additional file 3). The criteria used for the annotation were the relationship of CAZy database members and *T. reesei* CAZymes in the phylogenetic trees, the best blast hits in the CAZy database, predicted functional domains

Table 1 CAZyme genes of *T. reesei*

Gene ID[a]	Name[b]	CAZy[c]	Annotation	CBM module[d]	Ref[e]	Prot. cluster[f]	Functional subgroup[g]
A.							
107268		CE1	Cand. S-formylglutathione hydrolase		This study	2607	2607a
72072		CE1	Cand. esterase		[38]	6197	6197a
107850		CE1	Cand. esterase		[47]	6197	6197a
44366		CE3	Cand. esterase		This study	2806	2806a
70021		CE3	Cand. acetyl xylan esterase		[38]	2806	2806a
41248		CE3	Cand. acetyl xylan esterase		[47]	6113	6113a
107488		CE3	Cand. esterase		This study	6113	6113a
31227		CE3	Cand. esterase/suberinase		This study	13836	13836a
67678		CE4	Cand. chitin deacetylase	CBM18	[37]	302	302a
69490		CE4	Cand. chitin deacetylase	CBM18	[37]	302	302b
65215		CE4	Cand. imidase		[38]	2445	2445a
105072		CE4	Cand. polysaccharide deacetylase		[38], A	4972	4972a
60489		CE5	Cand. cutinase		[38]	540	540a
44214	axe2	CE5	Cand. acetyl xylan esterase		[16]	865	865a
54219		CE5	Cand. acetyl xylan esterase		[31]	865	865b
73632	axe1	CE5	Acetyl xylan esterase	CBM1	[28]	865	865b
79671		CE9	Cand. N-acetyl-glucosamine-6-phosphate deacetylase		[38]	2516	2516a
3101		CE14	Cand. N-acetylglucosaminylphosphatidylinositol de-N-acetylase		This study	2076	2076a
58550		CE14	Cand. N-acetylglucosaminylphosphatidylinositol de-N-acetylase		This study	7267	7267a
123940	cip2	CE15	Glucuronoyl esterase		[16]	4243	4243a
103825		CE16	Cand. acetyl esterase		This study	3851	3851a
121418	aes1	CE16	Acetyl esterase		[34]	3851	3851a
22197	cel1b	GH1	Cand. β-glucosidase		[16]	662	662a
120749	bgl2/cel1a	GH1	β-glucosidase		[21]	662	662b
5836		GH2	Cand. β-mannosidase		[38]	609	609a
59689		GH2	Cand. β-mannosidase		[38]	609	609a
69245		GH2	Cand. β-mannosidase		[38]	609	609a
62166		GH2	Cand. β-mannosidase		[38]	609	609b
57857		GH2	Cand. β-mannosidase		[38], A	609	609c
77299	gls93	GH2	Exo-β-D-glucosaminidase		[83]	2917	2917a
76852		GH2	Cand. β-galactosidase/β-glucuronidase		[38], A	4433	4433a
102909		GH2	Cand. GH2 protein		This study	125855	125855a
121735	cel3b	GH3	Cand. β-glucosidase		[16]	110	110a
47268	bgl3i	GH3	Cand. β-glucosidase		[62]	110	110b
66832		GH3	Cand. β-glucosidase		[37]	110	110c
76227	cel3e	GH3	Cand. β-glucosidase		[16]	110	110d
76672	bgl1/cel3a	GH3	β-glucosidase		[22]	110	110e
104797	bgl3j	GH3	Cand. β-glucosidase		[62]	110	110f
46816	cel3d	GH3	Cand. β-glucosidase		[16]	132	132a
82227	cel3c	GH3	Cand. β-glucosidase		[16]	132	132b
108671	bgl3f	GH3	Cand. β-glucosidase/glucan 1,4-β-glucosidase		[62], A	132	132c

Table 1 CAZyme genes of *T. reesei* *(Continued)*

69557		GH3	Cand. β-N-acetylglucosaminidase		[38], A	2317	2317a
79669		GH3	Cand. β-N-acetylglucosaminidase		[38], A	2317	2317b
58450	xyl3b	GH3	Cand. β-xylosidase		[62]	3274	3274a
121127	bxl1	GH3	β-xylosidase		[30]	3274	3274b
64375		GH5	Cand. glucan β-1,3-glucosidase		[38]	173	173a
64906		GH5	Cand. endo-β-1,6-glucanase		[38]	173	173b
77506	cel5d	GH5	Cand. β-glycosidase		[62]	732	732a
82616	cel5b	GH5	Cand. membrane bound endoglucanase		[16]	778	778a
120312	egl2/cel5a	GH5	Endo-β-1,4-glucanase	CBM1	[15]	778	778a
56996	man1	GH5	β-Mannanase	CBM1	[27]	1854	1854a
53731		GH5	Cand. endo-β-1,4-glucanase		[37]	8196	8196a
71554		GH5	Cand. β-1,3-mannanase/endo-β-1,4-mannosidase		[38], A	16589	16589a
72567	cbh2/cel6a	GH6	Cellobiohydrolase	CBM1	[10]	2966	2966a
122081	egl1/cel7b	GH7	Endo-β-1,4-glucanase	CBM1	[11]	470	470a
123989	cbh1/cel7a	GH7	Cellobiohydrolase	CBM1	[8]	470	470b
120229	xyn3	GH10	Endo-β-1,4-xylanase		[26]	429	429a
74223	xyn1	GH11	Endo-β-1,4-xylanase		[24]	643	643a
112392	xyn5	GH11	Cand. endo-β-1,4-xylanase		[58]	643	643a
123818	xyn2	GH11	Endo-β-1,4-xylanase		[24]	643	643b
123232	egl3/cel12a	GH12	Endo-β-1,4-glucanase		[12]	1708	1708a
77284		GH12	Cand. endo-β-1,4-glucanase		[38], A	7857	7857a
59578		GH13	Cand. α-glucosidase		[38]	316	316a
108477		GH13	Cand. α-glucosidase/oligo α-glucosidase		[38]	316	316b
105956		GH13	Cand. α-amylase		[38]	394	394a
57128		GH13	Cand. glycogen debranching enzyme		[38]	1465	1465a
123368		GH13	Cand. 1,4-α-glucan branching enzyme		[38]	1522	1522a
B.							
1885	gla	GH15	Glucoamylase		[84]	608	608a
65333		GH15	Cand. α-glycosidase (Glucoamylase and related glycosyl hydrolases)		[38], A	3601	3601a
121294		GH16	Cand. glucan endo-1,3(4)-β-D-glucosidase		[38]	169	169a
38536		GH16	Cand. glucan endo-1,3(4)-β-D-glucosidase		[38]	169	169b
49274		GH16	Cand. glucan endo-1,3(4)-β-D-glucosidase		[37]	169	169c
55886		GH16	Cand. glucan endo-1,3(4)-β-D-glucosidase		[38], A	169	169c
41768		GH16	Cand. cell wall glucanosyltransferase		[37], A	248	248
58239		GH16	Cand. cell wall glucanosyltransferase		[37], A	248	248a
66843		GH16	Cand. cell wall glucanosyltransferase		[47], A	248	248b
65406		GH16	Cand. cell wall glucanosyltransferase		[37], A	248	248c
50215		GH16	Cand. endo-1,3-β-D-glucosidase/1,3-glucan binding protein		[38], A	1176	1176a
76266		GH16	Cand. cell wall glucanosyltransferase	CBM18	[38], A	2089	2089a
122511		GH16	Cand. glucan endo-1,3(4)-β-D-glucosidase		[38]	2996	2996a
123726		GH16	Cand. glucan endo-1,3(4)-β-D-glucosidase		[38]	2996	2996a
39755		GH16	Cand. glucan endo-1,3(4)-β-D-glucosidase		[38], A	3545	3545a
70542		GH16	Cand. b-glycosidase (endo-beta-1,3(4)-β-D-glucanase)		[38], A	6108	6108a

Table 1 CAZyme genes of *T. reesei* *(Continued)*

71399		GH16	Cand. endo-1,3-β-glucanase		[38], A	9096	9096a
73101		GH16	Cand. glucan endo-1,3-1,4-β-D-glucosidase		[38], A	125817	125817a
49193		GH17	Cand. glucan 1,3-β-glucosidase		[38], A	536	536a
76700		GH17	Cand. glucan 1,3 β-glucosidase/ glucan endo-1,3-β-glucosidase		[38], A	2531	2531a
39942		GH17	Cand. glucan endo-1,3-β-glucosidase		[38]	2892	2892a
66792		GH17	Cand. glucan endo-1,3-β-glucosidase		[38], A	2892	2892b
59082	chi18-2	GH18	Cand. chitinase		[59]	115	115a
62704	chi18-3	GH18	Cand. chitinase		[59]	115	115a
2735	chi18-6	GH18	Cand. chitinase		[59]	115	115b
80833	chi46	GH18	Chitinase		[85]	115	115b
81598	chi18-7	GH18	Cand. chitinase		[59]	115	115c
56894	chi18-10	GH18	Cand. chitinase	CBM18	[59]	200	200a
108346	chi18-8	GH18	Cand. chitinase	CBM18	[59]	200	200b
53949	chi18-1	GH18	Cand. chitinase	CBM18	[59]	200	200c
72339	chi18-9	GH18	Cand. chitinase	CBM18	[59]	200	200d
119859	chi18-13	GH18	Cand. chitinase		[59]	410	410
43873	chi18-12	GH18	Cand. chitinase		[59]	410	410a
110317	chi18-17	GH18	Cand. chitinase	CBM1	[59]	410	410a
68347	chi18-16	GH18	Cand. chitinase	CBM1	[59]	410	410b
124043	chi18-14	GH18	Cand. chitinase	CBM1	[59]	410	410b
66041	chi18-18	GH18	Cand. chitinase		[59]	410	410c
65162	Endo T	GH18	Endo-N-acetyl-β-D-glucosaminidase		[86]	3445	3445a
121355	chi18-rel2	GH18	Cand. Endo-N-acetyl-β-D-glucosaminidase		[38], A	3445	3445a
56448	chi18-11	GH18	Cand. chitinase		[59]	3553	3553a
62645	chi18-4	GH18	Cand. chitinase		[59]	3553	3553a
59791	chi18-15	GH18	Cand. chitinase		[59]	125792	125792a
21725		GH20	Cand. exochitinase		[38]	1025	1025a
23346	nag2	GH20	Cand. exochitinase		[38]	1025	1025a
105931		GH20	Cand. N-acetyl-β-hexosaminidase		[38]	4197	4197a
109278		GH24	Cand. lysozyme		[38]	3728	3728a
103458		GH25	Cand. N,O-diacetylmuramidase		[38]	5086	5086a
55999		GH27	Cand. α-galactosidase		[58]	617	617a
65986		GH27	Cand. α-galactosidase		[38]	617	617a
72632	agl1	GH27	α-galactosidase		[32]	617	617b
27219		GH27	Cand. α-galactosidase		[38]	12458	12458a
27259		GH27	Cand. α-galactosidase		[38]	12458	12458a
59391		GH27	Cand. α-galactosidase		[38]	12458	12458a
72704	agl3	GH27	α-galactosidase		[32]	12458	12458a
75015		GH27	Cand. α-galactosidase		[38]	12458	12458a
70186		GH28	Cand. polygalacturonase/xylogalacturonan hydrolase		[38]	295	295a
112140	pgx1	GH28	Cand. exo-polygalacturonase		[58]	295	295b
122780	rgx1	GH28	Cand. exo-rhamnogalacturonase		[58]	295	295c
103049		GH28	Cand. endo-polygalacturonase		[38]	870	870a
69276		GH30	Cand. endo-β-1,4-xylanase		[38]	6176	6176a

Table 1 CAZyme genes of *T. reesei* *(Continued)*

111849	xyn4	GH30	Endo-β-1,4-xylanase		[87]	6176	6176a
3094		GH30	Cand. glucan endo 1,6-β-glucanase		[38]	7813	7813a
69736		GH30	Cand. glucan endo 1,6-β-glucanase		[38]	7813	7813a
110894		GH30	Cand. endo-β-1,6-galactanase		[38]	16621	16621a
82235		GH31	Cand. α-glucosidase		[38]	594	594a
121351	gls2	GH31	Glucosidase II alpha subunit		[88]	1379	1379a
69944		GH31	Cand. α-xylosidase/α-glucosidase		[38]	3846	3846a
60085		GH31	Cand. α-glucosidase		[38]	5176	5176a
80240	bga1	GH35	β-galactosidase		[89]	675	675a
64827		GH36	Cand. raffinose synthase domain protein		[38], A	3772	3772a
124016	agl2	GH36	α-galactosidase		[32]	4771	4771a
120676		GH37	Cand. α,α-trehalase		[38]	769	769a
123226		GH37	Cand. α,α-trehalase		[38]	4723	4723a
3196		GH38	Cand. α-mannosidase		[38]	1337	1337a
73102		GH39	Cand. β-xylosidase		[38], A	4388	4388a
3739		GH43	Cand. β-xylosidase/α-L-arabinofuranosidase		[38]	586	586a
68064		GH43	Cand. β-xylosidase/α-L-arabinofuranosidase		[47]	7306	7306a
49976	egl5/cel45a	GH45	Endo-β-1,4-glucanase	CBM1	[14]	12601	12601a
C.							
45717	mds1	GH47	α-1,2-mannosidase		[90]	70	70a
2662		GH47	Cand. α-1,2-mannosidase		[38]	70	70b
111953		GH47	Cand. α-1,2-mannosidase		[38]	70	70c
79960		GH47	Cand. α-1,2-mannosidase		[47]	70	70d
65380		GH47	Cand. α-1,2-mannosidase		[38]	70	70e
79044		GH47	Cand. α-1,2-mannosidase		[38]	70	70e
22252		GH47	Cand. α-1,2-mannosidase		[37]	70	70f
64285		GH47	Cand. α-mannosidase		[38]	1421	1421a
55319	abf3	GH54	Cand. α-L-arabinofuranosidase	CBM42	[31]	5126	5126a
123283	abf1	GH54	α-L-arabinofuranosidase I	CBM42	[30]	5126	5126a
121746	gluc78	GH55	Cand. exo-1,3-β-glucanase		[58]	235	235a
54242		GH55	Cand. β-1,3-glucanase		[38]	235	235b
70845		GH55	Cand. β-1,3-glucanase		[38]	235	235b
108776		GH55	Cand. β-1,3-glucanase		[38]	235	235b
56418		GH55	Cand. β-1,3-glucanase		[38]	235	235c
73248		GH55	Cand. exo-1,3-β-glucanase		[38]	235	235d
73643	egl4/cel61a	GH61	Cand. copper-dependent polysaccharide monooxygenase/endo-β-1,4-glucanase	CBM1	[13]	77	77a
120961	cel61b	GH61	Cand. copper-dependent polysaccharide monooxygenase		[16]	77	77b
22129		GH61	Cand. copper-dependent polysaccharide monooxygenase		[38]	515	515a
31447		GH61	Cand. copper-dependent polysaccharide monooxygenase		This study	515	515b
76065		GH61	Cand. copper-dependent polysaccharide monooxygenase		This study	515	515b
27554		GH61	Cand. copper-dependent polysaccharide monooxygenase		[31]	8230	8230a

Table 1 CAZyme genes of *T. reesei* *(Continued)*

76210	abf2	GH62	Cand. α-L-arabinofuranosidase		[16]	3103	3103a
22072		GH63	Cand. processing α-glucosidase		[38]	1244	1244a
75036		GH63	Cand. α-glucosidase		[38], A	2428	2428a
65137		GH64	Cand. endo-1,3-β-glucanase		[38]	3917	3917a
123639		GH64	Cand. endo-1,3-β-glucanase		[38]	3917	3917a
124175		GH64	Cand. endo-1,3-β-glucanase		[38]	3917	3917b
25224		GH65	Cand. α,α-trehalase		[38]	2891	2891a
123456		GH65	Cand. α,α-trehalase		[38]	2891	2891a
72526	glr1	GH67	α-Glucuronidase		[29]	5113	5113a
71532		GH71	Cand. α-1,3-glucanase		[37]	287	287a
108672		GH71	Cand. α-1,3-glucanase	CBM24	[38]	287	287b
120873		GH71	Cand. α-1,3-glucanase	CBM24	[38]	287	287b
73179		GH71	Cand. α-1,3-glucanase		[37]	287	287c
22914		GH72	Cand. β-1,3-glucanosyltransferase	CBM43	[38]	111	111a
82633		GH72	Cand. β-1,3-glucanosyltransferase		[37]	111	111a
77942		GH72	Cand. β-1,3-glucanosyltransferase		[38]	111	111b
78713		GH72	Cand. β-1,3-glucanosyltransferase		[47]	111	111c
123538		GH72	Cand. β-1,3-glucanosyltransferase		[38]	111	111d
49081	cel74a	GH74	Xyloglucanase	CBM1	[16]	4769	4769a
42152		GH75	Cand. chitosanase		[38]	3832	3832a
70341		GH75	Cand. chitosanase		[38]	3832	3832a
66789		GH75	Cand. chitosanase		[38]	3832	3832b
49409		GH76	Cand. α-1,6-mannanase		[38]	129	129a
67844		GH76	Cand. α-1,6-mannanase		[38]	129	129a
69123		GH76	Cand. α-1,6-mannanase		[38]	129	129b
122495		GH76	Cand. α-1,6-mannanase		[38]	129	129b
53542		GH76	Cand. α-1,6-mannanase		[38]	129	129c
55802		GH76	Cand. α-1,6-mannanase		[38]	129	129c
74807		GH76	Cand. α-1,6-mannanase		[47]	3765	3765a
27395		GH76	Cand. α-1,6-mannanase		[38]	4345	4345a
58887		GH78	Cand. α-L-rhamnosidase		[38]	3807	3807a
71394		GH79	Cand. β-glucuronidase		[38], A	5190	5190a
73005		GH79	Cand. β-glucuronidase		[38], A	5190	5190a
106575		GH79	Cand. β-glucuronidase		[38], A	5190	5190a
72568		GH79	Cand. β-glucuronidase		[37], A	5190	5190b
73256		GH81	Cand. endo-1,3-β-glucanase		[38]	722	722
79602		GH81	Cand. endo-1,3-β-glucanase		[38]	722	722a
58117		GH89	Cand. α-N-acetylglucosaminidase		[38]	6562	6562a
69700		GH89	Cand. α-N-acetylglucosaminidase		[38]	6562	6562a
74198		GH92	Cand. α-1,2-mannosidase		[38]	318	318a
111733		GH92	Cand. α-1,2-mannosidase		[38]	318	318a
79921		GH92	Cand. α-1,2-mannosidase		[38]	318	318b
60635		GH92	Cand. α-1,2-mannosidase		[38]	318	318c
55733		GH92	Cand. α-1,2-mannosidase		[38]	318	318d
57098		GH92	Cand. α-1,2-mannosidase		[47]	318	318e

Table 1 CAZyme genes of *T. reesei* *(Continued)*

69493		GH92	Cand. α-1,2-mannosidase	[38]	318	318f
72488		GH95	Cand. α-L-fucosidase	[38]	2951	2951
5807		GH95	Cand. α-L-fucosidase	[38]	2951	2951a
111138		GH95	Cand. α-L-fucosidase	[37]	2951	2951a
58802		GH95	Cand. α-L-fucosidase	[38]	2951	2951b
4221		GH105	Cand. rhamnogalacturonyl hydrolase	This study	4036	4036a
57179		GH105	Cand. rhamnogalacturonyl hydrolase	[37]	4065	4065a
79606		GH115	Cand. xylan-α-1,2-glucuronidase or α-(4-O-methyl)-glucuronidase	This study	3202	3202a
103033		PL7	Cand. alginate lyase	[38]	9526	9526a
110259		PL7	Cand. alginate lyase	[38]	9526	9526a
111245		PL8	Cand. chondroitin lyase	[38]	10699	10699a
53186	TrGl	PL20	Glucuronan hydrolase	[91]	5536	5536a
69189		PL20	Cand. endo-β-1,4-glucuronan lyase	This study	5536	5536a
108348		GH	-		10159	
105288		GH	-		29033	
121136		GH	-		29033	

Glycoside hydrolase, carbohydrate esterase (excluding CE10) and polysaccharide lyase genes of *T. reesei*, the annotation and functional diversification of the genes (a), gene identifier as in *T. reesei* v2.0 data base [38]; (b), name given to the gene in the publication/data base marked in the reference column; (c), family and class of the enzyme according to the CAZy classification [5]; (d), cellulose binding module present in the protein; (e), reference to previous studies or to *T. reesei* database versions 1.2 and 2.0.; (f), protein cluster the *T. reesei* protein was assigned to when the protein clusters were mapped to CAZy database by a blast search; (g), functional subgroups within the protein cluster determined according to phylogenetic analysis. A, a previous annotation has been specified/updated during this study.

in the proteins ([55,56], Additional file 4), together with the function predictions in the *T. reesei* v2.0 data base [37,38]. The list of the 49 fungi in the protein homology clusters, with their abbreviations are shown in Additional file 5. For selected cases, alignment of the candidate CAZyme against the PFAM profile of the CAZy family was also used to confirm the CAZy family of the candidate gene. By these means, the *T. reesei* CAZyme genes, excluding glycosyltransferase and the family CE10 genes, were concluded to contain 201 GH (glycoside hydrolase) genes, 22 CE (carbohydrate esterase genes) and 5 PL (polysaccharide lyase) genes. The outcome of the annotation process and the functional diversification observed is summarised in Table 1. Phylogenetic trees showing the functional diversifications discussed in more detail, are shown in Additional file 6. In this study we focus on putative CAZymes involved in degradation of plant cell wall derived material, whereas proteins with other functions such as glycosylation or degradation of cell wall components were given less attention. Especially, the function predictions obtained for genes without previous annotation, or genes for which only a general prediction was available, are discussed.

Cellulase protein homology clusters

A good example to support the functional diversification predictions based on a phylogenetic analysis was the diversification of the well characterised GH7 genes, the cellobiohydrolase *cbh1* and endoglucanase *egl1*. The encoded proteins are assigned to the same protein homology cluster but different functional subgroups in accordance with the known enzymatic activities of the proteins (Table 1A, Additional file 6). The third major component of the cellulolytic system, cellobiohydrolase CBHII belongs to the family GH6 and is the only member of the family in *T. reesei*. Endoglucanases can be found also in the glycoside hydrolase families GH5, GH12 and GH45. GH5 includes the characterised EGII, a candidate membrane-bound endoglucanase CEL5B in the same homology cluster and functional subgroup, and an additional endoglucanase candidate (53731) in a separate cluster (Table 1A). The other GH5 members include the β-mannanase MANI, and a candidate for a β-1,3-mannanase/endo-β-1,4-mannosidase (71554), glucan β-1,3-glucosidase (64375), endo-β-1,6-glucanase (64906) and a β-glycosidase (77506) without a specific functional prediction. GH12 family includes the characterised endoglucanase EGIII together with a candidate endoglucanase (77284) but in separate protein homology clusters (Table 1A). The endoglucanase EGV of the family GH45, as well as the xyloglucanase CEL74A of the family 74, are the only members of their families in *T. reesei*. GH61 family member EGIV/CEL61A has previously been described as an endoglucanase [13]. However, recently the family has been suggested to act in the degradation of lignocellulose material via an oxidative

mechanism [18]. Based on our analysis, *T. reesei* genome harbours five candidate GH61 members in addition to EGIV/CEL61A. The encoded proteins are divided into three protein homology clusters and four functional subgroups inside the clusters (Table 1C).

In the case of β-glucosidases, experimental evidence supporting functional differences has only been obtained for BGLI and BGLII, the first being the major extracellular β-glucosidase and the second being an intracellular enzyme. However, phylogenetic analysis suggests that further functional differences may exist within this group of enzymes (Table 1A). Altogether, the *T. reesei* genome encodes eleven characterised or predicted β-glucosidases, two belonging to the family GH1 and nine to the family GH3. The intracellular GH1 β-glucosidases of *T. reesei* (BGLII and CEL1B, the latter predicted to be intracellular due to the lack of predicted signal sequence) are in the same protein homology cluster but in different functional subgroups. The β-glucosidases of family GH3 are divided into two homology clusters, and furthermore, showed diversification within the clusters, so that the genes could be assigned to nine groups. Functional diversification of the GH3 β-glucosidases within the same protein homology cluster is visualised in Figure 1 and Additional file 6. The predicted β-glucosidases 47268, 66832 and 104797 are assigned to the same cluster as BGLI, CEL3B and CEL3E, and 108671 to the same cluster as CEL3C and CEL3D. The GH3 β-glucosidases BGLI, CEL3E, CEL3B, 66832, 104797 and 108671 are predicted to be secreted according to the signal sequence prediction (SignalP 4.0, [57]).

Hemicellulase and other CAZyme protein homology clusters

In addition to β-glucosidases, the family GH3 includes candidate β-N-acetylglucosaminidases (69557 and 79669) and β-xylosidases (BXLI and candidate 58450). The β-N-acetylglucosaminidases and β-xylosidases are found in separate protein homology clusters according to their (predicted) functions, and the β-xylosidases are also in separate functional subgroups (Table 1A). In addition to the GH3 β-xylosidases, *T. reesei* is predicted to encode a candidate β-xylosidase (73102) of the family GH39, and two proteins of the family GH43 predicted to have either β-xylosidase or α-L-arabinofuranosidase activity (68064, 3739), all in separate protein homology clusters (Table 1B). The characterised arabinofuranosidase ABFI and the candidate enzyme ABFIII of the family GH54 are not functionally diversed from each other according to the functional diversification criteria (Table 1C). In addition to these, the *T. reesei* genome also encodes a candidate arabinofuranosidase ABFII of the family GH62.

T. reesei has four characterized xylanases and two candidate xylanases. The characterized xylanases belong to families GH10 (XYNIII), GH11 (XYNI, XYNII) and GH30 (XYNIV). Candidate xylanases can be found from families GH11 (112392) and GH30 (69276). The candidate xylanase 112392 (also called XYNV in a recent publication, [58]) is assigned to the same protein homology cluster as XYNI and XYNII and is in the same functional subgroup as XYNI (Table 1A). Candidate xylanase 69276 is in the same homology cluster and functional subgroup as XYNIV (Table 1B).

The carbohydrate esterase families CE3 and CE5 contain known and/or candidate acetyl xylan esterases of *T. reesei*. The carbohydrate esterase family 5 includes both acetyl xylan esterases and cutinases (Table 1A). The candidate acetyl xylan esterases (54219 and AXEII) are in the same protein homology cluster with the characterized enzyme AXEI, the 54219 belonging to the same functional subgroup as AXEI. The candidate cutinase (60489) clusters together with known and candidate cutinase homologues of other fungi. Our study also revealed candidate members of CE3 family, including candidates for acetyl xylan esterases or esterase/suberinase. However, the ORF prediction (in the genome database [38]) for majority of the genes is unclear due to difficulties in prediction of N- or C-termini or intron positions, which hampers phylogenetic analysis of the genes.

Hemicellulases of *T. reesei* are also present in families GH2 and GH27. The family GH2 is one of the families including members with versatile enzymatic activities. The predicted members of GH2 of *T. reesei* include five candidate β-mannosidases (5836, 69245, 59689, 57857 and 62166) belonging to the same protein homology cluster but to three different functional subgroups within the cluster (Table 1A). GH2 members include also a candidate exo-β-D-glucosaminidase (77299), and a candidate enzyme with a predicted function as a β-galactosidase or β-glucuronidase (76852). The six candidate α-galactosidases (27219, 27259, 59391, 75015, 55999 and 65986) of family GH27 are divided to two protein homology clusters (Table 1B). Proteins encoded by genes 27219, 27259, 59391 and 75015 are assigned to the same cluster as AGLIII and are not functionally diversified either from AGLIII or from each other (cluster contained only *T. reesei* proteins). The remaining candidate α-galactosidases are in the same cluster as AGLI and are divided to two functional subgroups within the cluster.

Our study also suggested a new member of CE16 family and of PL20 family belonging to the same functional subgroup as the characterised AESI and TRGL, respectively (Table 1A and 1C). In addition to the known GH67 α-glucuronidase (GLRI), a GH115 type of α-

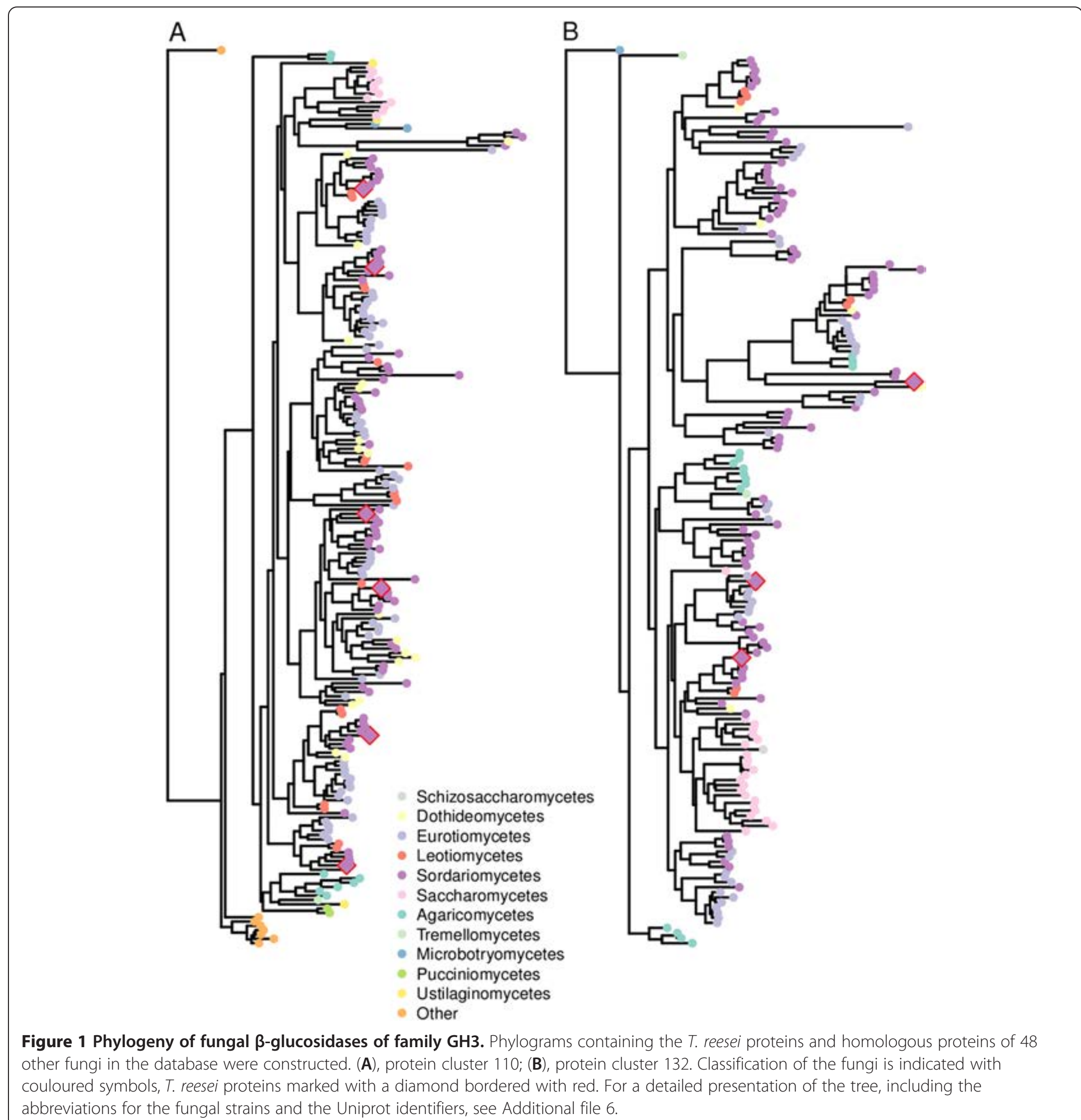


Figure 1 Phylogeny of fungal β-glucosidases of family GH3. Phylograms containing the *T. reesei* proteins and homologous proteins of 48 other fungi in the database were constructed. (**A**), protein cluster 110; (**B**), protein cluster 132. Classification of the fungi is indicated with couloured symbols, *T. reesei* proteins marked with a diamond bordered with red. For a detailed presentation of the tree, including the abbreviations for the fungal strains and the Uniprot identifiers, see Additional file 6.

glucuronidase (α-1,2- or α-(4-O-methyl)-glucuronidase) was predicted (79606). The *T. reesei* genome was also found to encode four candidate GH79 β-glucuronidases (71394, 106575, 72568, 73005) not identified previously, but which are probably involved in proteoglycan hydrolysis rather than lignocellulose degradation. Also an additional member of GH105 family, a predicted rhamnogalacturonyl hydrolase (4221), was identified in the study in addition to the previously predicted one (57179).

Comparison of *T. reesei* CAZyme homology clusters with other fungi

Comparison of *T. reesei* protein homology clusters with other fungi by looking at the number of genes per species in the clusters, revealed several interesting differences (Additional file 7). The cluster containing AGLIII and four candidate α-galactosidases is unique to *T. reesei*. This protein homology cluster is not found from any other of the 48 fungi included in this study. The cluster containing four candidate β-glucuronidases from family

GH79 is expanded in *T. reesei* as compared to other fungi. *T. reesei* has 4 genes encoding these proteins while the other 48 fungi in this study contained only 0 to 3 genes. Similarly, the cluster containing candidate extracellular or membrane bound chitinases is expanded to include 6 genes in total, while in most of the 48 fungi there are 0 to 5 genes. This cluster corresponds to phylogenetic group B as described in a previous publication on chitinase phylogeny [59]. The cluster also contains chitinase CHI18-18 which was not included in group B. Expansion of GH18 genes of *T. reesei* has been described previously and it has been suggested to be involved in functions related to pathogenicity to other fungi [37].

The cluster containing seven genes from family GH92 is not found in *Fusarium* species that are the closest relatives of *Trichoderma* in our data set. It is possible that *Fusarium* genes have diversified further apart while *Trichoderma* genes have retained the ancestral functions. One of the two clusters that contain genes encoding members of family GH43 is hugely reduced in *T. reesei* compared to other Pezizomycotina species, especially *Fusarium* spp. *T. reesei* has only one gene in this cluster while *Fusarium oxysporum* has 12. Reduction is also visible in two protein homology clusters containing members from the family GH61. The reduction is especially notable in the cluster number 77 where *T. reesei* has only two genes while the number of genes in other Pezizomycotina can be as high as 43. These two reductions were already noticed during the initial genome analysis of *T. reesei* [37].

Horizontal gene transfer

Our study revealed also several cases of putative horizontal gene transfer from bacteria. As mentioned above, proteins encoded by genes 59791 and 73101 were assigned to protein homology clusters with no other fungal proteins i.e. they had no significant homologues in the fungal genomes included in the clustering. However, the proteins had homologues in the closely related *Hypocrea* (*Trichoderma*) species as well as a large number of bacterial homologues. 73101 had several bacterial GH16 family proteins as homologues, and a candidate endo-β-1,3-1,4-glucanase of *Bacillus subtilis* as the best blast hit in the CAZy database. 59791was closely related to GH18 family chitinases of *Hypocrea* (*Trichoderma*) species, and had homologues also especially in *Streptomyces* and *Bacillus* species. The phylogeny of the genes thus suggests possible gene transfer from bacteria. The possibility of horizontal gene transfer of 59791 has also been discussed in a previous publication [59]. In addition, a candidate β-glucosidase (108671) of GH3 was shown to be possibly a result of horizontal gene transfer. Although the protein is assigned to the same homology cluster as CEL3D and CEL3C beside proteins from other *Trichoderma* species, the closest homologous for this protein are from bacteria. The phylogenetic analysis of these genes are represented in additional files (59791 in Additional file 8, 73101 in Additional file 9, and 108671 in Additional file 10).

Identification of similarly expressed genes by visualization of the expression data

The induction of *T. reesei* CAZyme genes was studied by cultivating the fungus in the presence of different inducing substrates and analysing the transcriptional responses at different time points of induction using oligonucleotide microarrays. The transcript signals in induced cultures were compared to the ones in uninduced control cultures at the same time point. The microarray expression data on the CAZyme genes is represented in detail in the Additional file 11 (including the normalised log2 scale signal intensity of the genes in each condition, the fold change of the signal in the induced culture as compared to the uninduced control cultures at the same time point, and statistical significance of the difference in expression as compared to the control cultures at a corresponding time point). The substrates used in the study were differentially pretreated bagasse materials (ground bagasse without further pretreatment, bagasse pretreated by steam explosion, and bagasse hydrolysed enzymatically after steam explosion pretreatment), pretreated wheat straw and pretreated spruce, birch xylan, oat spelt xylan, Avicel cellulose and sophorose. Based on liquid chromatographic analysis of the carbohydrate content of the pretreated complex substrates (Table 2) and information obtained from the manufacturers of the other substrates used, bagasse and

Table 2 Carbohydrate composition of the pre-treated biomass substrates

Sample	Rhamnose	Arabinose	Galactose	Glucose	Xylose	Mannose	Fructose
BS	< 0.1	0.54	0.15	59.5	12.59	0.24	< 0.1
BE	< 0.1	0.62	0.14	33	9.68	0.38	< 0.1
WH	< 0.1	0.38	< 0.1	61.6	2.96	0.12	-
SP	< 0.1	< 0.1	< 0.1	46	< 0.1	< 0.1	< 0.1

The amounts of the different carbohydrates are shown as mg/100 mg or dry matter. BS, steam exploded bagasse; BE, enzymatically hydrolysed bagasse; WH, wheat straw; SP, spruce.

wheat straw are the most complex ones. Steam exploded wheat straw and pretreated bagasse material contain cellulose and arabinoxylan, but also polysaccharides with other substitutions, such as galactose or mannose units, whereas pretreated spruce consists mostly of cellulose. Sophorose is a disaccharide, 2-O-β-D-Glucopyranosyl-α-D-Glucose, that can be produced via transglycosylation reaction from cellobiose, a cleavage product of cellulose.

The majority (179 genes) of the 228 CAZyme genes (excluding GT and CE10 family genes) were induced by at least one of the substrates used (higher expression level in the induced cultures as compared to the uninduced cultures at the same time point). The largest number of genes and CAZy families induced was detected in the cultures with the hemicellulosic material, bagasse, xylans and wheat straw (68–124 genes in 39–47 CAZy families), whereas cultivation in the presence of cellulosic or cellulose derived materials, Avicel cellulose, pretreated spruce, or sophorose, resulted in a clearly smaller number of genes induced (43–58 genes in 28–36 families). The number of induced genes within each CAZy family is represented in Figure 2.

Based on the microarray data, the common core of genes induced in the presence of the lignocellulose substrates (as judged by induction of gene expression by both cellulose and xylan, and induction in the presence of at least 70% of the substrates used) included genes encoding characterised or predicted functions as GH6 cellobiohydrolase, GH5 endoglucanase, xylanases of families GH10, GH11 and GH30, GH5 β-mannanase, GH3 family β-glucosidases and β-xylosidases, GH27 α-galactosidases, GH2 β-mannosidases, acetyl xylan esterases of families CE3 and CE5, glucuronoyl and acetyl esterases of families CE15 and CE16, GH31 α-glucosidases/α-xylosidases, GH54 and GH43 α-L-arabinofuranosidases (or β-xylosidase/α-L-arabinofuranosidases), GH61 polysaccharide monooxygenases, GH55 β-1,3-glucanases, GH67 α-glucuronidase, GH79 β-glucuronidase, GH105 rhamnogalacturonyl hydrolase, GH95 α-L-fucosidase, GH89 α-N-acetylglucosaminidase, and chitinases of families GH18 and GH20.

The analysis revealed also a more refined pattern of co-expressed genes. To visualise the co-expressed gene groups as well as differences in the induction pattern of the genes in the presence of different substrates, a heatmap representation was used (Figure 3). The heatmap was generated using the fold changes of the transcript signals in the induced cultures vs. the uninduced cultures at the same time point for each of the genes.

In the heatmap, the branches A, B, C and D represent genes that are induced by the presence of most of the substrates, and include many of the known cellulolytic and hemicellulolytic genes. Branch A contains genes that are rather evenly induced by all the substrates (*egl4/cel61a*, *bxl1*, *xyn1*, *xyn4*). The genes give strong signals in the inducing conditions, and have moderately high basal signal levels also in the uninduced conditions. The genes in branch B have an especially pronounced induction by the presence of Avicel cellulose and wheat, but reduced expression at the late time points of xylan cultures (candidate GH28 exo-polygalacturonase *pgx1*, candidate GH3 β-glucosidase *cel3d*, GH5 β-mannanase *man1*, candidate GH2 β-mannosidase, candidate GH61 polysaccharide monooxygenase *cel61b*). The genes in branch C are moderately induced by most of the substrates, but similarly to the branch B, many of the genes show reduction of the signal at the late time points of xylan cultures (candidate GH3 β-glucosidase *bgl3f*, candidate CE3 acetyl xylan esterase 41248, candidate GH18 chitinase *chi18-7*, CE16 acetyl esterase *aes1*, candidate GH105 rhamnogalacturonyl hydrolase 57179, candidate GH95 α-L-fucosidase 5807, GH27 α-galactosidase *agl3*, candidate GH55 β-1,3-glucanases 56418 and 54242). The branch D represents genes induced by most of the substrates but especially by the presence of xylans, steam exploded bagasse and wheat straw. In accordance with the strong induction by xylans, the branch D includes both known and candidate hemicellulase genes (GH67 α-glucuronidase *glr1*, GH11 endo-β-1,4-xylanase *xyn2*, candidate GH11 endo-β-1,4-xylanase *xyn5*, CE5 acetyl xylan esterase *axe1* and a candidate CE5 acetyl xylan esterase).

Other distinctive groups of co-regulated genes are induced by subset of the substrates studied. The branch F is characterised by genes giving rather low or moderate signals in the array analysis and showing induction by birch xylan, steam exploded and enzymatically treated bagasse already at the early time points. This group includes e.g. candidates of GH16, GH18, GH27 and GH92 family genes among others. The genes in branch E show moderate expression and induction levels, mostly in the presence of wheat straw, but subgroups of the branch also with bagasse, spruce, oat spelt xylan or sophorose. The branch G represents genes induced especially by sophorose. This group contains many genes known to encode enzymes with activity on lignocellulose substrates, (GH3 β-glucosidase *bgl1/cel3a*, a candidate GH3 β-glucosidase *cel3e*, GH7 endo-β-1,4-glucanase *egl1/cel7b*, GH10 endo-β-1,4-xylanase *xyn3*, GH74 xyloglucanase *cel74a*, CE15 glucuronoyl esterase *cip2*) as well as candidates for GH30 endo-β-1,4-xylanase, and GH79 β-glucuronidase among others. Branch H contain genes that have moderately high signal levels and are induced in the presence of the cellulosic substrates, spruce, sophorose and Avicel, but whose signal levels are reduced in the presence of pretreated bagasse, wheat straw and birch xylan. This group contains especially candidate α-1,2-mannosidase genes of the family GH47, but also genes of families GH3 (β-glucosidase *cel3c)*,

	BO	BS	BE	WH	XO	XB	AV1	AV0.75	SP	SO	Induced by LC	Genes in genome
GH18	4	9	13	3	3	8	3	3	2	3	16	20
GH3	9	7	10	5	4	7	8	6	4	8	12	13
GH16	6	4	6	6	3	3	0	0	3	1	10	16
GH2	4	6	6	2	1	3	3	1	2	2	8	8
GH27	4	5	6	4	3	2	3	1	4	3	7	8
GH92	2	3	4	1	4	3	1	0	0	1	7	7
GH55	3	4	4	3	4	6	3	0	3	0	6	6
GH5	3	3	4	3	1	3	3	3	3	2	5	8
GH30	3	4	5	3	3	3	2	2	2	3	5	5
GH61	3	3	3	3	3	3	3	2	1	2	5	6
GH76	1	1	1	2	0	2	0	0	1	1	5	8
CE3	1	2	3	1	2	2	1	0	1	1	4	5
CE5	2	2	3	4	3	2	2	2	1	1	4	4
GH31	3	2	2	2	1	3	3	1	2	2	4	4
GH79	1	3	3	1	2	4	0	1	0	1	4	4
CE1	0	1	2	1	0	2	0	0	0	1	3	3
GH11	3	3	3	3	3	3	3	3	2	3	3	3
GH28	2	3	3	3	1	2	1	1	2	1	3	4
GH47	0	0	0	1	1	1	0	1	0	1	3	8
GH64	0	1	2	0	0	3	0	0	0	1	3	3
GH71	1	3	2	1	2	1	1	0	1	0	3	4
GH75	1	1	3	1	2	1	0	0	0	0	3	3
GH95	2	1	2	2	2	1	1	1	0	2	3	4
CE4	1	2	2	0	0	2	0	0	0	0	3	4
CE14	1	1	1	1	0	0	0	0	1	0	2	2
CE16	1	1	1	1	1	1	1	1	1	1	2	2
GH1	0	0	0	1	0	0	0	1	1	1	2	2
GH7	2	2	2	1	0	0	2	2	2	2	2	2
GH13	1	0	1	0	1	1	2	0	0	0	2	5
GH15	1	0	1	1	0	1	0	1	0	1	2	2
GH20	1	2	2	2	2	1	1	0	1	0	2	3
GH36	2	0	2	0	1	1	0	0	1	0	2	2
GH43	1	0	1	2	1	1	1	1	1	1	2	2
GH54	1	1	1	2	1	1	2	1	2	1	2	2
GH89	1	1	2	1	1	0	1	1	1	0	2	2
GH105	1	1	1	1	1	2	0	1	1	1	2	2
PL20	0	1	1	1	1	1	1	0	0	0	2	2
PL7	1	2	2	1	1	1	0	0	0	0	2	2
CE9	1	0	0	1	1	1	1	0	1	0	1	1
CE15	1	1	0	1	1	1	1	1	0	1	1	1
GH6	1	1	1	1	1	0	1	1	1	1	1	1
GH10	1	1	1	1	0	1	1	1	0	1	1	1
GH12	0	1	1	0	0	1	0	1	0	1	1	2
GH24	1	0	1	0	1	0	0	0	0	0	1	1
GH25	1	1	1	0	1	1	0	0	0	0	1	1
GH35	1	1	0	1	1	1	0	0	0	1	1	1
GH39	0	1	1	0	0	1	0	0	0	0	1	1
GH45	1	1	0	1	0	0	1	1	1	1	1	1
GH62	1	1	1	1	0	1	0	0	0	1	1	1
GH63	1	0	0	0	0	0	0	0	0	0	1	2
GH67	1	1	1	1	1	1	1	1	1	1	1	1
GH72	0	0	0	1	0	0	0	0	0	0	1	5
GH74	1	0	0	0	1	1	0	0	0	0	1	1
GH78	0	1	1	0	0	1	0	0	0	0	1	1
GH81	0	0	1	0	0	0	0	0	0	0	1	2
GH115	0	0	1	1	0	0	0	0	0	1	1	1
PL8	1	0	0	1	1	1	0	0	1	0	1	1
GH17	0	0	0	0	0	0	0	0	0	0	0	4
GH37	0	0	0	0	0	0	0	0	0	0	0	2
GH38	0	0	0	0	0	0	0	0	0	0	0	1
GH65	0	0	0	0	0	0	0	0	0	0	0	2
GH, not defined	1	2	3	1	0	1	0	0	0	0	3	3
No of genes	87	98	124	82	68	94	58	43	51	57	179	228
No of families	45	43	47	45	39	46	30	28	31	36	57	61

Figure 2 CAZy family members induced in the presence of the different lignocellulose substrates. The number of genes in each CAZy family induced in the presence of the different lignocellulose substrates: BO, ground bagasse; BS, steam exploded bagasse; BE, enzymatically hydrolysed steam exploded bagasse; XO, oat spelt xylan; XB, birch xylan; AV1, 1% Avicel cellulose; AV0.75, 0.75% Avicel cellulose; WH, steam exploded wheat straw; SP, steam exploded spruce; SO, sophorose. The intensity of the colour represents the amount of genes induced. The total number of genes in each CAZy family induced by at least one of the substrates, and the total number of the CAZy family members in *T. reesei* genome are shown on the right. The total number of genes induced by the substrate and the number of CAZy families the genes belong to, are shown at the bottom. LC, lignocellulose substrate.

GH5, GH16, GH17, GH27, GH55, GH76, and GH95. The branch I contains genes with the strongest induction at the late time point of xylan cultures (a candidate GH3 β-xylosidase *xyl3b*, a candidate GH3 β-glucosidase *cel3b*, GH36 α-galactosidase *agl2*, a candidate GH79 β-glucuronidase, a candidate PL7 polysaccharide lyase, chitinase genes of families GH18 and GH20 and a candidate GH31 α-glucosidase among others). The branch J contains genes with rather strong constitutive signals. Among others, the group contains both GH1 family β-glucosidase genes, a candidate GH3 family β-glucosidase gene, two GH5 family genes (*cel5b* and *cel5d*), a

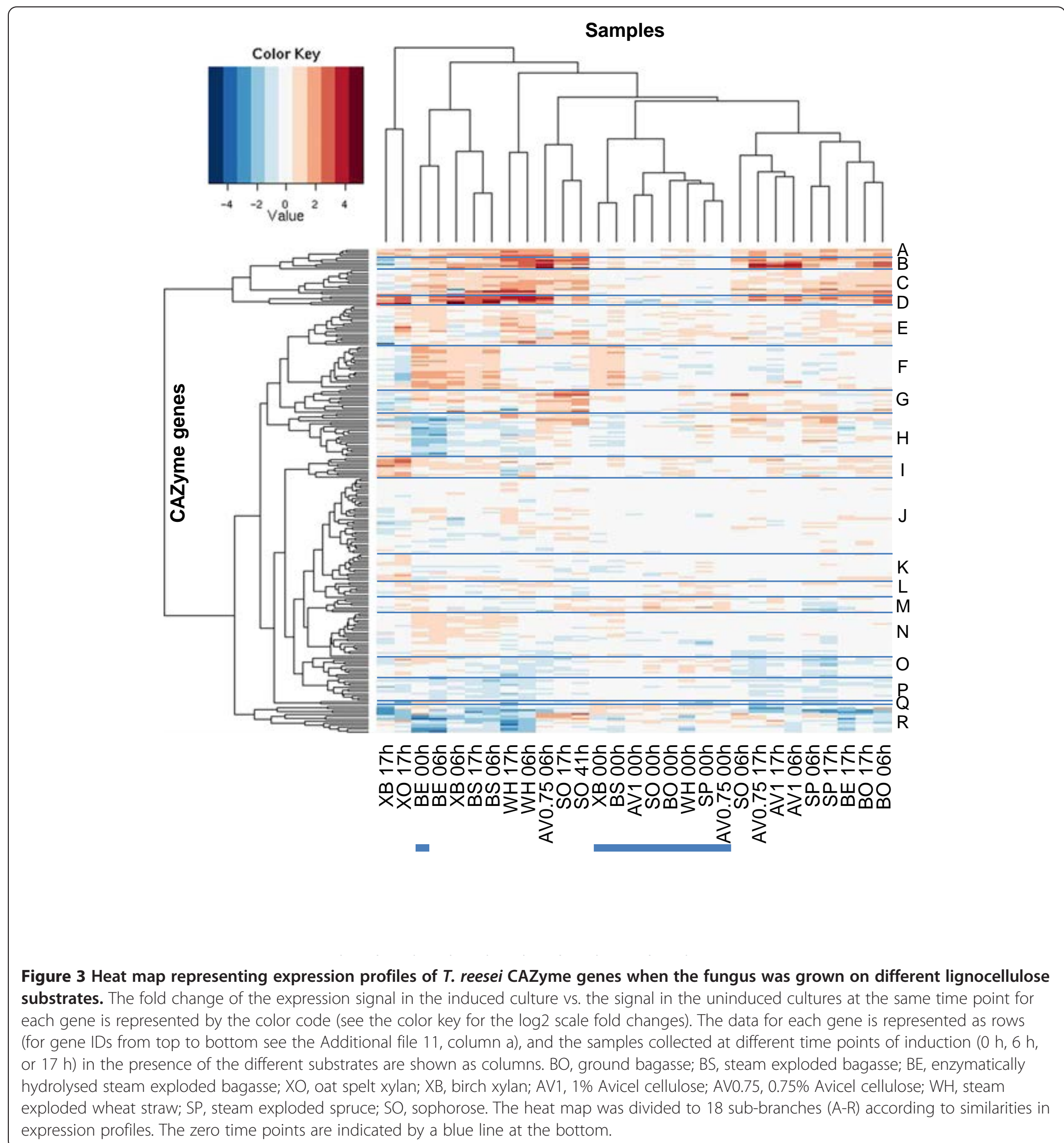


Figure 3 Heat map representing expression profiles of *T. reesei* CAZyme genes when the fungus was grown on different lignocellulose substrates. The fold change of the expression signal in the induced culture vs. the signal in the uninduced cultures at the same time point for each gene is represented by the color code (see the color key for the log2 scale fold changes). The data for each gene is represented as rows (for gene IDs from top to bottom see the Additional file 11, column a), and the samples collected at different time points of induction (0 h, 6 h, or 17 h) in the presence of the different substrates are shown as columns. BO, ground bagasse; BS, steam exploded bagasse; BE, enzymatically hydrolysed steam exploded bagasse; XO, oat spelt xylan; XB, birch xylan; AV1, 1% Avicel cellulose; AV0.75, 0.75% Avicel cellulose; WH, steam exploded wheat straw; SP, steam exploded spruce; SO, sophorose. The heat map was divided to 18 sub-branches (A-R) according to similarities in expression profiles. The zero time points are indicated by a blue line at the bottom.

candidate CE5 acetyl xylan esterase *axe2*, a candidate CE3 family esterase, a candidate GH43 family β-xylosidase/arabinofuranosidase and a novel candidate GH115 family gene. For detailed gene content of the branches, see the Additional file 11.

The major cellobiohydrolase genes, *cbh1* and *cbh2*, are erroneously included to branch J due to saturated signal level in the microarray analysis. The expression of these genes was studied separately using q-PCR (see below).

The microarray study also revealed a group of genes whose expression level was increased immediately after addition of the substrate, but induction (as compared to the control cultures) ceased soon after that. These genes can be found especially in heatmap branches I, L, M, O, and R (Figure 3). These included e.g. genes encoding two candidate β-mannosidases (5836 and 69245), candidate β-xylosidase (58450), a chitinase (*chi18-3*), two α-L-arabinofuranosidases (*abf1* and *abf3*), two candidate α-

1,2-mannosidases (74198 and 60635), a candidate endo-polygalacturonase (103049), endoglucanase (*egl3*), a candidate GH2 family protein (102909), two α-galactosidases (*agl1* and *agl2*) and a glycoside hydrolase for which a family could not be assigned (105288). From these genes *chi18-3*, *egl3*, *abf3* and genes 74198, 5836, 69245, 102909, 105288 and 60635 were induced at the early time point by at least five of the substrates. It is notable that several genes induced at early time point encode enzymes that are involved in hemicellulose degradation (mannose backbone degradation, releasing side chains from hemicellulose and digesting oligosaccharides derived from hemicellulose).

Due to the importance of β-glucosidases in the total hydrolysis of lignocellulose biomass and the observation that members of GH3 are abundantly induced by different substrates and are functionally very diverse according to the phylogenetic analysis, the induction of GH3 β-glucosidase genes was inspected in more detail (Figure 4). The expression patterns of the GH3 β-glucosidase genes differed clearly from one another. A set of the genes was induced most strongly in the presence of sophorose and Avicel cellulose, but not that much on the xylans, whereas some of the genes showed equal or higher induction in the presence of bagasse and xylans as in the presence of the cellulosic materials, and some lacked induction by the substrates. In addition, the induction pattern of the GH3 β-glucosidase genes showed gene specific features. Gene *cel3d* showed strong induction by all the substrates except for the xylans. The strongest induction was obtained in the presence of Avicel, wheat, spruce and sophorose. Genes *bgl1/cel3a* and *cel3e* were strongly induced on sophorose, but also showed a milder induction by many other substrates, especially in the presence of Avicel. Gene *cel3c* was induced by sophorose and Avicel, but also by spruce and ground bagasse. Genes *cel3b* and 108671 were moderately induced by the majority of the substrates except for *cel3b* on wheat and sophorose and 108671 on oat spelt xylan and spruce. Gene 104797 was mildly induced by all the substrates, most by oat spelt xylan and sophorose. Genes 47268 and 66832 were hardly at all induced by the substrates studied.

47 CAZyme genes did not show up-regulation in the presence of any of the substrates as compared to the control cultures (Additional file 11). Most of the uninduced genes are predicted to encode functions other than lignocellulose degradation. Especially, the uninduced genes included genes encoding proteins likely to be involved in processing of cell wall components, chitin processing/degradation, utilisation of storage carbon or 1,4-α-glucan substrates, or protein glycosylation. Expression information could not be obtained for four genes that are absent from the Rut-C30 genome due to a large deletion in scaffold 15 (the genes 25224, 64906, 65215 and 122780) [60,61].

Identification of expression of highly induced genes and confirmation of microarray results by quantitative PCR

The major cellulase genes, *cbh1* and *cbh2,* are among the most abundantly expressed genes in *T. reesei* (within the top 0.5% of the genes in the dataset). In the microarray analysis, the induction of the genes was barely detectable due to saturation of the signal levels. It was also suspected that the magnitude of induction of the major endoglucanase genes, *egl1* and *egl2,* might be too low in the microarray analysis. In order to study the expression of these genes, a quantitative PCR analysis of samples collected at the 17 hour time point of induction was carried out. According to the qPCR analysis, *cbh1* and *cbh2* were induced in the presence of all other substrates except for *cbh2* on birch xylan, and *cbh1* on neither of the xylans (Figure 5). Endoglucanase gene *egl1* was induced in the presence of ground and steam exploded bagasse, Avicel, spruce and sophorose and *egl2* was induced especially in the presence of steam exploded bagasse, but also in the presence of Avicel, wheat, spruce, sophorose and oat spelt xylan. Furthermore, the array data on the

	BG	BS	BE	XO	XB	AV0.75	AV1	WH	SP	SO
cel3d	1,3	1,3	1,1	-0,4	0,1	3,4	1,7	3,1	2,4	2,8
bgl1	0,6	0,5	0,7	-0,1	1,2	1,9	0,6	0,9	0,4	3,2
cel3e	0,4	0,5	0,9	-0,6	0,6	1,2	1,1	0,7	0,7	2,3
cel3c	1,0	0,6	0,1	-2,3	0,2	1,6	1,2	0,3	1,5	1,7
cel3b	1,3	0,9	0,8	0,8	1,0	0,6	1,3	-0,3	1,1	0,3
108671	1,4	0,9	1,4	0,4	0,9	0,9	0,6	1,3	0,2	1,1
104797	0,8	0,7	0,9	1,1	0,5	0,5	0,8	0,8	0,8	1,1
47268	0,4	0,3	0,6	-0,6	0,4	0,2	0,0	0,4	0,1	0,5
66832	0,4	0,2	-0,2	0,1	0,1	0,2	0,3	0,4	0,2	0,6

Figure 4 Expression profiles of *T. reesei* GH3 β-glucosidase genes. The maximal induction level in each of the culture conditions is shown as the fold change of the signal in the induced cultures vs. the signal in the uninduced cultures at the time point when the induction was the highest (log2 scale). The blue and red colours represent negative and positive changes in the expression, respectively. The intensity of the colour is proportional to the magnitude of induction/repression. BO, ground bagasse; BS, steam exploded bagasse; BE, enzymatically hydrolysed steam exploded bagasse; XO, oat spelt xylan; XB, birch xylan; AV1, 1% Avicel cellulose; AV0.75, 0.75% Avicel cellulose; WH, steam exploded wheat straw; SP, steam exploded spruce; SO, sophorose.

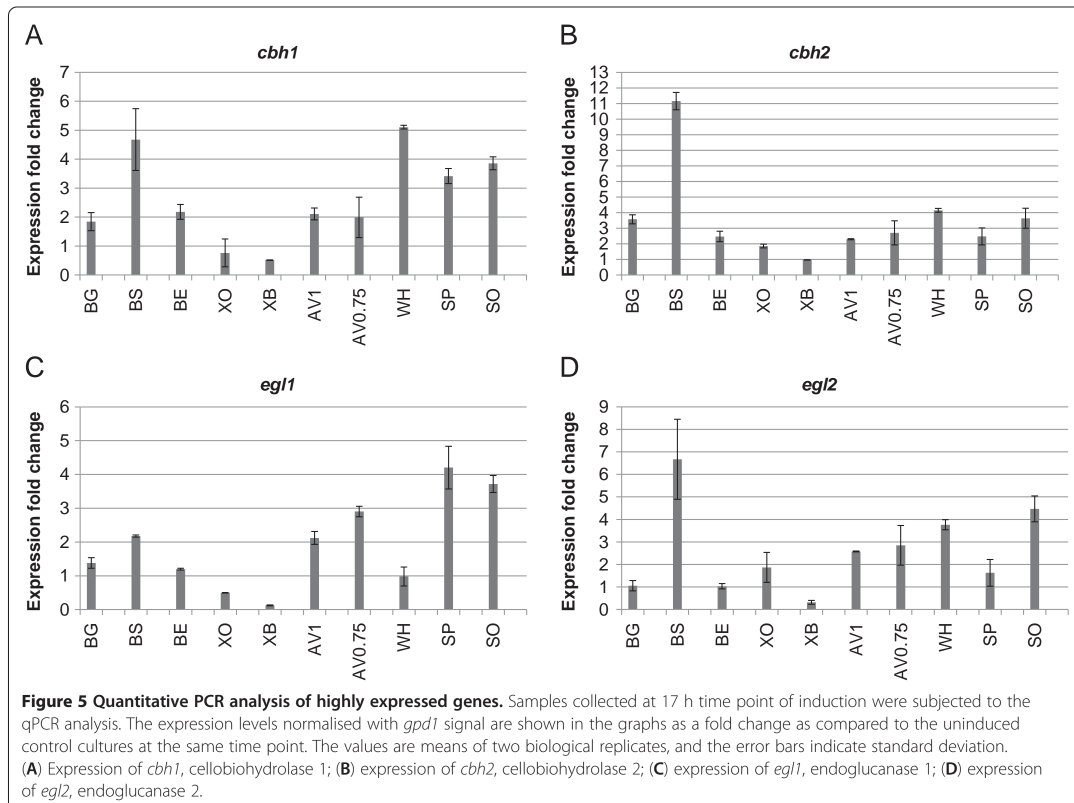


Figure 5 Quantitative PCR analysis of highly expressed genes. Samples collected at 17 h time point of induction were subjected to the qPCR analysis. The expression levels normalised with *gpd1* signal are shown in the graphs as a fold change as compared to the uninduced control cultures at the same time point. The values are means of two biological replicates, and the error bars indicate standard deviation. (**A**) Expression of *cbh1*, cellobiohydrolase 1; (**B**) expression of *cbh2*, cellobiohydrolase 2; (**C**) expression of *egl1*, endoglucanase 1; (**D**) expression of *egl2*, endoglucanase 2.

earlier time points of induction showed induction of *egl2* also on ground and enzymatically pretreated bagasse as well as on birch xylan.

In addition, a set of genes covering both abundantly and moderately expressed genes were included in the qPCR analysis in order to validate the microarray method and to investigate the detection limits of the two methods. The additional set of genes in the qPCR analysis included *abf1, axe1, bxl1, cel3c, cel3d, cel61b, swo1, xyn1, xyn2, xyn3* and *xyn4.* A scatter plot comparing the log2 signal intensities of the microarray data to the Cp values of qPCR data is represented in the Additional file 12. The log2 signal intensities of the microarray data correlate reasonably well with qPCR data at microarray signal level below 15, above which the microarray signal start to become saturated.

Discussion

Although *T. reesei* is an important producer of enzymes for industry and biorefinery applications, little is known about the expression of the enzyme genes in the presence of complex biomass substrates. In this study, the expression of the CAZyme genes of *T. reesei* was studied using several substrates as inducers of gene expression. Substrates included complex biomass materials that are of interest from a biorefinery point of view, as well as purified polysaccharides and a simple inducing disaccharide. In addition, the annotations and functional predictions of *T. reesei* CAZymes were updated using computational and manual methods, including phylogenetic information. This was done in order to assist deeper understanding of *T. reesei* plant biomass degrading enzymes, their regulation and identification of essential enzymatic activities and enzyme genes for complete biomass degradation. After the initial annotation of the *T. reesei* genome [37], attempts to identify all *T. reesei* cellulolytic and hemicellulolytic genes has been done using genome version 1.2 [62] but to our knowledge this is the first publication after the initial annotation in which the *T. reesei* genome v2.0 has been searched for the CAZyme genes and the phylogenetic data has been thoroughly explored to assist the annotation process.

A BLAST based method together with phylogenetic information was used to identify 201 glycoside hydrolase genes, 22 carbohydrate esterase genes (CE10 genes were left out) and 5 polysaccharide lyase genes in total.

Detected discrepancies between the annotations of the genome versions 1.2 [47] and 2.0 [38] and published literature were corrected, and additional refined functional predictions were made for a set of CAZymes based on the analyses. In total 13 putatively new *T. reesei* CAZyme genes were identified during this study (Table 1). Several of these genes belonged to the carbohydrate esterase class which indicates that this group of *T. reesei* enzymes is still less studied than the glycoside hydrolases. Two additional candidate GH61 genes were found emphasizing the possible importance of GH61 enzymes as accessory enzymes in cellulose degradation, although the number of GH61 genes of *T. reesei* is still reduced as compared to other fungi ([37], Additional file 7). For 31 *T. reesei* CAZyme genes the annotation was either refined or a new annotation was given (Table 1). Updated annotations were abundant especially in families GH16, GH17 and GH79. Updated annotations revealed, among others, a fifth candidate GH2 β-mannosidase, a putative ninth GH3 β-glucosidase, a putative second GH12 endoglucanase and the first candidate GH39 β-xylosidase.

In the annotation process, protein homology clusters from 49 fungal species (including *T. reesei*) were mapped to the CAZy database for CAZy family assignment and functional prediction of the genes/gene products, and phylograms of the homology clusters were constructed to assist the annotation of *T. reesei* CAZymes. The phylogenetic relationship of the genes/proteins within the clusters was used to predict further functional diversification of the genes.

Several known and candidate lignocellulose degrading enzymes of *T. reesei* displayed functional diversification within the protein homology clusters, even in the cases in which the enzymes belonged to the same CAZy family and for whom similar activity was predicted based on the closest homologues. Particularly *T. reesei* β-glucosidases of family GH3 and α-galactosidases of GH27 were functionally diverse (Table 1, Figure 1, Additional file 6). It is also worth noting, that GH18 chitinases were extremely diverse by dividing in to as many as five different protein homogy clusters and into 12–13 functional subgroups (Table 1, Additional file 6). A group of chitinases was induced by the lignocellulose substrates used in the study. It is possible that some of the genes encode functions other than merely chitin degradation, since most of the *T. reesei* chitinases are not biochemically characterized. It could also be hypothesized that the saprophytic way of life and pathogenicity towards other fungi would share common regulation mechanisms.

In most cases where the phylogenetic analysis suggested functional diversification, the expression of the genes on different substrates differed as well, as judged based on clustering of the expression profiles (Table 1, Figure 3). A good example of this is the family GH3 β-glucosidases, which were all divided in separate functional subgroups, and whose expression patterns differed from each other. Tight co-regulation was relatively rare among the genes that belonged to the same functional subgroup, also indicating diverted regulation of these genes. However, in a few cases genes of the same functional subgroup were co-regulated. An example of such a case is the GH2 candidate β-mannosidases (5836 and 69245), belonging to the same functional subgroup and showing co-induction immediately after addition of the substrates. The observation that functional diversification is rather common for the CAZymes of *T. reesei* and that the diversification can be seen in differential expression, suggests that the diversified enzymes might be involved in substrate specific processes and/or have different biochemical properties.

The expression analysis of the CAZyme genes in the presence of different substrates revealed distinct groups of co-expressed genes (Figure 3). Part of the CAZyme genes were induced in the presence of both xylan and cellulose type of substrates. The group of genes showing the most consistent induction in the presence of all the substrates used in the experiment contained many genes related to xylanolytic activities. The majority of the CAZyme genes showed differential expression patterns on different types of substrates. Some of the genes exhibited induction especially on the xylan containing material, either on the pure xylans or the complex material containing xylan. The induction of some of the genes was dependent on the type of xylan used in the experiment, suggesting that side chains on xylan may play a role in the induction process. The different types of side chains may contribute also to induction of the genes by different biomass material. The induction of the genes in the xylan cultures also showed temporal differences, some of the genes were induced at the late time points of the induction experiment and some were specific to early stages in xylan cultures or cultures with the xylan containing complex material. A group of genes was induced especially on the cellulosic material, either Avicel or pretreated spruce, and on other complex material to different extent. Part of the cellulose induced genes was induced also in the presence of sophorose. Sophorose can be generated as a transglycosylation product from cellulose degradation product, cellobiose, and therefore could act as a primary inducer in cellulose cultures. Interestingly, a set of genes were induced prímarily by sophorose and only to a lesser extent by the more complex materials. Furthermore, a number of examples were detected where the genes showed stronger induction in the presence of the complex material as compared to the purified polymers, or where the genes

were only induced in the presence of the complex material.

Our data also showed a group of CAZyme genes that were induced at the early time points immediately after addition of the inducing substrates, after which their expression declined. Induction only at early time points followed by a decline at later time points could indicate that the enzyme is either required to initialise hydrolysis of the substrate or to be involved in recognizing the polymer substrate and cutting inducing monomers from the substrate.

The results suggest that several regulatory mechanisms, depending on the inducers present, may act on the CAZyme gene promoters simultaneously, and in some cases also in an additive manner. The complex material may also provide other inducing components than the xylan and cellulose derived inducers. Complete hydrolysis of complex biomass derived material most likely requires action of the enzymes as a cascade. At the initial stages certain components are exposed to act as inducers or as sources for inducer formation and certain linkages are accessible for the enzymatic cleavage. As the degradation proceeds, additional components and linkages are exposed requiring other enzymatic activities for cleavage and, different induction mechanisms to produce the enzymes. The regulation of genes encoding xylanolytic enzymes of the model organism *Neurospora crassa* has been suggested to involve several regulatory groups. Xylanase regulator XLR-1 was suggested to work alone or in combination with other unknown regulators and a XLR-1 independent group of genes was also suggested to exist [63]. The results of our study support the theory of several different regulatory groups which may be partly overlapping.

Comparison of transcriptional profiling data sets reveals the partly different regulatory mechanisms employd by different fungi. The most notable differences between the Avicel regulons of *T. reesei* and *N. crassa* are the larger number of *T. reesei* GH3 genes induced as compared to *N. crassa* and the larger amount of *N. crassa* CE1 and GH61 genes induced as compared to *T. reesei*. There are also differences between the xylan induced genes of the two fungi which are partly due to the fact that in contrast to *T. reesei*, *N. crassa* cellulase genes are not induced by xylan [63]. *N. crassa* cellulase genes are also not induced by sophorose [64].

Conclusions

Computational and manual approaches, also including phylogenetic analysis, was used to update and refine annotation of the CAZyme gene content of *T. reesei* and to study the functional diversification of *T. reesei* CAZyme genes. As an outcome of this study several putatively new CAZyme genes of *T. reesei* were detected, discrepancies between the annotations of the different genome versions and published literature were corrected, and additional refined functional predictions were made for a set of CAZymes. In addition, phylogenetic analysis revealed functional diversification within the CAZy families and enzyme activity groups.

The analysis of *T. reesei* CAZyme gene expression in the presence of different licgnocellulose materials showed a complex pattern of co-regulated groups of genes. Both substrate dependent and temporal differences in the induction of the different groups of genes were detected. The results suggest that several regulatory mechanisms, depending on the inducers present, may act on the CAZyme gene promoters simultaneously, and in some cases the different mechanisms may also act in an additive manner. The complex regulatory system may be required to accomplish complete hydrolysis of biomass derived material by the enzymes produced. Different sets of enzymes are likely to be required to hydrolyse different materials at the different stages of the hydrolysis, thus setting a demand for complex regulatory mechanisms to ensure energetically cost-effective enzyme production in the cells.

Identification of the CAZyme content of *T. reesei* genome together with the expression analysis in the presence of different lignocellulose materials has given evidence for the importance of several yet uncharacterized enzymes in the degradation of biomass substrates and also new information on the enzymes needed for the complete degradation of different lignocellulose substrates. Furthermore, the information on the co-regulated groups of genes can be utilised in further studies to elucidate the regulatory mechanisms of the genes.

Methods

Strains, media and culture conditions

The strain used for the transcriptional profiling was *Trichoderma reesei* Rut-C30 (ATCC 56765, VTT-D-86271, [65]) obtained from VTT Culture Collection. For preparation of spore suspension, the fungus was grown on potato-dextrose plates (Difco) for 5 days. The spores were dislodged, suspended in a buffer containing 0.8% NaCl, 0.025% Tween20 and 20% glycerol, filtered through cotton, and stored at −80°C.

For the induction experiments, *T. reesei* was first cultivated on minimal medium ($(NH_4)_2SO_4$ 7.6 g l^{-1}, KH_2PO_4 15.0 g l^{-1}, 2.4 mM $MgSO_4 \cdot 7H_2O$, 4.1 mM $CaCl_2 \cdot H_2O$, $CoCl_2$ 3.7 mg l^{-1}, $FeSO_4 \cdot 7H_2O$ 5 mg l^{-1}, $ZnSO_4 \cdot 7H_2O$ 1.4 mg l^{-1}, $MnSO_4 \cdot 7H_2O$ 1.6 mg l^{-1}, pH adjusted to 4.8 with KOH) supplemented with 2% (w/v) of sorbitol as a carbon source. The medium was inoculated with 8×10^7 spores per 200 ml aliquots of the medium, and cultivated in shake flasks at 28°C, with shaking at 250 rpm, until biomass dry weight in the cultures was close to

0.9 g/l (4 days). In order to get equal starting material for all the inducing conditions, the preculture aliquots were first mixed together, then divided again in 200 ml aliquots in shake flasks, and let to recover for 30 min at 28°C with shaking at 250 rpm before addition of the inducing substrates. The inducing substrates (for details, see below), suspended in 100 ml of minimal medium, were combined with the 200 ml aliquots of the preculture to start the induction, and the cultivation was continued under the same conditions (28°C, 250 rpm) for 3 days. In uninduced control cultures, 100 ml of minimal medium without inducer was added, and the control cultures were treated similarly to the induced cultures throughout the experiment. Samples for RNA isolation were collected after 0 min, 6 hours, 17 hours, 41 hours and 65 hours of the onset of the induction. The time points were selected to be long enough for induction but to minimize the possible changes in the gene expression related to the growth of the fungus. For the RNA isolation, the mycelium was collected by filtering, washed with equal volume of 0.7% NaCl, frozen immediately in liquid nitrogen, and stored at −80°C. In addition samples for determining the biomass dry weight were withdrawn from the precultures and separate uninduced control cultures during the induction. Biomass dry weight was determined by drying the mycelium samples, collected as above, to constant weight. pH of the cultures was measured throughout the cultivation.

In this study, the induction experiment was carried out in two parts. In the first cultivation set, the inducing substrates used were 0.75% (w/v) Avicel cellulose (Fluka BioChemika), 1% (dry matter w/v) pretreated wheat straw, 1% (dry matter w/v) pretreated spruce, or 0.75 mM α-sophorose (Serva). In the second cultivation set the inducing substrates were 1% (w/v) Avicel cellulose (Fluka BioChemika), 1% (w/v) bagasse ground to homogenous composition, 1% (dry matter w/v) bagasse pretreated using steam explosion, 1% (dry matter w/v) enzymatically hydrolysed pretreated bagasse, 1% (w/v) birch xylan (Roth 7500), 1% (w/v) oat spelt xylan (Sigma-Aldrich, XO627). Uninduced control cultures were included in both cultivation sets. The inducing substrates were added at the same phase of active growth in both cultivation sets. In the first set the biomass dry weight was 0.81 g/l by the addition of the inducers, and 0.97 g/l in the second set, being less than 25% of the maximal biomass in the experiment. Biomass dry weight measurement showed that growth continued logarithmically in the control cultures of both sets during the mock induction. The same specific growth rate, 0.024/h, was measured for both cultivation sets (Figure 6). Thus, it was concluded that the induction took place at a similar growth phase in both cultivation sets. Fungal biomass dry weight could not be measured in the induced cultures due to the insoluble substrates added.

Preparation of the inducing substrates

Steam exploded spruce was kindly provided by Guido Zacchi (Lund University, Sweden). Steam explosion had been done as in [66]. Steam exploded wheat straw was obtained from IFP Energies Nouvelles (France). Bagasse was pretreated by steam explosion using steam pressure 14.5 bar at 200°C (10 l kettle) for 5 minutes without the addition of acid or SO_2 (kindly provided by Anne Kallioinen, VTT). The steam exploded material was washed with distilled water (1 kg/200 ml) after which the insoluble material was filtered, washed with hot tap water (1 kg/2000 ml) and filtered again to obtain an insoluble fibre fraction. The filtered spruce and bagasse materials were washed further twice with 2 l of 82°C distilled water (filtered between washes) and once with 400 ml of 85°C distilled water after which the material was filtered.

Enzymatically pretreated bagasse was obtained by incubating the washed fibre fraction of the steam exploded bagasse with the cellulase mixture Celluclast 1.5 L FG (Novozymes) (50 FPU/g cellulose in the material) and β-glucosidase Novozym 188 DCN00206 (Novozymes) (500 nkat /g cellulose in the material) in 50 mM sodium acetate buffer, pH 4.8, first for 24 h at 45°C with shaking 160 rpm. After the initial 24 h of incubation, insoluble material was collected by centrifugation (20 min, 5300 g, 20°C, Sorvall RC12BP H12000 rotor), resuspended in fresh buffer, and incubation was continued with newly added enzymes (as above) for further 48 h. After the incubation, insoluble material was recovered by centrifugation as above and washed three times with distilled water (pH adjusted to 2.5 with HCl). After the cellulase treatment, the material was resuspensed in final concentration of 5% (w/v) in 50 mM Na_2HPO_4, pH 6.0, and incubated with Protease N (PRW12511N, Amano) (100 nkat / substrate dry weight) for 24 h at 40°C with magnetic stirring. Insoluble fraction was collected by centrifugation, resuspended in 80°C distilled water, and incubated at 80°C for 15 min to inactivate the protease. The insoluble material was washed three times with 1 volume of distilled water.

The carbohydrate composition of the pre-treated substrates (mg/100 mg of dry matter) is shown in Table 2.

Isolation of total RNA, preparation of cDNA, and microarray analytics

Frozen mycelium was ground under liquid nitrogen using mortar and pestle, and total RNA was isolated using Trizol reagent (Invitrogen Life Technologies, Carlsbad, CA, USA) according to manufacturer's instructions. RNA was purified using RNeasy Mini Kit

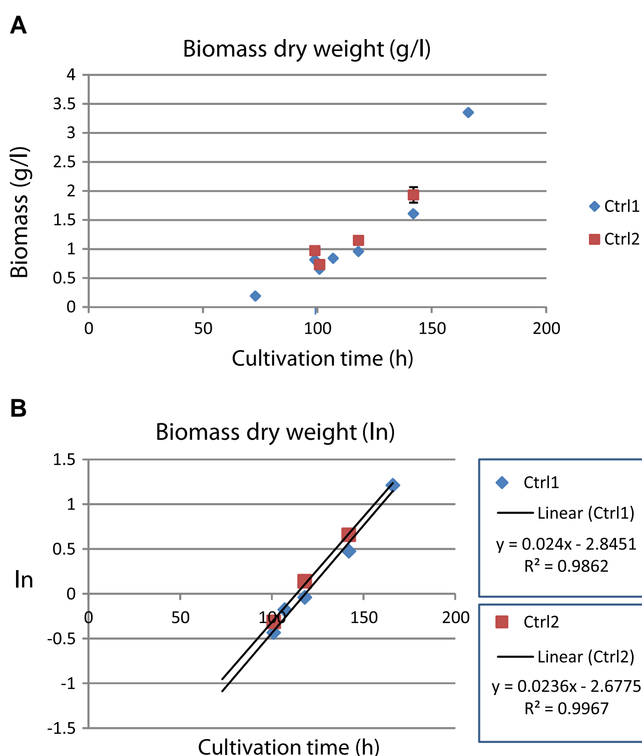


Figure 6 Growth curves of the control cultures in the induction experiments. Biomass dry weight at different time points of cultivation is shown in a linear scale (**A**), and in a logarithmic scale (**B**) for calculation of the specific growth rate. Equations for the linear trend lines of the growth curve after the induction time point, 100 h, are shown in the legend of panel B.

(Qiagen, Hilden, Germany) and RNA concentration was measured using NanoDrop ND-1000 (NanoDrop Technologies Inc. Wilmington, DE, USA). Integrity of RNA was analysed using Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA).

Microarray analysis of total RNA isolated from the first cultivation set (with Avicel, wheat straw, spruce and sophorose as inducing substrates) was carried out by Roche NimbleGen (Roche-NimbleGen, Inc., Madison, WI, USA) as part of their array service, including the synthesis and labelling of cDNA, hybridization of cDNA on microarray slides, and scanning of the slides to produce the raw data files. For the microarray analysis of the second cultivation set, the total RNA samples were processed essentially according to the instructions by RocheNimblegen. The double-stranded cDNA was synthesised using Superscript Double-Stranded cDNA synthesis Kit (Invitrogen), and the integrity of the double-stranded cDNA was analyzed using Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). The double-stranded cDNA was labelled with Cy3 fluorescent dye, hybridized to microarray slides (Roche-NimbleGen, Inc., Madison, WI, USA) and scanned using Roche NimbleGen Microarray scanner according to the instructions of the manufacturer.

The probe design and manufacturing of the microarray slides was carried out by RocheNimbleGen. For the first cultivation set the design was based on the *T. reesei* genome version 1.2 [47] as described in [67]. For the second cultivation set, an array design based on the *T. reesei* genome version 2.0 [38] was used. In the latter array format, six 60mer probes were designed for each of the genes.

The microarray data was analysed using the R package Oligo for preprocessing of the data and the package Limma for identifying differentially expressed genes [68-70]. In the analysis of the differentially expressed genes, the signals in the samples of the induced cultures were compared to the ones of uninduced control cultures at the corresponding time point. Four biological replicates were analysed for each condition and each time point. The cut-off used for statistical significance was p-value <

0.01, and an additional cut-off for the log2-scale fold change (>0.4) was set.

Quantitative PCR

Total RNA isolated from samples collected at the induction time point 17 h were subjected to qPCR analysis of a selected set of genes. cDNA was synthesized using Transcriptor High Fidelity cDNA synthesis kit (Roche), 2 μg of total RNA as a template. A dilution of 1:100 from the cDNA sample was used for assays. qPCR reaction was performed using LightCycler 480 SYBR Green I Master kit (Roche) and Light Cycler 480 II instrument according to the instructions of the manufacturer. The primers used in the qPCR are listed in Table 3. The results were analysed with LightCycler 480 Software release 1.5.0. (version 1.5.0.39) using *gpd1* signal for normalisation. The results are shown as a fold change as compared to the uninduced control cultures.

Mapping *T. reesei* proteome to CAZy database

Information in the CAZy database [5,6] was downloaded family by family (November 2010). The sequence information of reported CAZy family members was then downloaded from NCBI [71] using the database identifier listed on CAZy site. For sequences that did not have a NCBI idenfier, a Uniprot identifier was used instead. Sequences that did not have either identifier were left out. Finally, a searchable BLAST database was created from the protein sequence information using formatdb command from the BLAST program suite [48]. The information on the CAZy family annotation was retained during the construction of the local CAZy BLAST database. The local CAZy BLAST database was queried with each *T. reesei* protein using blastp [48]. Only blast matches having E-value smaller than 10^{-11} were retained.

Each *T. reesei* gene with significant similarity to CAZy database genes was mapped to protein homology clusters described in [49] and updated to include 49 fungi [50], and all the protein members of the found clusters were mapped to CAZy by blastp [48]. A cluster member protein was identified as a CAZyme (carbohydrate active enzyme) if it had a hit in CAZy dabase of at least 97% identity covering over 200 amino acids in the blastp alignment (Additional file 2).

Interpro protein domains [55] from all protein sequences were predicted using InterproScan [72].

Table 3 Primers used in the quantitative PCR method

Gene	5' forward primer	3' reverse primer
cbh1	GCGGATCCTCTTTCTCAG	ATGTTGGCGTAGTAATCATCC
cbh2	TCCTGGTTATTGAGCCTGAC	GCAACATTTGGAAGGTTCAG
egl1	GTCTACTACGAACTCGAC	GTAGTAGTCGTTGCTATACTG
egl2	CTGTACCACAGATGGCAC	ATCATACTTGGAAATGCTCG
xyn1	AAACTACCAAACTGGCGG	TTGATGGGAGCAGAAGATCC
xyn2	CGGCTACTTCTACTCGTACTG	TTGATGACCTTGTTCTTGGTG
xyn4	TTTGACATTGCGACATGGC	GCCGCTATAATCCCAGGT
abf1	ATATCCTTCCGATGCAACAG	AGAGATTGACGAACCGAC
axe1	TAAAGCAGCAATCTTCATGG	GCAGTAAGACTTGATCTTGG
cel3c	ACATCAAGCATTTCATCGCC	ACACTATCCATAAAGGGCCA
cel3d	AGCATATCTCAACTACGCCA	GAAGGTAGCGTAAGACAGG
bxl1	GTCACTCTTCCAAGCTCAG	ATCGTTACCTCTTCTCCCA
cel61b	TGAACTTCTTGCTGCCCA	TAGAGCTGAGTTGCAGGAG
xyn3	TACAAGGGCAAGATTCGTG	ACTGGCTTCCAATACCGT
gpd1	TCCATTCGTGTCCCTACC	AGATACCAGCCTCAATGTC
swo1	ATTACTACACCCAATTCTGGTC	GACAGCCGTATTTGAAGTC

Phylogenetic analysis of *T. reesei* CAZymes

Reconstruction of phylograms of protein homology clusters was carried out as in [73] i.e. the sequences of the proteins in the clusters were aligned with MAFFT [74,75], the alignments trimmed with trimAL [76] and phylogenetic trees constructed with RAxML version 7.2.8 [77], except that due to the large number of trees only 100 bootstraps per tree were made. Trees were visualised using the R [78] library ape [79].

Aligning *T. reesei* CAZymes against the PFAM profiles of a CAZy family

Sequences belonging to selected CAZy families were clustered to help annotate *T. reesei* proteins belonging to the family. The members of the CAZy family, including *T. reesei* candidates, were aligned to the PFAM [80] profile of the family using hmmalign from the HMMer program package [81]. The alignment was then fed into the same pipeline that was used for the construction of phylograms of protein homology clusters. Namely, the alignment was trimmed with trimAL [76] and phylogenetic trees constructed with RAxML version 7.2.8 [77], except that due to the large number of trees only 100 bootstraps per tree were made. Typically the sequences grouped in the phylogenetic tree by order (fungal, bacterial) and by function. Annotation of *T. reesei* proteins was conducted by studying the members of the CAZy family assigned to the same branch as the *T. reesei* protein and by studying whether the *T. reesei* protein contained the conserved amino acids typical for the members of the CAZy family.

Additional files

Additional file 1: Results of mapping the *T. reesei* proteome to the CAZy database. Results shown are the best blast matches for the genes. Results have been sorted according to the e value. Columns C-L are from the default output of blastp search. Column M has the gene descriptions from the CAZy database with the possible EC numbers and column B

has the same information as column M with also the possible CAZy family and information whether the protein is characterized (cha) and if its structure has been determined (str).

Additional file 2: Cut-offs of mapping protein sequences to CAZy database member proteins. (A) Scatterplot of blastp results of all protein sequences from protein clusters of 49 fungi with a *T. reesei* candidate CAZyme. Only values for best hit are shown. Each sequence is represented by the majority vote predicted CAZy family identifier of the protein cluster. Y axis shows the identity percentage from blastp alignment and X axis the length of the alignment as amino acids. Protein was said to be found in CAZy if it had a hit of at least 97% identity which covered over 200 amino acids. (B) Scatterplot of blastp results of protein cluster averages of protein clusters with a CAZy database protein. For each protein only the value of the best hit was considered for counting the cluster averages. Each cluster is represented by the majority vote predicted CAZy family identifier of the protein cluster. Y axis shows the average identity percentage from blastp alignment and X axis the length of the alignment as amino acids. Clusters above the red line and shown in red were accepted for further analysis. (C) Scatterplot of protein cluster averages of protein clusters without a CAZy database protein. See further details from panel B.

Additional file 3: Annotation of *T. reesei* CAZymes. *T. reesei* glycoside hydrolase, carbohydrate esterase (excluding CE10) and polysaccharide lyase genes, the annotation of the genes and the bases used for annotation. (a), gene identifier as in *T. reesei* v2.0 data base [38]; (b), name given to the gene in the publication/data base marked in the reference column; (c), reference to previous studies or to *T. reesei* database versions 1.2 and 2.0. ("A", a previous annotation has been specified/updated during this study); (d), other names used for the gene in marked references; (e), annotation given for the gene in *T. reesei* v2.0 data base; (f), best match for the *T. reesei* CAZyme when *T. reesei* proteome was mapped with blast search to the protein sequences of the CAZy database. The gi identifiers refer to the NCBI protein database; (g), best characterized match for the *T. reesei* CAZyme when *T. reesei* proteome was mapped with blast search to the protein sequences of the CAZy database. The gi identifiers refer to the NCBI protein database.; (h), protein cluster the *T. reesei* CAZyme was assigned to when the protein clusters were mapped to CAZy database by a blast search; (i), functional subgroups within the protein cluster determined according to phylogenetic analysis; (j), characterized protein from another fungus and/ or other *T. reesei* proteins closest to the *T. reesei* CAZyme in a phylogenetic tree constructed from the members of a protein cluster. Uniprot protein identifier is preceded by a code that specifies the species (Additional file 5); (k), CAZy family assigned to the *T. reesei* CAZyme based on the CAZy family members present in the protein cluster; (l), existence of a functional domain/domains for the *T. reesei* CAZyme that supports the CAZy prediction. All the functional domains of the *T. reesei* CAZymes are found from Additional file 4; (m), protein closest to the *T. reesei* CAZyme in a phylogenetic tree constructed from alignment against PFAM of a CAZy family. Protein identifier is preceded by a six letter code that specifies the species and a possible EC number of the enzyme is given. Fusequ = *Fusarium equiseti*, Coccar = *Cochliobolus carbonum*, Hyplix = *Hypocrea lixii*, Bifbif = *Bifidobacterium bifidum*, Acrimp = *Acremonium implicatum*, Penchr = *Penicillium chrysogenum*, Phachr = *Phanerochaete chrysosporium*, Maggri = *Magnaporthe grisea*, Glolin = *Glomerella lindemuthiana*, Aspnid = *Aspergillus nidulans*, Hypvir = *Hypocrea virens*, Flacol = *Flavobacterium columnare*, Aspfum = *Aspergillus fumigatus*, Azocau = *Azorhizobium caulinodans*, Enthis = *Entamoeba histolytica*, Hypjec = *Hypocrea jecorina*, Clothe = *Clostridium thermocellum*.

Additional file 4: Functional Interpro domains of *T. reesei* CAZymes. Gene identifiers are as in *T. reesei* database 2.0 [38]. Domain identifiers and annotations are as in InterPro database [55].

Additional file 5: Fungal species from the protein clusters. Abbreviations specifying the 49 fungal species belonging to the protein homology clusters and the taxonomy of the species.

Additional file 6: Phylogenetic trees for *T. reesei* CAZymes. Trees are constructed from the protein clusters of 49 fungi including *T. reesei* CAZymes. Proteins are named with an uniprot protein identifier which is preceded by a code that specifies the species (Additional file 5).

Additional file 7: Heatmap comparing the protein cluster content of different fungi. Each row is a protein cluster (marked with T and a number) and each column is a fungal species. The colouring of the cells is proportional to the count of proteins. A phylogram of the species is shown above the heatmap together with a colour bar coloured by the taxon of the species. Species abbreviations below the heatmap are explained in Additional file 5. On the left, the colour bar named IDp shows the identity percentage of the genes belonging to the same protein cluster. The darker the colour, the more identical the proteins are.

Additional file 8: Phylogeny of *T. reesei* CAZyme gene 59791. Tree was constructed from the results of blastp against the non-redundant proteinsequences database [82] using BLAST pairwise alignment. The tree method used was fast minimum evolution. Maximum sequence difference was 0.85 and distance model used was Grishin.

Additional file 9: Phylogeny of *T. reesei* CAZyme gene 73101. Tree was constructed from the results of blastp against the non-redundant proteinsequences database [82] using BLAST pairwise alignment. The tree method used was fast minimum evolution. Maximum sequence difference was 0.85 and distance model used was Grishin.

Additional file 10: Phylogeny of *T. reesei* CAZyme gene 108671. Tree was constructed from the results of blastp against the non-redundant proteinsequences database [82] using BLAST pairwise alignment. The tree method used was fast minimum evolution. Maximum sequence difference was 0.85 and distance model used was Grishin.

Additional file 11: Fold changes, signal intensities and significance test for the differential expression of *T. reesei* CAZyme genes. BO, ground bagasse; BS, steam exploded bagasse; BE, enzymatically hydrolysed steam exploded bagasse; XO, oat spelt xylan; XB, birch xylan; AV1, 1% Avicel cellulose; AV0.75, 0.75% Avicel cellulose; WH, steam exploded wheat straw; SP, steam exploded spruce; SO, sophorose; CO1: uninduced control from the first cultivation set; CO2, uninduced control from the second cultivation. Columns marked "Fold change" show the fold change of the signal in the induced culture vs. the signal in the uninduced cultures at the time point (log2 scale). The intensity of the red colour and blue colour indicates the strength of positive and negative fold changes, respectively. Columns marked "Significance" show the results of a significance test (R package limma, p-value < 0.01, log2 fold change > 0.4), 1 indicates induction and −1 repression. Columns marked "Signal intensity" show the signal intensities (log2 scale) from the microarray analysis. Colours indicate different intensities of signals, red represents the strongest signals and green the weakest signals. (a), gene identifier as in *T. reesei* v2.0 data base [38]; (b), class of the protein according to the CAZy classification [5]; (c), family of the protein according to the CAZy classification; (d), protein cluster the *T. reesei* CAZyme was assigned to when the protein clusters were mapped to CAZy database by a blast search; (e), functional subgroups within the protein cluster determined according to phylogenetic analysis; (f), heat map branch (A-R) the gene was assigned to according to the expression profile of the gene; (g), the order in which the genes appear in the heat map representing the expression profiles; (h), induction of the gene by the presence of any of the substrates tested in the study is indicated by "1".

Additional file 12: Comparison of relative transcript signals obtained using microarray or qPCR detection. Pre-processed and normalised microarray signals (log2 scale) were plotted against the relative expression signals obtained using qPCR analysis of the same samples (shown as -ΔCp, normalised using the signals of *gpd1*). Expression data of the genes *xyn4, xyn3, egl2, cel61b, bxl1, egl1, abf1, xyn2, cbh1, swo1, cel3d, cbh2, axe1, xyn1* and *cel3c* were combined and plotted.

Abbreviations

CAZy: Carbohydrate active enzyme database; CAZyme: Carbohydrate active enzyme; CBM: Cellulose binding module; CE: Carbohydrate esterase; GH: Glycoside hydrolase; GT: Glycosyl transferase; PL: Polysaccharide lyase.

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

Authors' contributions
MH carried out fungal cultivations and microarray detection of the expression signals, and participitated in the phylogenetic analysis of CAZyme genes, annotation of the CAZymes as well as in the analysis and interpretation of the microarray data, and drafted the manuscript, MA and MO participated in designing the computational analysis required for gene annotations and carried out mapping of *T. reesei* proteome to CAZy database and phylogenetic analysis of the CAZyme genes, NA carried out qPCR analysis of transcript levels, MP and MS conceived of the study, participated in its design and coordination, and TMP participated in the design and coordination of the study, carried out microarray data analysis, and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements
Aili Grundström is acknowledged for extremely skillful technical assistance. The work was co-funded by the European Commission within the Sixth Framework Programme (2002–2006) (NILE, New Improvements for Lignocellulosic Ethanol, Contract No 019882), Tekes – the Finnish Funding Agency for Technology and Innovation (SugarTech, Hydrolysis technology to produce biomass-based sugars for chemical industry raw materials, Tekes 1503/31/2008), and Academy of Finland (The regulatory network of the cellulolytic and hemicellulolytic system of Trichoderma reesei, Decision no 133455). Work of MA was funded by Academy of Finland Postdoctoral Researcher's fellowship 127715.

Received: 29 June 2012 Accepted: 22 September 2012
Published: 4 October 2012

References
1. Sánchez C: **Lignocellulosic residues: Biodegradation and bioconversion by fungi.** *Biotechnol Adv* 2009, **27**(2):185–194.
2. McKendry P: **Energy production from biomass (part 1): overview of biomass.** *Bioresour Technol* 2002, **83**(1):37–46.
3. Jordan DB, Bowman MJ, Braker JD, Dien BS, Hector RE, Lee CC, Mertens JA, Wagschal K: **Plant cell walls to ethanol.** *Biochem J* 2012, **442**(2):241–252.
4. Sukumaran RK, Singhania RR, Mathew GM, Pandey A: **Cellulase production using biomass feed stock and its application in lignocellulose saccharification for bio-ethanol production.** *Renew Energy* 2009, **34**(2):421–424.
5. *Carbohydrate active enzymes database.* http://www.cazy.org/.
6. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The Carbohydrate-active enxymes database (CAZy): an expert resource for glycogenomics.** *Nucleic Acids Res* 2009, **37**(suppl 1):D233–D238.
7. Schuster A, Schmoll M: **Biology and biotechnology of *Trichoderma*.** *Appl Microbiol Biotechnol* 2010, **87**(3):787–799.
8. Teeri T, Salovuori I, Knowles J: **The molecular cloning of the major cellulase gene from *Trichoderma reesei*.** *Nat Biotech* 1983, **1**(8):696–699.
9. Shoemaker S, Schweickart V, Ladner M, Gelfand D, Kwok S, Myambo K, Innis M: **Molecular cloning of exo-cellobiohydrolase I derived from *Trichoderma reesei* strain L27.** *Nat Biotech* 1983, **1**(8):691–696.
10. Teeri TT, Lehtovaara P, Kauppinen S, Salovuori I, Knowles J: **Homologous domains in *Trichoderma reesei* cellulolytic enzymes: Gene sequence and expression of cellobiohydrolase II.** *Gene* 1987, **51**(1):43–52.
11. Penttilä M, Lehtovaara P, Nevalainen H, Bhikhabhai R, Knowles J: **Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence of the endoglucanase I gene.** *Gene* 1986, **45**(3):253–263.
12. Okada H, Tada K, Sekiya T, Yokoyama K, Takahashi A, Tohda H, Kumagai H, Morikawa Y: **Molecular characterization and heterologous expression of the gene encoding a low-molecular-mass endoglucanase from *Trichoderma reesei* QM9414.** *Appl Environ Microbiol* 1998, **64**(2):555–563.
13. Saloheimo M, Nakari-Setälä T, Tenkanen M, Penttilä M: **cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast.** *Eur J Biochem* 1997, **249**(2):584–591.
14. Saloheimo A, Henrissat B, Hoffrén A, Teleman O, Penttilä M: **A novel, small endoglucanase gene, *egl5*, from *Trichoderma reesei* isolated by expression in yeast.** *Mol Microbiol* 1994, **13**(2):219–228.
15. Saloheimo M, Lehtovaara P, Penttilä M, Teeri TT, Ståhlberg J, Johansson G, Pettersson G, Claeyssens M, Tomme P, Knowles JKC: **EGIII, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme.** *Gene* 1988, **63**(1):11–21.
16. Foreman PK, Brown D, Dankmeyer L, Dean R, Diener S, Dunn-Coleman NS, Goedegebuur F, Houfek TD, England GJ, Kelley AS, Meerman HJ, Mitchell T, Mitchinson C, Olivares HA, Teunissen PJM, Yao J, Ward M: **Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*.** *J Biol Chem* 2003, **278**(34):31988–31997.
17. Grishutin SG, Gusakov AV, Markov AV, Ustinov BB, Semenova MV, Sinitsyn AP: **Specific xyloglucanases as a new class of polysaccharide-degrading enzymes.** *Biochimica et Biophysica Acta (BBA) - General Subjects* 2004, **1674**(3):268–281.
18. Langston JA, Shaghasi T, Abbate E, Xu F, Vlasenko E, Sweeney MD: **Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61.** *Appl Environ Microbiol* 2011, **77**(19):7007–7015.
19. Fowler T, Brown RD: **The *bgl1* gene encoding extracellular β-glucosidase from *Trichoderma reesei* is required for rapid induction of the cellulase complex.** *Mol Microbiol* 1992, **6**(21):3225–3235.
20. Saloheimo M, Kuja-Panula J, Ylosmaki E, Ward M, Penttila M: **Enzymatic properties and intracellular localization of the novel *Trichoderma reesei* β-glucosidase BGLII (Cel1A).** *Appl Environ Microbiol* 2002, **68**(9):4546–4553.
21. Takashima S, Nakamura A, Hidaka M, Masaki H, Uozumi T: **Molecular cloning and expression of the novel fungal β-glucosidase genes from *Humicola grisea* and *Trichoderma reesei*.** *J Biochem* 1999, **125**(4):728–736.
22. Barnett CC, Berka RM, Fowler T: **Cloning and amplification of the gene encoding an extracellular β-glucosidase from *Trichoderma reesei*: evidence for improved rates of saccharification of cellulosic substrates.** *Nat Biotech* 1991, **9**(6):562–567.
23. Saloheimo M, Paloheimo M, Hakola S, Pere J, Swanson B, Nyyssönen E, Bhatia A, Ward M, Penttilä M: **Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials.** *Eur J Biochem* 2002, **269**(17):4202–4211.
24. Tenkanen M, Puls J, Poutanen K: **Two major xylanases of *Trichoderma reesei*.** *Enzyme Microb Technol* 1992, **14**(7):566–574.
25. Torronen A, Mach RL, Messner R, Gonzalez R, Kalkkinen N, Harkki A, Kubicek CP: **The two major xylanases from *Trichoderma reesei*: characterization of both enzymes and genes.** *Nat Biotech* 1992, **10**(11):1461–1465.
26. Xu J, Takakuwa N, Nogawa M, Okada H, Okada H: **A third xylanase from *Trichoderma reesei* PC-3-7.** *Appl Microbiol Biotechnol* 1998, **49**(6):718–724.
27. Stalbrand H, Saloheimo A, Vehmaanpera J, Henrissat B, Penttila M: **Cloning and expression in *Saccharomyces cerevisiae* of a *Trichoderma reesei* beta-mannanase gene containing a cellulose binding domain.** *Appl Environ Microbiol* 1995, **61**(3):1090–1097.
28. Margolles-Clark E, Tenkanen M, Soderlund H, Penttila M: **Acetyl xylan esterase from *Trichoderma reesei* contains an active-site serine residue and a cellulose-binding domain.** *Eur J Biochem* 1996, **237**(3):553–560.
29. Margolles-Clark E, Saloheimo M, Siika-aho M, Penttilä M: **The α-glucuronidase-encoding gene of *Trichoderma reesei*.** *Gene* 1996, **172**(1):171–172.
30. Margolles-Clark E, Tenkanen M, Nakari-Setala T, Penttila M: **Cloning of genes encoding alpha-L-arabinofuranosidase and beta-xylosidase from *Trichoderma reesei* by expression in *Saccharomyces cerevisiae*.** *Appl Environ Microbiol* 1996, **62**(10):3840–3846.
31. Herpoël-Gimbert I, Margeot A, Dolla A, Jan G, Mollé D, Lignon S, Mathis H, Sigoillot J, Monot F, Asther M: **Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains.** *Biotechnol Biofuels* 2008, **1**:18.
32. Margolles-Clark E, Tenkanen M, Luonteri E, Penttilä M: **Three α-galactosidase genes of *Trichoderma reesei* cloned by expression in yeast.** *Eur J Biochem* 1996, **240**(1):104–111.
33. Zeilinger S, Kristufek D, Arisan-Atac I, Hodits R, Kubicek CP: **Conditions of formation, purification, and characterization of an alpha-galactosidase of *Trichoderma reesei* RUT C-30.** *Appl Environ Microbiol* 1993, **59**(5):1347–1353.
34. Li X, Skory CD, Cotta MA, Puchart V, Biely P: **Novel family of carbohydrate esterases, based on identification of the *Hypocrea jecorina* acetyl esterase gene.** *Appl Environ Microbiol* 2008, **74**(24):7482–7489.
35. Pokkuluri PR, Duke NEC, Wood SJ, Cotta MA, Li X, Biely P, Schiffer M: **Structure of the catalytic domain of glucuronoyl esterase Cip2 from**

Hypocrea jecorina. *Proteins: Structure, Function, and Bioinformatics* 2011, **79**(8):2588–2592.
36. Li X, Špániková S, de Vries RP, Biely P: **Identification of genes encoding microbial glucuronoyl esterases.** *FEBS Lett* 2007, **581**(21):4029–4035.
37. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EGJ, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS: **Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*).** *Nat Biotech* 2008, **26**(5):553–560.
38. *Trichoderma reesei genome database v2.0*. http://genome.jgi-psf.org/Trire2/Trire2.home.html.
39. Aro N, Pakula T, Penttilä M: **Transcriptional regulation of plant cell wall degradation by filamentous fungi.** *FEMS Microbiol Rev* 2005, **29**(4):719–739.
40. Schmoll M, Kubicek CP: **Regulation of *Trichoderma* cellulase formation: lessons in molecular biology from an industrial fungus.** *Acta Microbiol Immunol Hung* 2003, **50**(2):125–145.
41. Kubicek C, Mikus M, Schuster A, Schmoll M, Seiboth B: **Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*.** *Biotechnol Biofuels* 2009, **2**(1):19.
42. Margolles-clark E, Ilmen M, Penttilä M: **Expression patterns of ten hemicellulase genes of the filamentous fungus *Trichoderma reesei* on various carbon sources.** *J Biotechnol* 1997, **57**(1–3):167–179.
43. Verbeke J, Coutinho P, Mathis H, Quenot A, Record E, Asther M, Heiss-Blanquet S: **Transcriptional profiling of cellulase and expansin-related genes in a hypercellulolytic *Trichoderma reesei*.** *Biotechnol Lett* 2009, **31**(9):1399–1405.
44. Ilmen M, Saloheimo A, Onnela M, Penttila M: **Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*.** *Appl Environ Microbiol* 1997, **63**(4):1298–1306.
45. Nogawa M, Goto M, Okada H, Morikawa Y: **l -Sorbose induces cellulase gene transcription in the cellulolytic fungus *Trichoderma reesei*.** *Curr Genet* 2001, **38**(6):329–334.
46. Mach-Aigner AR, Gudynaite-Savitch L, Mach RL: **L-Arabitol is the actual inducer of xylanase expression in *Hypocrea jecorina* (*Trichoderma reesei*).** *Appl Environ Microbiol* 2011, **77**(17):5988–5994.
47. *Trichoderma reesei genome database v1.2*. http://genome.jgi-psf.org/trire1/trire1.home.html.
48. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**(17):3389–3402.
49. Arvas M, Kivioja T, Mitchell A, Saloheimo M, Ussery D, Penttila M, Oliver S: **Comparison of protein coding gene contents of the fungal phyla Pezizomycotina and Saccharomycotina.** *BMC Genomics* 2007, **8**(1):325.
50. Gasparetti C, Faccio G, Arvas M, Buchert J, Saloheimo M, Kruus K: **Discovery of a new tyrosinase-like enzyme family lacking a C-terminally processed domain: production and characterization of an *Aspergillus oryzae* catechol oxidase.** *Appl Microbiol Biotechnol* 2010, **86**(1):213–226.
51. Wolfe KH, Shields DC: **Molecular evidence for an ancient duplication of the entire yeast genome.** *Nature* 1997, **387**(6634):708–713.
52. Kellis M, Birren BW, Lander ES: **Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*.** *Nature* 2004, **428**(6983):617–624.
53. Costenoble R, Picotti P, Reiter L, Stallmach R, Heinemann M, Sauer U, Aebersold R: **Comprehensive quantitative analysis of central carbon and amino-acid metabolism in *Saccharomyces cerevisiae* under multiple conditions by targeted proteomics.** *Mol Syst Biol* 2011, **7**:464.
54. Taylor JW, Berbee ML: **Dating divergences in the fungal tree of life: review and new analyses.** *Mycologia November/December* 2006, **98**(6):838–849.
55. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duquenne L, Finn RD, Fraser M, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, McMenamin C, Mi H, Mutowo-Muellenet P, Mulder N, Natale D, Orengo C, Pesseat S, Punta M, Quinn AF, Rivoire C, Sangrador-Vegas A, Selengut JD, Sigrist CJA, Scheremetjew M, Tate J, Thimmajanarthanan M, Thomas PD, Wu CH, Yeats C, Yong S: **InterPro in 2011: new developments in the family and domain prediction database.** *Nucleic Acids Res* 2012, **40**(D1):D306–D312.
56. *NCBI conserved domains database*. http://www.ncbi.nlm.nih.gov/cdd.
57. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Meth* 2011, **8**(10):785–786.
58. Metz B, Seidl-Seiboth V, Haarmann T, Kopchinskiy A, Lorenz P, Seiboth B, Kubicek CP: **Expression of biomass-degrading enzymes is a major event during conidium development in *Trichoderma reesei*.** *Eukaryot Cell* 2011, **10**(11):1527–1535.
59. Seidl V, Huemer B, Seiboth B, Kubicek CP: **A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases.** *FEBS J* 2005, **272**(22):5923–5939.
60. Le Crom S, Schackwitz W, Pennacchio L, Magnuson JK, Culley DE, Collett JR, Martin J, Druzhinina IS, Mathis H, Monot F, Seiboth B, Cherry B, Rey M, Berka R, Kubicek CP, Baker SE, Margeot A: **Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei* using massively parallel DNA sequencing.** *Proc Natl Acad Sci* 2009, **106**(38):16151–16156.
61. Seidl V, Gamauf C, Druzhinina I, Seiboth B, Hartl L, Kubicek C: **The *Hypocrea jecorina* (*Trichoderma reesei*) hypercellulolytic mutant RUT C30 lacks a 85 kb (29 gene-encoding) region of the wild-type genome.** *BMC Genomics* 2008, **9**(1):327.
62. Ouyang J, Yan M, Kong D, Xu L: **A complete protein pattern of cellulase and hemicellulase genes in the filamentous fungus *Trichoderma reesei*.** *Biotechnol J* 2006, **1**(11):1266–1274.
63. Sun J, Tian C, Diamond S, Glass NL: **Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in *Neurospora crassa*.** *Eukaryot Cell* 2012, **11**(4):482–493.
64. Znameroski EA, Coradetti ST, Roche CM, Tsai JC, Iavarone AT, Cate JHD, Glass NL: **Induction of lignocellulose-degrading enzymes *in Neurospora crassa* by cellodextrins.** *Proc Natl Acad Sci* 2012, **109**(16):6012–6017.
65. Montenecourt BS, Eveleigh DE: **Preparation of mutants of *Trichoderma reesei* with enhanced cellulase production.** *Appl Environ Microbiol* 1977, **34**(6):777–782.
66. Stenberg K, Tengborg C, Galbe M, Zacchi G: **Optimisation of steam pretreatment of SO2-impregnated mixed softwoods for ethanol production.** *J Chem Technol Biotechnol* 1998, **71**(4):299–308.
67. Arvas M, Pakula T, Smit B, Rautio J, Koivistoinen H, Jouhten P, Lindfors E, Wiebe M, Penttila M, Saloheimo M: **Correlation of gene expression and protein production rate - a system wide study.** *BMC Genomics* 2011, **12**:616.
68. Bolstad BM, Irizarry RA, Åstrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** *Bioinformatics* 2003, **19**(2):185–193.
69. *Bioconductor, open source software for bioinformatics*. http://www.bioconductor.org/.
70. Smyth GK, Michaud J, Scott HS: **Use of within-array replicate spots for assessing differential expression in microarray experiments.** *Bioinformatics* 2005, **21**(9):2067–2075.
71. *National center for biotechnology information*. http://www.ncbi.nlm.nih.gov/.
72. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R: **InterProScan: protein domains identifier.** *Nucleic Acids Res* 2005, **33**(suppl 2):W116–W120.
73. Koivistoinen OM, Arvas M, Headman JR, Andberg M, Penttilä M, Jeffries TW, Richard P: **Characterisation of the gene cluster for l-rhamnose catabolism in the yeast *Scheffersomyces* (*Pichia*) *stipitis*.** *Gene* 2012, **492**(1):177–185.
74. Katoh K, Misawa K, Kuma K, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids Res* 2002, **30**(14):3059–3066.
75. Katoh K, Kuma K, Toh H, Miyata T: **MAFFT version 5: improvement in accuracy of multiple sequence alignment.** *Nucleic Acids Res* 2005, **33**(2):511–518.
76. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T: **trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses.** *Bioinformatics* 2009, **25**(15):1972–1973.
77. Stamatakis A: **RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006, **22**(21):2688–2690.
78. Team RDC: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2008.

79. Paradis E, Claude J, Strimmer K: **APE: Analyses of phylogenetics and evolution in R language.** *Bioinformatics* 2004, **20**(2):289–290.
80. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD: **The Pfam protein families database.** *Nucleic Acids Res* 2012, **40**(D1):D290–D301.
81. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching.** *Nucleic Acids Res* 2011, **39**(suppl 2):W29–W37.
82. *BLAST: Basic local alignment search tool.* http://blast.ncbi.nlm.nih.gov/Blast.cgi.
83. Ike M, Isami K, Tanabe Y, Nogawa M, Ogasawara W, Okada H, Morikawa Y: **Cloning and heterologous expression of the exo-β-d-glucosaminidase-encoding gene (*gls93*) from a filamentous fungus, *Trichoderma reesei* PC-3-7.** *Appl Microbiol Biotechnol* 2006, **72**(4):687–695.
84. Dunn-Coleman N, Neefe-Kruithof P, Pilgrim CE, Van Solingen P, Ward DE: ***Trichoderma reesei* glucoamylase and homologs thereof.** *United States Patent* 2009, **11/510:**892 (7494685).
85. Ike M, Nagamatsu K, Shioya A, Nogawa M, Ogasawara W, Okada H, Morikawa Y: **Purification, characterization, and gene cloning of 46 kDa chitinase (Chi46) from *Trichoderma reesei* PC-3-7 and its expression in *Escherichia coli*.** *Appl Microbiol Biotechnol* 2006, **71**(3):294–303.
86. Stals I, Samyn B, Sergeant K, White T, Hoorelbeke K, Coorevits A, Devreese B, Claeyssens M, Piens K: **Identification of a gene coding for a deglycosylating enzyme in *Hypocrea jecorina*.** *FEMS Microbiol Lett* 2010, **303**(1):9–17.
87. Saloheimo M, Siika-Aho M, Tenkanen M, Penttila M: **Xylanase from *Trichoderma reesei*, method for production thereof, and methods employing this enzyme.** *Patent* 2003, **09/658:**772. 6555335.
88. Geysens S, Pakula T, Uusitalo J, Dewerte I, Penttilä M, Contreras R: **Cloning and characterization of the glucosidase II alpha subunit gene of *Trichoderma reesei*: a frameshift mutation results in the aberrant glycosylation profile of the hypercellulolytic strain Rut-C30.** *Appl Environ Microbiol* 2005, **71**(6):2910–2924.
89. Seiboth B, Hartl L, Salovuori N, Lanthaler K, Robson GD, Vehmaanpera J, Penttila ME, Kubicek CP: **Role of the *bga1*-encoded extracellular β-galactosidase of *Hypocrea jecorina* in cellulase induction by lactose.** *Appl Environ Microbiol* 2005, **71**(2):851–857.
90. Maras M, Callewaert N, Piens K, Claeyssens M, Martinet W, Dewaele S, Contreras H, Dewerte I, Penttilä M, Contreras R: **Molecular cloning and enzymatic characterization of a *Trichoderma reesei* 1,2-α-d-mannosidase.** *J Biotechnol* 2000, **77**(2–3):255–263.
91. Konno N, Igarashi K, Habu N, Samejima M, Isogai A: **Cloning of the *Trichoderma reesei* cDNA encoding a glucuronan lyase belonging to a novel polysaccharide lyase family.** *Appl Environ Microbiol* 2009, **75**(1):101–107.

doi:10.1186/1475-2859-11-134
Cite this article as: Häkkinen *et al.*: Re-annotation of the CAZy genes of *Trichoderma reesei* and transcription in the presence of lignocellulosic substrates. *Microbial Cell Factories* 2012 **11**:134.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

PUBLICATION II

Screening of candidate regulators for cellulase and hemicellulase production in *Trichoderma reesei* and identification of a factor essential for cellulase production

In: Biotechnology for Biofuels 7:14.
Copyright 2014 Authors.

Häkkinen *et al. Biotechnology for Biofuels* 2014, **7**:14
http://www.biotechnologyforbiofuels.com/content/7/1/14

Biotechnology for Biofuels

RESEARCH **Open Access**

Screening of candidate regulators for cellulase and hemicellulase production in *Trichoderma reesei* and identification of a factor essential for cellulase production

Mari Häkkinen*, Mari J Valkonen, Ann Westerholm-Parvinen, Nina Aro, Mikko Arvas, Marika Vitikainen, Merja Penttilä, Markku Saloheimo and Tiina M Pakula

Abstract

Background: The soft rot *ascomycetal* fungus *Trichoderma reesei* is utilized for industrial production of secreted enzymes, especially lignocellulose degrading enzymes. *T. reesei* uses several different enzymes for the degradation of plant cell wall-derived material, including 9 characterized cellulases, 15 characterized hemicellulases and at least 42 genes predicted to encode cellulolytic or hemicellulolytic activities. Production of cellulases and hemicellulases is modulated by environmental and physiological conditions. Several regulators affecting the expression of cellulase and hemicellulase genes have been identified but more factors still unknown are believed to be present in the genome of *T. reesei*.

Results: We have used transcriptional profiling data from *T. reesei* cultures in which cellulase/hemicellulase production was induced by the addition of different lignocellulose-derived materials to identify putative novel regulators for cellulase and hemicellulase genes. Based on this induction data, supplemented with other published genome-wide data on different protein production conditions, 28 candidate regulatory genes were selected for further studies and they were overexpressed in *T. reesei*. Overexpression of seven genes led to at least 1.5-fold increased production of cellulase and/or xylanase activity in the modified strains as compared to the parental strain. Deletion of gene 77513, here designated as *ace3*, was found to be detrimental for cellulase production and for the expression of several cellulase genes studied. This deletion also significantly reduced xylanase activity and expression of xylan-degrading enzyme genes. Furthermore, our data revealed the presence of co-regulated chromosomal regions containing carbohydrate-active enzyme genes and candidate regulatory genes.

Conclusions: Transcriptional profiling results from glycoside hydrolase induction experiments combined with a previous study of specific protein production conditions was shown to be an effective method for finding novel candidate regulatory genes affecting the production of cellulases and hemicellulases. Recombinant strains with improved cellulase and/or xylanase production properties were constructed, and a gene essential for cellulase gene expression was found. In addition, more evidence was gained on the chromatin level regional regulation of carbohydrate-active enzyme gene expression.

Keywords: Carbohydrate active enzymes, Cellulase, Co-regulation, Gene regulation, Hemicellulase, Transcription factors, Transcriptional profiling

* Correspondence: mari.hakkinen@vtt.fi
VTT Technical Research Centre of Finland, PO Box 1000 Tietotie 2, Espoo FI-02044, VTT, Finland

© 2014 Häkkinen et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

Background

Plant biomass, consisting mostly of cellulose, hemicellulose and lignin, is the most abundant renewable energy source on earth. Degradation of the biomass and continuation of the carbon cycle is maintained mainly by microbial action, especially by fungi of different species. The biomass-degrading enzymes produced by these organisms also have applications in different fields of industry, including biorefinery applications [1]. *Trichoderma reesei* (an anamorph of *Hypocrea jecorina*) is an extremely efficient producer of cellulose- and hemicellulose-degrading enzymes, and is therefore widely employed by the enzyme industry for the production of its own enzymes as well as for producing proteins from other sources [2,3]. The genome of *T. reesei* encodes nine characterized cellulase enzymes and 15 characterized hemicellulase enzymes. In addition, a large number of genes encoding candidate carbohydrate-active enzymes (CAZy) [4,5] have been identified from the genome [6,7]. According to an updated annotation, the genome encodes 201 glycoside hydrolase genes, 22 carbohydrate esterase genes and 5 polysaccharide lyase genes, of which at least 66 are known or predicted to encode cellulolytic and hemicellulolytic activities [8].

Energy efficient production of cellulases and hemicellulases is achieved by tight gene regulation governed by inducer-dependent expression of the genes and by repression of the genes in the presence of fast metabolized carbon sources (for reviews see [9,10]). In addition to the type of carbon source, additional environmental conditions are known to affect protein production together with the physiological state of the cells, such as pH [11], light [12], the specific growth rate and cell density of the fungus [13,14], and the physiological state of the mitochondria [15]. Furthermore, the expression of many cellulase and hemicellulase genes is shown to be under a feedback regulation mechanism that functions under conditions in which the capacity of the cells to fold and secrete proteins is limited and transcriptional down-regulation is required to reduce the amount of secreted protein produced [16].

The variety of environmental and physiological factors affecting the enzyme production of *T. reesei* infers that a complex signaling cascade and regulatory network is needed for the accurate timing of hydrolytic enzyme production. Several regulatory factors for cellulase and hemicellulase genes have been characterized, the most extensively studied of which are the transcription factor CRE1, which mediates carbon catabolite repression [17], and the major regulator needed for expression, XYR1 [18]. Other characterized factors are the positively acting ACE2 [19] and HAP2/3/5 complex [20], and the negatively acting factor ACE1 [21,22]. Recently, novel factors possibly affecting the regulation of genes encoding hydrolytic enzymes have been found from *Trichoderma* and other fungi. F-box proteins that have been suggested to be involved in the regulation of plant cell wall-degrading enzymes have been identified from *Aspergillus* and *Fusarium* [23,24]. Two putative regulators of cellulase and hemicellulase genes named CLR-1 and CLR-2 have been identified from *Neurospora crassa* [25] and a transcription factor BglR has been suggested to regulate β-glucosidase genes of *T. reesei* [26]. Another recent finding is that the putative methyltransferase LAE1 is essential for the formation of *T. reesei* cellulases and hemicellulases, although the precise mechanism is still unclear [27]. In the light of recent findings from *Trichoderma* and other fungi, it can be assumed that not all regulatory factors have been identified yet and that additional regulatory genes can still be found in the genome of *T. reesei*.

In this study, transcriptional profiling data from *T. reesei* cultivated in the presence of several lignocellulose substrates as well as other genome-wide data from different types of protein production conditions were used to identify putative regulators for cellulase and hemicellulase genes. Several candidate regulatory genes were identified, and shown to have an effect on cellulase and hemicellulase production when overexpressed in *T. reesei*. Furthermore, the genomic context of the CAZy genes and co-regulated candidate regulatory genes were analyzed. The data revealed co-regulated regions containing candidate regulatory genes and CAZy genes, as well as other genes relevant for the utilization of the carbon source, such as transporter genes. The relevance of the regions is discussed in the paper.

Results

Analysis of transcriptome data to identify candidates for regulators of cellulase and hemicellulase genes

Transcriptome analysis has previously been carried out to study the expression of CAZy genes in *T. reesei* cultures that were induced by the addition of different types of lignocellulose material, purified carbohydrate polymers or disaccharides (Avicel cellulose, pretreated wheat straw, pretreated spruce or sophorose) [8]. In the present study, data from the previous work were further analyzed and explored to identify candidate regulators for CAZy genes and, in particular, for cellulase and hemicellulase genes. The expression data were clustered using Mfuzz [28,29] to reveal groups of co-regulated genes. The majority of the genes encoding characterized enzymes and accessory factors involved in lignocellulose degradation were found in two clusters. Cluster 10 contained the major cellulase and β-glucosidase genes (*cbh1, cbh2, egl1, egl2, egl3, egl5, bgl1* and *bgl2*) together with a set of hemicellulase genes (*abf1, bga1, cip2, cel74a* and *xyn3*). Cluster 35 contained predominantly hemicellulase genes (*agl1, agl3, man1, aes1, axe1, bxl1, glr1, xyn1 and xyn4*) (Figure 1; for gene names, see [8]).

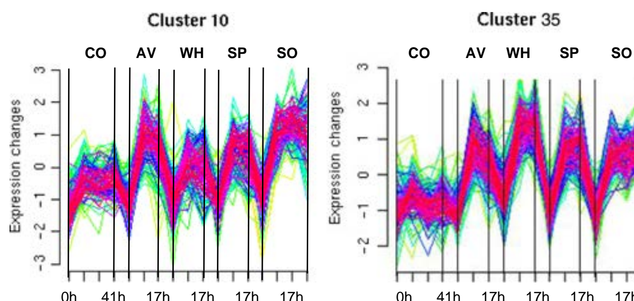


Figure 1 Expression profiles of the clusters from Mfuzz clustering containing the majority of the cellulase and hemicellulase genes. The expression array dataset on *T. reesei* cultures induced with Avicel cellulose, pretreated wheat, pretreated spruce or sophorose (described in [8]) were clustered using Mfuzz. AV, Avicel cellulose; CO, control cultivation; SO, sophorose; SP, spruce; WH, wheat straw.

Only a few characterized hemicellulase genes were found outside these clusters (α-galactosidase genes 1 and 2, and xylanase 2 gene). A large number of putative regulatory genes clustered together with the known cellulase and hemicellulase genes. In particular, many genes encoding putative fungal C6 zinc finger-type transcription factors (containing InterPro domains IPR001138 fungal transcriptional regulatory protein, N-terminal and/or IPR007219 transcription factor, fungi [30]) were enriched within the clusters (P = 0.00027). Within the clusters, 5.9% of the genes encoded the predicted fungal type transcription factors, whereas only 2.5% of the total genome content belonged to this class. In addition, the clusters contained genes encoding candidates for other types of Zinc finger proteins, kinases and proteins involved in chromatin remodeling or organization, as well as proteins with InterPro domains indicating different regulatory or signal transduction functions (for the classes of the genes, see Table 1). A few known regulators were among the co-expressed genes, such as *xyr1,* the major regulator for cellulase and hemicellulase expression [18], and the homologues for *N. crassa clr-2* [25], *Aspergillus nidulans creC* [31] and *Fusarium oxysporum frp1* [24].

To cover putative regulatory genes induced by the substrates but showing different temporal patterns and extent of induction (and therefore not clustered together with the characterized cellulase and hemicellulase genes), the differentially expressed genes at each of the time points were identified by comparing the expression level in the induced cultures to the level in the uninduced control cultures (using Limma package (R, Bioconductor) [28,32], and the cut-off P <0.01 in the statistical analysis). Altogether, 89 genes with putative regulatory functions were either co-clustered with the characterized cellulase and hemicellulase genes or showed increased signal level in most of the inducing conditions studied.

In order to get further support for the relevance of the 89 candidate genes in cellulase and hemicellulase production and to narrow down the number of genes to be selected for further studies, the expression of the candidate genes was compared in additional datasets on different protein production conditions. Transcriptome and proteome data from chemostat cultures with different specific growth rate, cell density and specific protein production rate [14] were explored for expression of the candidate genes and production of the corresponding proteins. Expression of 14 candidate genes showed either positive or negative correlation (absolute value <0.5) to the specific protein production rate in the chemostat cultures. Proteome analysis of the same cultures [14] showed that the candidate GCN5-related N-acetyltransferase (123668) was more abundant in the cultures with higher protein (and cellulase) production level, whereas the SEC14-domain protein 81972 and the candidate GCN5-related N-acetyltransferase (120120) were more abundant in the cultures with low protein (cellulase) production. The results are in accordance with the positive and negative correlation of the expression of genes 123668 and 81972 with the specific protein production rate, respectively. Gene 120120 showed a slightly negative correlation with the specific protein production rate.

The CAZy genes are not randomly positioned in the genome. It has been reported that 41% of CAZy genes are found in 25 discrete regions ranging from 14 kb to 275 kb in length, and cases of co-expressed adjacent or nearly adjacent genes have been shown [7]. The regions of high CAZy gene density were found to contain genes encoding proteins involved in secondary metabolism. Our study also revealed the presence of regulatory genes in close vicinity to CAZy genes. In some cases, co-expression of these regulatory genes with CAZy genes was also detected. This information was used in the selection of candidate regulatory genes for further studies. For example, genes 76677 and 121130 are located in a broad, partly co-regulated region containing several CAZy genes. These genes include a candidate GH27 α-galactosidase

Table 1 Classes, functional domains and domain descriptions of the candidate regulators encoded by the genes that are co-regulated with cellulase and/or hemicellulase genes

Class	InterPro domain	Domain description	Candidate regulatory genes
Fungal transcription factors (Zn2-C6 type)	IPR001138	Fungal transcriptional regulatory protein, N-terminal	31
	IPR007219	Transcription factor, fungi	
Transcription factors, basic-leucine zipper type	IPR004827	Basic-leucine zipper transcription factor	1
	IPR011616	Basic-leucine zipper transcription factor, bZIP-1	
Transcription factor, Tcf25 type repressor	IPR006994	Transcription factor 25	1
Zinc finger, C2H2 type	IPR007087	Zinc finger, C2H2-type	2
Zinc finger, other types	IPR002893	Zinc finger, MYND-type	1
	IPR000058	Zinc finger, AN1-type	1
	IPR002867	Zinc finger, C6HC-type	1
	IPR008913	Zinc finger, CHY-type	1
	IPR000571	Zinc finger, CCCH-type	1
Chromatin level regulation/remodeling	IPR013256	Chromatin SPT2	1
	IPR000953	Chromo domain	1
	IPR008251	Chromo shadow	
	IPR000182	GCN5-related N-acetyltransferase	6
	IPR013178	Histone H3-K56 acetyltransferase, RTT109	1
	IPR001214	SET domain	2
	IPR000330	SNF2-related	1
	IPR001025	Bromo adjacent homology domain	1
	IPR001487	Bromodomain	1
	IPR000210	BTB/POZ-like	
Protein kinases	IPR000719	Protein kinase, catalytic domain	7
	IPR011009	Protein kinase-like domain	
G protein signaling	IPR011021	Arrestin-like, N-terminal	1
	IPR011022	Arrestin-like, C-terminal	
	IPR000832	G protein-coupled reseptor, family 2, secretin-like	1
	IPR000342	Regulator of G protein signaling	2
Other regulators	IPR000095	PAK-box/P21-Rho-binding	
	IPR000387	Dual-specific/protein-tyrosine phosphatase, conserved region	1
	IPR000791	GPR1/FUN34/yaaH	1
	IPR009057	Homeodomain-like	1
	IPR001611	Leucine-rich repeat	1
	IPR008030	NmrA-like	1
	IPR008914	Phosphatidylethanolamine-binding protein	1
	IPR012093	Pirin	2
	IPR011989	Armadillo-like helical	2
	IPR001313	Pumilio RNA-binding repeat	
	IPR001251	CRAL-TRIO domain	1
	IPR005511	Senescence marker protein-30	1
	IPR001810	F-box domain, cyclin-like	1
	IPR003892	Ubiquitin system component Cue	1
	IPR001680	WD40 repeat	4

gene (59391), a candidate GH2 β-mannosidase gene (59689) and characterized β-glucosidase (*bgl1*) and β-xylosidase (*bxl1*) genes (data not shown). Gene 102499 has an interesting location between a very tightly co-regulated region of CAZy genes and putative secondary metabolism genes (Figure 2, region 1). Gene 120120 is located in a co-expressed region including four genes of the hemicellulase gene-enriched cluster (cluster 35), and close to a second co-expressed region containing the candidate regulatory genes 74765, 55422 and the repressor gene *cre1* (Figure 2, region 2).

Interestingly, we found several loci where a β-glucosidase and/or putative sugar transporter gene is located next to a gene with a putative regulatory function and co-expressed with it. Genes 77513, 105263 and 121121 are located next to candidate β-glucosidase genes *cel1b*, *cel3e* and *cel3d*, respectively. The regions including genes 77513, 121121 and 26163 (the closest homologue for *N. crassa clr-2*) contain a putative sugar transporter gene (Figure 2).

The focus in selection of candidate regulatory genes for further studies was on the genes encoding putative

Region	Gene ID	Overexpression construct	Strand	Scaffold no.	Start	End	Class	Description	Mfuzz cluster	AV 0h	AV 6h	AV 17h	WH 0h	WH 6h	WH 17h	SP 0h	SP 6h	SP 17h	SO 0h	SO 6h	SO 17h	SO 41h
	73618		-	1	482161	490278	Secondary metabolism	Candidate polyketide synthase	24	0,7	-0,4	2,6	2,1	-4,2	-4,5	1,7	-3,5	-4,3	1,1	0,8	2,4	0,6
	73621		+	1	491045	499128	Secondary metabolism	Candidate polyketide synthase	24	0,4	0,1	2,3	1,1	-2,9	-4,1	0,8	-2,9	-3,8	0,7	1,0	2,1	0,6
	73623		+	1	499862	501401	Secondary metabolism	Candidate flavoprotein monooxygenase	24	0,1	0,2	2,1	1,3	-3,5	-4,8	1,3	-2,7	-3,6	0,8	0,9	1,8	0,9
	43701		-	1	501681	503623	Transport	Candidate multidrug-resistance transporter, MFS	32	-0,1	-0,3	0,7	0,4	-2,0	-3,6	-0,1	-1,8	-3,1	0,1	0,0	0,6	0,2
	102497		+	1	504139	506467	Regulatory functions	Candidate fungal transcriptional regulatory protein	24	-0,1	-0,4	1,1	0,0	-1,5	-2,5	0,1	-1,6	-2,6	0,3	0,1	1,1	0,3
1.	73631		-	1	506653	508537	Metabolism	Candidate FAD linked oxidase	15	0,1	0,1	1,7	1,2	-4,1	-4,0	1,2	-2,2	-3,1	0,5	0,8	2,0	0,8
	102499	**pMH12**	-	1	510531	512829	Regulatory functions	Candidate fungal transcriptional regulatory protein	25	0,1	1,5	3,4	0,7	-0,5	1,2	0,5	-0,4	1,1	-0,2	1,8	2,7	1,8
	102500		-	1	518168	518762	Secreted	Hypothetical protein	44	-0,4	0,6	0,8	0,4	2,0	2,3	0,8	0,0	1,2	0,5	-0,1	-0,2	-0,5
	73632		-	1	520228	521337	Carbohydrate esterase family 5	axe1, acetyl xylan esterase	35	-0,9	3,1	2,4	-0,7	3,6	3,2	-1,4	0,9	1,7	-2,1	0,9	1,3	2,2
	73638		-	1	523039	524217	Carbohydrate binding, CBM	cip1	35	-0,4	1,5	1,3	0,2	1,5	1,3	-0,1	0,9	1,4	-0,2	1,2	1,4	1,8
	73643		-	1	524767	526118	Glycoside hydrolase family 61	egl4/cel61a, candidate polysaccharide monooxygenase/endo-β-1,4-glucanase	35	0,0	1,5	1,3	0,8	2,1	2,1	0,3	1,3	1,8	0,0	0,9	1,3	1,4
	74765	**pMH25**	-	2	765243	767029	Nucleasome	Candidate BTB/POZ domain protein, possibly involved in histone deacetylation	23	0,0	1,8	1,0	0,1	2,0	1,9	0,1	1,4	1,6	0,0	1,4	0,8	0,3
	103653		+	2	767953	768491	Hypothetical protein	Hypothetical protein	23	-0,1	1,5	0,1	-0,2	1,8	0,5	-0,3	0,3	0,4	0,1	0,6	0,3	0,3
	55422	**pMH27**	+	2	773609	774500	Regulation & other	Candidate kinase domain protein	23	-0,1	0,9	0,6	0,2	1,0	1,1	-0,1	0,7	1,1	0,1	0,8	0,4	0,3
	103655		-	2	775052	776674	Replication and repair	Candidate RecA family protein	44	0,2	0,4	-0,1	0,6	0,8	0,7	0,4	0,4	0,3	0,1	0,7	0,3	-0,1
2.	120117		+	2	785322	788379	Regulatory functions	cre1, carbon catabolite repressor	31	0,2	0,2	0,1	0,1	0,4	0,1	0,4	0,1	0,3	0,0	0,0	-0,1	-0,2
	103660		-	2	794133	795265	Hypothetical protein	Hypothetical protein	35	0,0	0,4	0,0	0,6	1,2	0,5	0,3	0,9	0,7	0,9	0,6	0,9	0,4
	55790		-	2	798986	800882	RNA	Candidate N2,N2-dimethylguanosine tRNA methyltransferase protein	35	0,0	0,5	0,3	-0,1	0,5	0,9	-0,1	0,2	0,6	0,1	0,6	0,4	0,2
	55478		+	2	802255	802927	Hypothetical protein	Hypothetical protein	35	0,0	1,9	1,4	0,1	1,4	1,6	-0,1	0,5	0,8	0,5	1,3	1,4	1,1
	120120	**pMH22**	-	2	803502	804334	Regulatory functions	Candidate GCN5-related N-acetyltransferase, chromatin remodeling	35	-0,3	0,8	0,8	0,1	1,1	0,9	0,0	0,4	0,5	-0,1	0,5	0,9	1,1
	105263	**pMH16**	+	5	132264	134116	Regulation	Candidate fungal transcriptional regulatory protein	10	0,1	0,0	0,1	0,2	0,2	0,6	-0,1	0,2	0,4	0,4	0,3	0,9	0,8
3.	76227		+	5	134792	137417	Glycoside hydrolase family 3	cel3e, candidate β-glucosidase	10	0,4	0,8	1,2	0,7	0,1	0,6	0,3	0,7	0,6	0,4	1,2	2,1	2,3
	58701		-	5	137510	138803	Metabolism/protein transport	Candidate hexose-6-phosphate isomerase	10	0,3	0,6	0,9	0,2	0,4	0,7	0,0	0,5	0,3	0,0	0,9	1,6	2,1
	76659		-	6	6460	7804	Hypothetical protein	Hypothetical protein	35	-0,2	1,3	0,7	1,1	1,7	1,2	0,5	1,2	1,3	-0,5	1,0	0,8	1,8
4.	121121	**pMH10**	+	6	10751	13028	Regulation	Candidate fungal transcriptional regulatory protein	35	-0,1	1,0	0,5	0,3	1,3	1,5	-0,3	0,7	1,1	0,1	0,5	1,1	1,2
	46816		-	6	13220	15861	Glycoside hydrolase family 3	cel3d, candidate β-glucosidase	35	0,4	3,4	2,5	1,1	3,0	3,1	0,3	2,1	2,4	0,3	2,1	2,8	2,7
	46819		+	6	16768	18679	Transport	Candidate sugar transporter, MFS	10	-0,3	3,7	2,8	0,0	2,1	1,7	-0,9	1,5	1,1	-0,2	2,2	3,1	3,5
	106164		+	6	1352069	1353346	Metabolism	Candidate short-chain dehydrogenase/reductase, SDR	35	0,0	1,8	1,2	0,1	1,7	1,8	0,5	2,1	2,0	0,0	1,7	1,4	1,4
5.	3405		+	6	1355561	1357332	Transport	Candidate sugar transporter, MFS	10	-0,1	0,3	0,5	0,2	0,2	0,4	-0,2	0,3	0,5	-0,2	0,4	0,6	0,4
	26163	**pMH9**	+	6	1359971	1361974	Regulation	Candidate fungal specific transcription factor (clr2 homologue)	10	0,3	1,0	0,9	0,4	0,6	0,9	0,3	0,9	0,9	0,1	1,3	1,5	0,9
	77513	**pMH15**	+	8	422130	424168	Regulatory functions	Candidate fungal specific transcription factor	10	0,1	1,0	0,5	-0,2	0,1	-0,4	0,2	0,7	0,7	0,2	1,0	1,0	0,9
6.	22197		-	8	425315	427046	Glycoside hydrolase family 1	cel1b, candidate β-glucosidase	10	0,1	0,5	0,3	0,1	0,1	0,0	0,0	0,4	0,2	0,2	0,6	0,6	0,6
	77517		+	8	428282	430101	Transport	Candidate hexose transporter, MFS	10	0,0	1,0	0,6	-0,2	0,7	0,7	-0,1	0,8	0,9	0,1	1,0	0,9	0,6
	107857		-	10	699048	700756	Hypothetical protein	Hypothetical protein	35	0,1	0,7	0,8	0,3	0,7	1,1	-0,2	0,5	0,8	-0,4	0,8	0,8	-0,1
7.	107858	**pMH21**	-	10	701330	703726	Regulatory functions	Candidate fungal transcriptional regulatory protein	35	0,0	0,8	0,8	0,0	0,6	2,2	-0,3	0,7	1,0	-0,1	0,8	1,0	0,8

Figure 2 Tightly co-expressed genomic regions with candidate regulatory genes. The expression array dataset described in [8] was searched for genomic regions with co-expressed genes. The regions containing a selected candidate regulatory gene with adjacent genes belonging to the same Mfuzz gene expression clusters as the major cellulase and hemicellulase genes are shown. The genomic location of the genes is indicated as scaffold number, start and end position, and strand in the scaffold as in *T. reesei* database 2.0 [45]. Gene annotation is as in *T. reesei* database 2.0. The expression data of the genes in the induction dataset with cellulose, wheat and spruce material, and sophorose is shown as the expression cluster number (Mfuzz) and fold change of the transcript signals in the induced cultures as compared to the uninduced control cultures at the same time point. The intensity of the red color and blue color indicates the strength of positive and negative fold changes as compared to the uninduced control cultures, respectively. AV, Avicel cellulose; SO, sophorose; SP, spruce; WH, wheat straw.

transcription factors. The selected genes fulfilled several of the following criteria: induction by three or more of the cellulase- or hemicellulase-inducing substrates used in the study; co-clustering with the characterized cellulase and hemicellulase genes in the Mfuzz clustering of the expression data; correlation of the expression signal with specific protein production rate in the chemostat study [14]; increased signal of the corresponding protein under good protein-producing conditions in proteome analysis of the chemostat cultures [14]; and co-localization with cellulase and hemicellulase genes in the genome and, preferably, also co-expression of the co-localized genes. In addition, representatives of genes with functional domains indicating different regulatory functions and fulfilling the same criteria were selected. Altogether 28 genes were selected for further studies (Table 2).

The expression profiles of the selected candidate regulatory genes together with characterized cellulase and hemicellulase genes are represented as a heatmap in Figure 3. The heatmap shows fold change data of the signals in the induced cultures versus the signals in the uninduced cultures at the corresponding time points. Expression values of an additional dataset on cultures induced with a broader set of lignocellulose material (differently pretreated bagasse, oat spelt and birch xylans [8]) are also included. In the heatmap, the candidate regulatory genes are divided into three major groups. Genes 122523, 80291, 74765 and 123668 are co-expressed together with the gene cluster containing many of the known hemicellulase genes (cluster 35). The genes are moderately induced in the presence of the majority of the substrates used, but especially on wheat and spruce. The second group of candidate regulatory genes showed modest induction by the majority of the substrates (IDs 73792, 107858, 70351, 121130, 123019, 62244, 55422, 76677, 121121 and 56077). The third group clusters together with many of the

Table 2 Putative regulatory genes chosen for further studies and the functional domains present in the encoded proteins

Gene ID	Construct	InterPro ID	Description
108381	pMH8	IPR001138	Fungal transcriptional regulatory protein, N-terminal
26163	pMH9	IPR001138, IPR007219	Fungal transcriptional regulatory protein, N-terminal; Transcription factor, fungi
121121	pMH10	IPR001138	Fungal transcriptional regulatory protein, N-terminal
70351	pMH11	IPR001138	Fungal transcriptional regulatory protein, N-terminal
102499	pMH12	IPR001138, IPR007219	Fungal transcriptional regulatory protein, N-terminal; Transcription factor, fungi
62244	pMH13	IPR001138, IPR007219	Fungal transcriptional regulatory protein, N-terminal; Transcription factor, fungi
111742	pMH14	IPR001138,	Fungal transcriptional regulatory protein, N-terminal,
77513	pMH15	IPR007219	Transcription factor, fungi
105263	pMH16	IPR001138	Fungal transcriptional regulatory protein, N-terminal
112524	pMH17	IPR001138, IPR007219	Fungal transcriptional regulatory protein, N-terminal; Transcription factor, fungi
123668	pMH18	IPR000182	GCN5-related N-acetyltransferase
73792	pMH19	IPR001138	Fungal transcriptional regulatory protein, N-terminal
80291	pMH20	IPR001138, IPRO007219	Fungal transcriptional regulatory protein, N-terminal; Transcription factor, fungi
107858	pMH21	IPR001138	Fungal transcriptional regulatory protein, N-terminal
120120	pMH22	IPR000182	GCN5-related N-acetyltransferase
47317	pMH24	IPR001138, IPR007219	Fungal transcriptional regulatory protein, N-terminal; Transcription factor, fungi
74765	pMH25	IPR001487, IPR000210	Bromodomain; BTB/POZ-like
76677	pMH26	IPR001138, IPR007219	Fungal transcriptional regulatory protein, N-terminal; Transcription factor, fungi
55422	pMH27	IPR011009	Protein kinase-like domain
121130	pMH28	IPR001138	Fungal transcriptional regulatory protein, N-terminal
122523	pMH29	IPR001138	Fungal transcriptional regulatory protein, N-terminal
123019	pMH30	IPR000719	Protein kinase, catalytic domain
54703	pMH32	IPR007087	Zinc finger, C2H2-type
56077	pMH33	IPR001138	Fungal transcriptional regulatory protein, N-terminal
60215	pMH34	IPR001138	Fungal transcriptional regulatory protein, N-terminal
66966	pMH35	IPR001680	WD40 repeat
64608	pMH36	IPR001680	WD40 repeat
81972	pMH37	IPR001251	Cellular retinaldehyde-binding/triple function, C-terminal

Gene IDs are as in *T. reesei* database 2.0.

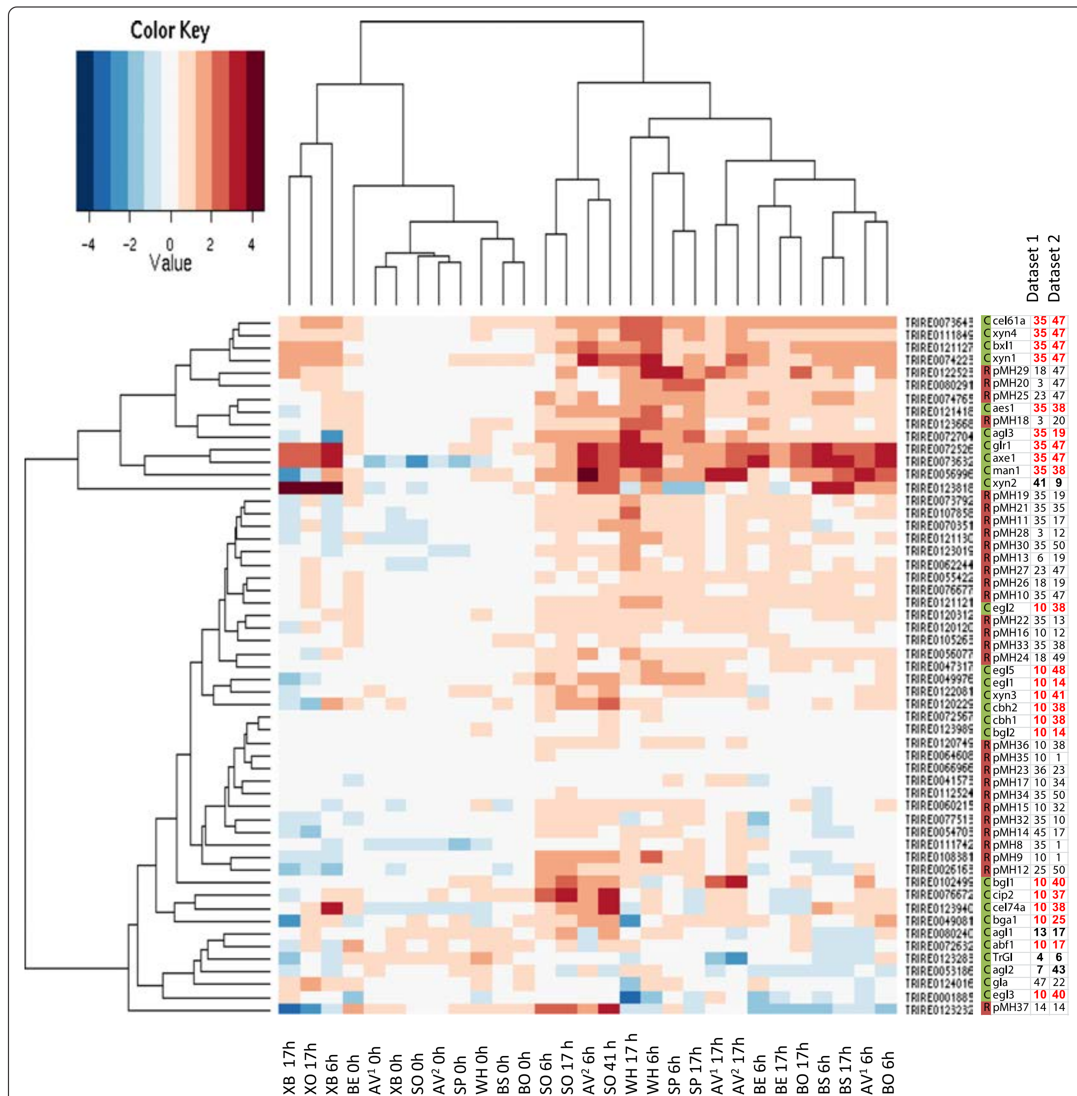


Figure 3 Heat map visualization of expression data on the known cellulase and hemicellulase genes and the putative regulatory genes in cultures induced with different lignocellulose substrates. The color key indicates the log2 scale fold change of the transcript signals in the induced cultures versus the uninduced control cultures at the same time point. The genes are shown as rows and the samples as columns. The legend on the right shows the gene ID and the cluster membership of the gene in Mfuzz clustering of the expression datasets. Dataset 1: Induction experiment with Avicel cellulose (0.75%), pretreated wheat straw, pretreated spruce or sophorose; Dataset 2: Induction experiment with Avicel cellulose (1%), bagasse, or xylans [8]. C: CAZy gene, R: regulatory gene. The legend below indicates the lignocellulose substrate in the culture and time point after addition of the substrate. AV^1, 0.75% Avicel cellulose; AV^2, 1% Avicel cellulose; BE, enzymatically hydrolyzed bagasse material; BO, untreated bagasse material; BS, steam-exploded bagasse material; SO, sophorose; SP, spruce; WH, wheat straw; XB, birch xylan; XO, oat spelt xylan.

genes in the cellulase-enriched cluster (cluster 10). This group includes genes induced mainly by sophorose, Avicel cellulose, wheat or spruce, but not with bagasse material, and genes hardly induced at all. Detailed transcriptional data of the genes is presented in Additional file 1.

Primary screening of the effects of the candidate regulatory genes on the cellulase and xylanase production of *T. reesei*

In order to investigate the effects of the putative regulatory genes chosen from the data, *T. reesei* QM9414

strains overexpressing the genes were constructed. The genes were cloned to an expression vector under the *A. nidulans gpdA* promoter and the expression plasmids were transformed to QM9414. A β-glucan plate assay was used for preliminary evaluation of enzyme production by the transformants and for selection of representative clones from the transformation for further analysis. The recombinant strains were cultivated in shake flasks on lactose containing rich medium to analyze the effect of the genetic modification on growth and protein production. Produced cellulase and xylanase activities (Figure 4) were measured throughout the cultivation. The growth of the strain transformed with the construct pMH12 was clearly defective as compared to the parental strain and to other recombinant strains, and was therefore omitted from further studies. The enzyme activity produced during the cultivation of the recombinant strains as compared to the activity produced in the cultures of the parental strain is summarized in Figure 4. Detailed information on production of the enzymatic activities during the time course of cultivation is shown in Additional file 2.

The strains overexpressing genes 77513, 74765, 80291, 66966, 123668, 64608 and 122523 (constructs pMH15, pMH25, pMH20, pMH35, pMH18, pMH36 and pMH29) produced cellulase and/or xylanase activity over 1.5-fold as compared to the parental strain in the shake flask cultures. The integrity of these seven strains and overexpression of the genes were confirmed by southern and northern blot analysis, respectively (Additional files 3 and 4). Most of the modified strains tested had the overexpression construct integrated as a single copy. The strain overexpressing the construct pMH35 had one to two copies according to the Southern hybridization. For the construct pMH15, both a single-copy and a double-copy transformant were analyzed (Figures 5 and 6). Northern analysis showed 1.4- to 23.6-fold overexpression of the gene for the strains analyzed (Additional file 4), except for gene 123668 (pMH18), which was expressed at a low level both in the overexpression strain and in the parental strain and therefore was not quantified. In addition, a number of the recombinant strains (transformed with constructs pMH8, pMH13, pMH21, pMH22, pMH24, pMH26 and pMH37) produced clearly less enzymatic activity than the parental strain. These genes were omitted from further studies.

Overexpression of gene 77513 (construct pMH15) had the most consistent and statistically significant (t-test; $P < 0.05$) positive effect on both cellulase and xylanase production by *T. reesei*. The strain produced in the initial screening 3- to 4-fold cellobiohydrolase 1 (CBHI) activity, 2- to 2.5-fold endoglucanase 1 (EGI) activity and 2- to 3-fold combined activity as measured against the 4-methylumbelliferyl-β-D-lactoside (MUL) substrate (Figure 4). The strain also produced 2- to 3-fold more xylanase activity as measured against the parental strain.

The strain overexpressing gene 80291 (construct pMH20) produced 2.5-times more CBHI activity, 2-times more EGI activity and 2.5-times more total activity against the MUL substrate. However, the xylanase activity was only slightly improved in this recombinant strain (less than 1.5-fold) as compared to the parental strain. The change in the production levels by pMH20 overexpression was statistically significant (t-test; $P < 0.05$).

The overexpression of gene 74765 (construct pMH25) produced the largest amount of cellulase activity as measured volumetrically against the substrate MUL, as compared to the other recombinant strains and to the parental strain (almost 3.5-times more than the parental strain). Production of xylanase activity was also increased more than 1.5 times in the recombinant strain. However, *T. reesei* EGI (CEL7B) has been shown to have activity against xylans as well and thus the increase in xylanase activity could be partly due to the increase in EGI production [33].

Quantitative PCR of cellulase and hemicellulase genes

Based on the preliminary enzyme activity measurements, strains overexpressing genes 77513, 80291 and 74765 (constructs pMH15, pMH20 and pMH25) were selected for further studies. For clarity, the recombinant strains will be referred to by the construct names. A quantitative PCR analysis of *axe1*, *bxl1*, *xyn1*, *xyn2*, *xyn3*, *cbh1*, *cbh2*, *egl1*, *bgl1* and *xyr1* was carried out. The results are shown as a fold change of the signals as compared to the parental strain QM9414 (Figure 7). For all the strains, the expression of *cbh1*, *cbh2* and *egl1* was improved as compared to the parental strain, although for pMH20 and pMH25 the effect was more moderate and was detected for pMH20 only at the 3-day time point. The expression of the major β-glucosidase gene *bgl1* was clearly improved by the pMH15 and pMH25 constructs but not by pMH20. Similarly, the expression of the three xylanase genes was improved by pMH15 and pMH25. Regarding xylanase gene expression, the overexpression of gene 77513 (pMH15) seemed to have most effect on *xyn3*, whereas the two other candidate regulatory genes were more specific to *xyn1* (only xylanase gene with improved expression by pMH20). Particularly, overexpression of gene 74765 (pMH25) had a major effect on the transcription of *xyn1*. The expression of *bxl1* was moderately improved with pMH15 and pMH25. The clearest increase in *axe1* expression was seen with pMH25. The expression of *xyr1*, which encodes the major regulator of cellulase and hemicellulase genes, was higher in pMH15 than in the parental strain but was not affected in the other two strains.

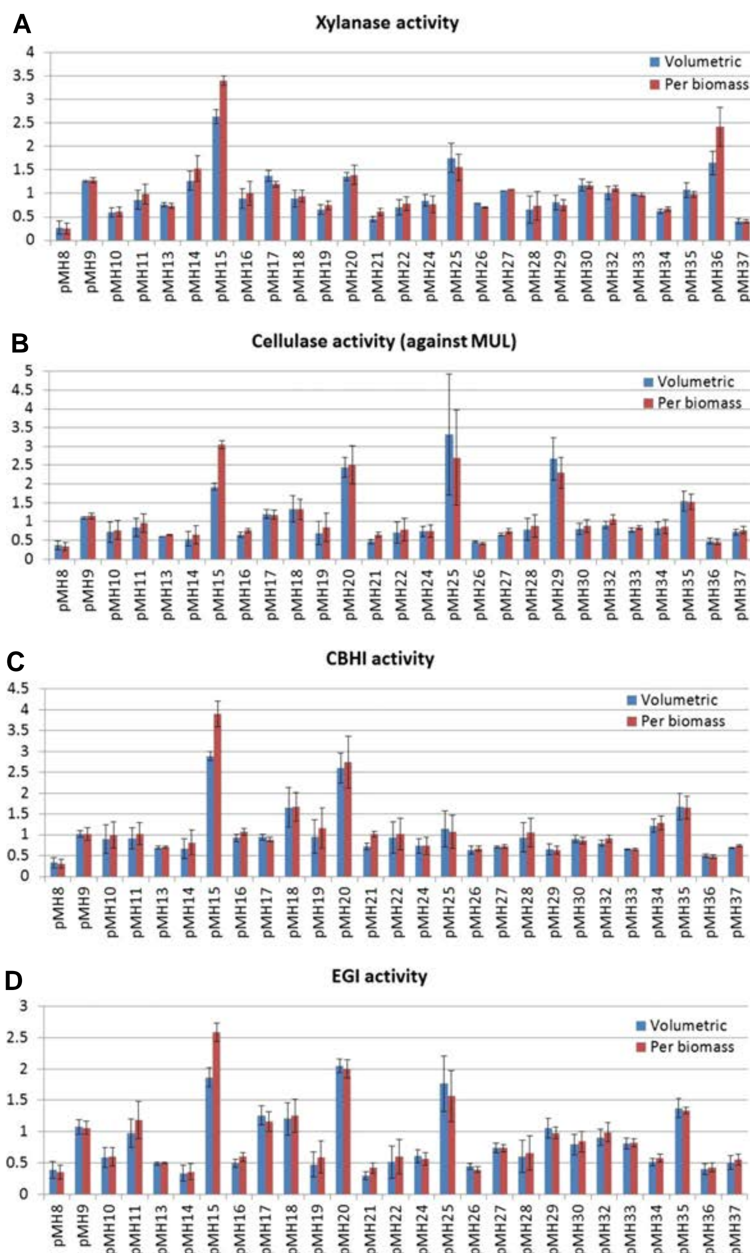


Figure 4 Cellulase and xylanase production by *T. reesei* QM9414 recombinant strains overexpressing the candidate regulatory genes. The volumetric enzyme production (blue bars) and production per biomass dry weight (red bars) are shown as the fold change of the maximum amount of activity produced in the cultures of the recombinant strains as compared to the maximum activity produced in the cultures of the parental strain. The values are means of three biological replicates. Error bars show the standard error of the mean. Panels **A** and **B** show the total xylanase activity against birch glucuronoxylan substrate and cellulase activity against 4-methylumbelliferyl-β-D-lactoside substrate, respectively. Panels **C** and **D** show the specific enzymatic activity produced by cellobiohydrolase 1 and endoglucanase 1. Detailed time course data on enzyme production in the cultures is shown in the Additional file 2. CBHI, cellobiohydrolase 1; EGI, endoglucanase 1; MUL, 4-methylumbelliferyl-β-D-lactoside.

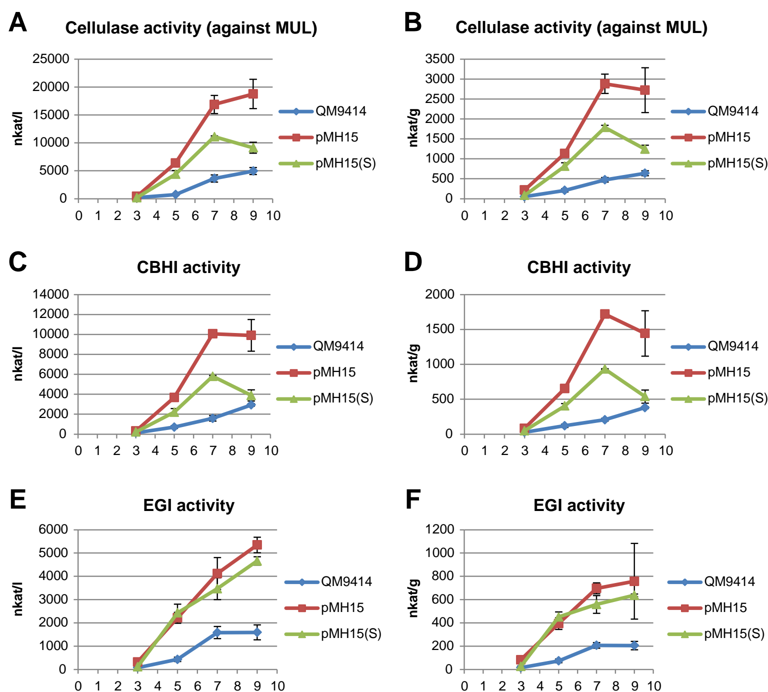


Figure 5 Production of cellulase activity by two different transformants overexpressing gene 77513. Transformants harboring the overexpression cassette as a single-copy (pMH15(S)) or as a double-copy (pMH15) were cultivated in shake flasks with lactose as a carbon source. Enzyme activity was measured at four different time points (3, 5, 7 and 9 days). The values are means of three biological replicates. Error bars show the standard error of the mean. Panels **A** and **B** show the volumetric and production per biomass dry weight of total cellulase activity against MUL substrate, respectively. Panels **C-F** show the specific enzymatic activity produced by CBHI and EGI. CBHI, cellobiohydrolase 1; EGI, endoglucanase 1; MUL, 4-methylumbelliferyl-β-D-lactoside.

Overexpression and deletion of gene 77513, designated as *ace3*

Based on the quantitative PCR and enzyme production results of the recombinant strain overexpressing the construct pMH15, gene 77513 was selected for more detailed studies. A recombinant strain was constructed from which gene 77513 was deleted (designated Del77513). We also analyzed enzyme production by strains having both one (pMH15(S)) or two (pMH15) copies of the overexpression cassette and in the 77513 deletion strain (all the constructs were confirmed by Southern and Northern analyses, Additional files 3 and 4). Both overexpression

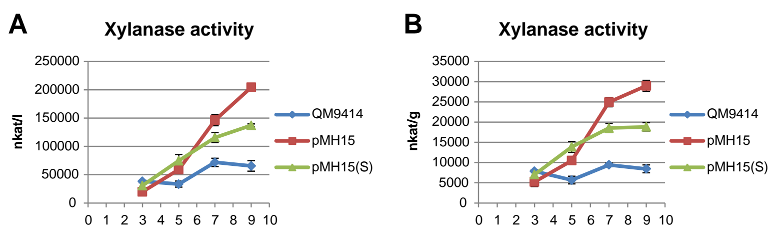


Figure 6 Production of xylanase activity by two different transformants overexpressing gene 77513. Transformants harboring the overexpression cassette as a single-copy (pMH15(S)) or as a double-copy (pMH15) were cultivated in shake flasks with lactose as a carbon source. Xylanase activity was measured at four different time points (3, 5, 7 and 9 days). The values are means of three biological replicates. Error bars show the standard error of the mean. Panels **A** and **B** show the volumetric and production per biomass dry weight of xylanase activity, respectively.

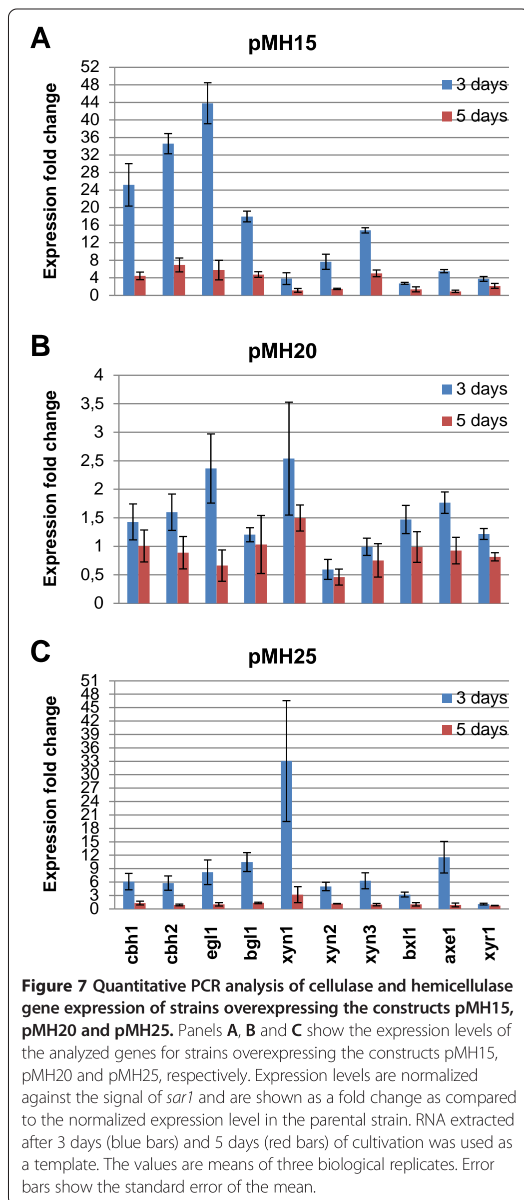


Figure 7 Quantitative PCR analysis of cellulase and hemicellulase gene expression of strains overexpressing the constructs pMH15, pMH20 and pMH25. Panels **A**, **B** and **C** show the expression levels of the analyzed genes for strains overexpressing the constructs pMH15, pMH20 and pMH25, respectively. Expression levels are normalized against the signal of *sar1* and are shown as a fold change as compared to the normalized expression level in the parental strain. RNA extracted after 3 days (blue bars) and 5 days (red bars) of cultivation was used as a template. The values are means of three biological replicates. Error bars show the standard error of the mean.

strains were cultivated in parallel with the deletion strain and the parental strains. Produced cellulase activity against the MUL substrate and xylanase activity were measured throughout the cultivation.

Both overexpression strains produced significantly (t-test; P <0.05) more total MUL activity, CBHI, EGI and xylanase activity as compared to the parental strain (Figures 5 and 6). The improvement in cellulase and xylanase production was higher in the double-copy strain than in the single-copy strain, indicating that the possible double-integration of the expression cassette also amplified the positive effect of the overexpressed gene to cellulase and xylanase production. When gene 77513 was deleted, the production of total cellulase activity against the MUL substrate was abolished completely (Figure 8). Interestingly, production of xylanase activity decreased to approximately half that of the parental strain (most significant decrease at day 7), indicating that gene 77513 is not essential for the production of xylanase activity but does modulate it (Figure 8).

A quantitative PCR analysis of *axe1*, *bxl1*, *xyn1*, *xyn2*, *xyn3*, *cbh1*, *cbh2*, *egl1*, *bgl1* and *xyr1* was carried out for samples collected from the cultivation of strains pMH15, pMH15(S) and Del77513. Due to the different parental strains of the overexpression strains and the deletion strain, the results are shown normalized with the signal of *sar1* (Figures 9 and 10). The expression of *cbh1*, *cbh2*, *egl1*, *bgl1*, *xyn1*, *xyn2*, *xyn3* and *xyr1* was higher in the overexpression strains as compared to the parental strain. In accordance with the enzymatic activity measurements, the increase in the gene expression was higher in the double-copy strain than in the single-copy strain. The expression of *bxl1* was improved only in the double-copy strain.

Expression of *cbh1*, *cbh2*, *egl1*, *axe1* and *xyn3* was almost undetectable in the deletion strain as compared to the parental strain. The expression of *bxl1*, *xyn1*, *xyn2*, *bgl1* and *xyr1* was also lower as compared to the parental strain. In the light of the enzymatic activity and quantitative PCR results for the two strains overexpressing gene 77513 and for the strain with the gene deleted, this gene was named *activator of cellulase expression 3 (ace3)*.

Discussion

The double-lock gene regulation mechanism, in which a master transcription factor regulates an additional trans-acting regulatory factor gene together with its actual target genes, is well-documented in filamentous fungi. In particular, carbon catabolite repression has been reported to be mediated by such a mechanism. In the model organism *A. nidulans*, the carbon catabolite repressor CREA regulates the ethanol utilization genes by repressing both the positively acting regulatory gene *alcR* and its target, *alcA* [34]. CREA also regulates lignocellulolytic genes by repressing the major activator (*xlnR*) as well as many of its target genes, for example, *xlnD* and *xlnB* [35]. Similarly, the major regulator of cellulolytic and xylanolytic genes in *T. reesei* (*xyr1*, a homologue of *xlnR)* is repressed by the carbon catabolite repressor CRE1 together with many *xyr1* target genes [18,36].

In this study, we utilized the principle of the double-lock mechanism to find new regulators of cellulase and

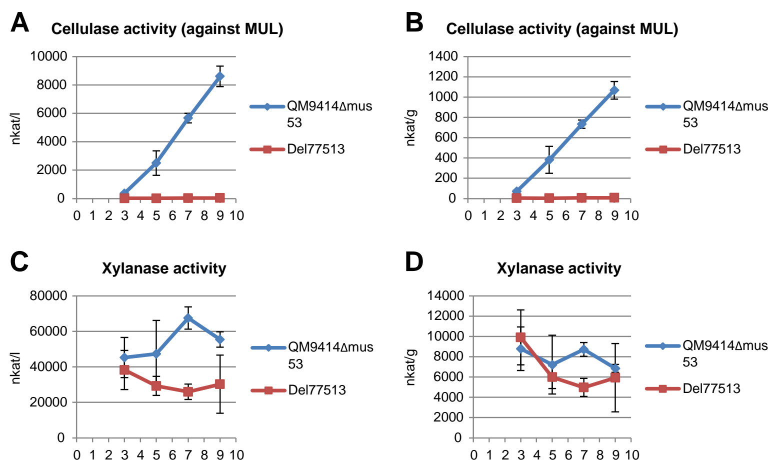


Figure 8 Production of cellulase and xylanase activity by the 77513 deletion strain. Del77513 was cultivated in shake flasks in lactose containing medium in parallel with the parental strain QM9414Δmus53. Enzyme activity was measured at four different time points (3, 5, 7 and 9 days). The values are means of three biological replicates. Error bars show the standard error of the mean. Panels **A** and **B** show the volumetric and production per biomass dry weight of total cellulase activity against MUL substrate, respectively. Panels **C** and **D** show the volumetric and production per biomass dry weight of xylanase activity, respectively. MUL, 4-methylumbelliferyl-β-D-lactoside.

hemicellulase genes, presuming that these regulators would be regulated in a similar manner as their target genes. We analyzed transcriptome data from *T. reesei* cultures induced with different lignocellulose-derived substances to search for candidate regulatory genes. This led to identification of 89 candidate genes that were co-induced with many of the known cellulase or hemicellulase genes in the presence of different lignocellulose-derived materials. We selected 28 genes for overexpression screening by taking into account supporting evidence from other genome-wide datasets, such as transcriptome and proteome analysis of chemostat cultures with different protein production rates [14], as well as location of the genes in the genome.

Clustering of the biosynthesis genes for fungal secondary metabolites together with their regulatory genes in the genome, as well as the regulatory cascades including chromatin-mediated regulation of the genomic regions, is relatively well-characterized in fungi (for a review, see [37]). Recent studies have indicated that chromatin level regulation also takes place in the regulation of CAZy genes of *T. reesei*. The putative methyltransferase LAE1, a homologue of LaeA functioning in chromatin level regulation of secondary metabolism in *Aspergilli*, has been shown to be involved in controlling cellulase gene expression in *T. reesei,* although the actual mechanism is not fully understood [38]. Furthermore, genes with significant up- or down-regulation during conidiation [39] as well as genes whose expression levels correlate with the specific production rate of extracellular proteins [14] have been shown to be non-randomly distributed in the *T. reesei* genome. Genes encoding, for example, secondary metabolism proteins, CAZys, putative transporters and putative transcription factors have been identified from such genomic clusters. In addition, the protein families of these regulators and the protein families of CAZys and secondary metabolism-related enzymes have recently expanded in the evolution of filamentous fungi, (*Pezizomycotina)* [40]. Thus, positioning of the regulatory genes in the close vicinity of their target genes (or other genes involved in the same process) may not be limited to the secondary metabolism genes, but could involve the genes active in lignocellulose degradation as well.

The transcriptome data on the cultures induced with different lignocellulosic material showed genomic regions that are co-regulated in an inducer-specific manner. Of the genes that were co-expressed with the major cellulase and hemicellulase genes according to the Mfuzz clustering, 22.7% were located in enriched genomic regions (≥ three genes within a window of nine genes, with a maximal distance of five genes). Of these, 9.1% (32 genes) were located next to each other in patches of three or more genes and were tightly co-regulated.

In addition to the known regulatory gene for hemicellulase and cellulase genes, *xyr1*, nine candidate regulatory genes were located in these tightly co-regulated regions or within close vicinity (Figure 2). Interestingly, four of the genes were located next to a putative sugar

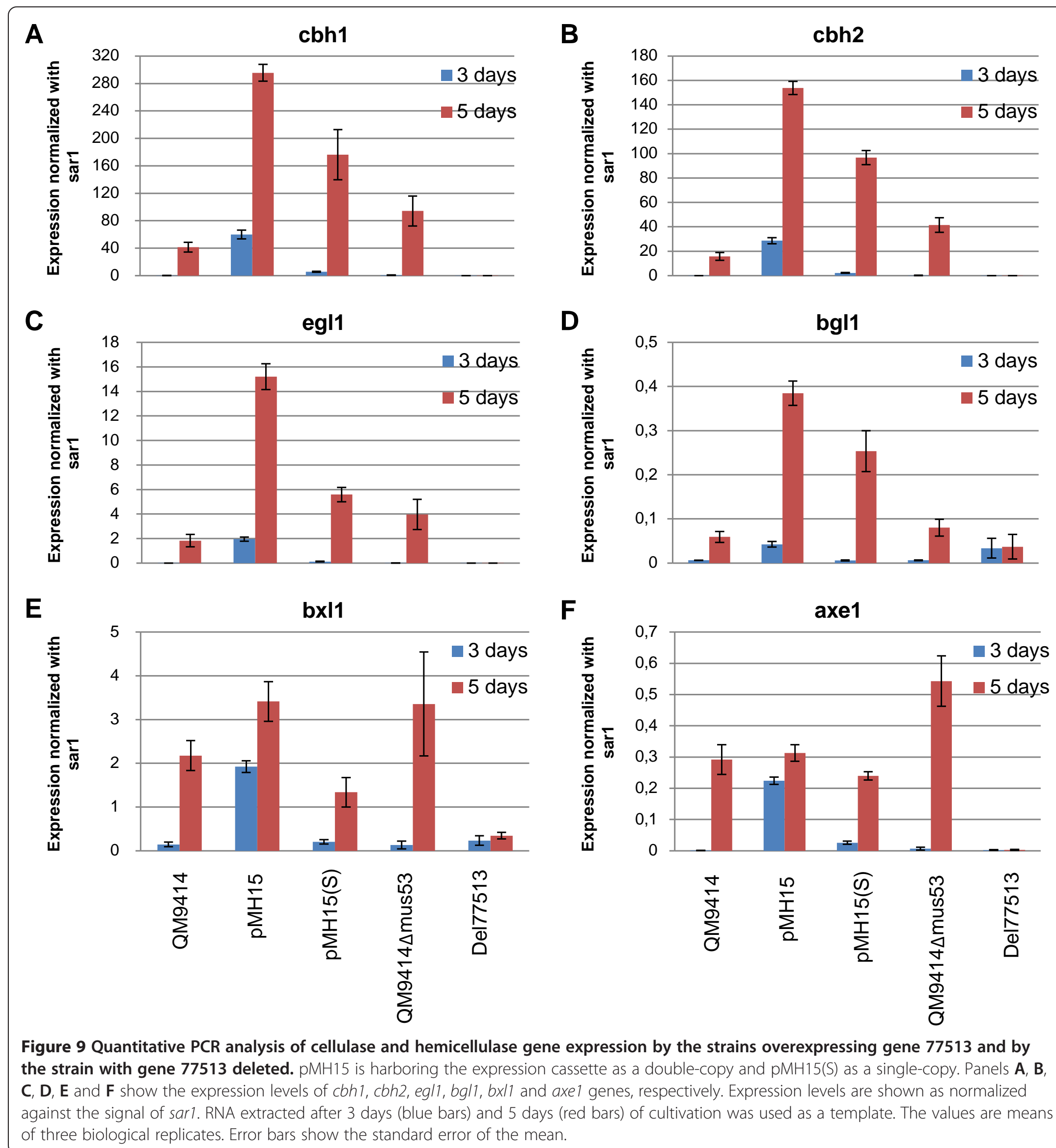


Figure 9 Quantitative PCR analysis of cellulase and hemicellulase gene expression by the strains overexpressing gene 77513 and by the strain with gene 77513 deleted. pMH15 is harboring the expression cassette as a double-copy and pMH15(S) as a single-copy. Panels **A**, **B**, **C**, **D**, **E** and **F** show the expression levels of *cbh1*, *cbh2*, *egl1*, *bgl1*, *bxl1* and *axe1* genes, respectively. Expression levels are shown as normalized against the signal of *sar1*. RNA extracted after 3 days (blue bars) and 5 days (red bars) of cultivation was used as a template. The values are means of three biological replicates. Error bars show the standard error of the mean.

transporter and/or a β-glucosidase gene. In addition to the release of glucose from cellobiose by extracellular β-glucosidases and transport of sugars into the cells, the sugar transporters and β-glucosidases may have a special role in the onset of the CAZy gene induction. Sugar units derived from the complex carbon source may be transported inside the cells and further modified by intracellular β-glucosidases to form an inducing compound, such as sophorose via a transglycosylation reaction. The gene *cel3e,* located next to gene 105263 (pMH16), encodes a predicted extracellular β-glucosidase. By contrast, *cel3d* and *cel1b,* located next to genes 121121 (pMH10) and 77513/*ace3* (pMH15), respectively, are predicted to encode intracellular enzymes. Interestingly, the sugar transporter genes located next to genes *ace3* and 26163 have recently been suggested to be involved in lactose uptake and cellulase production in lactose-containing media [41,42].

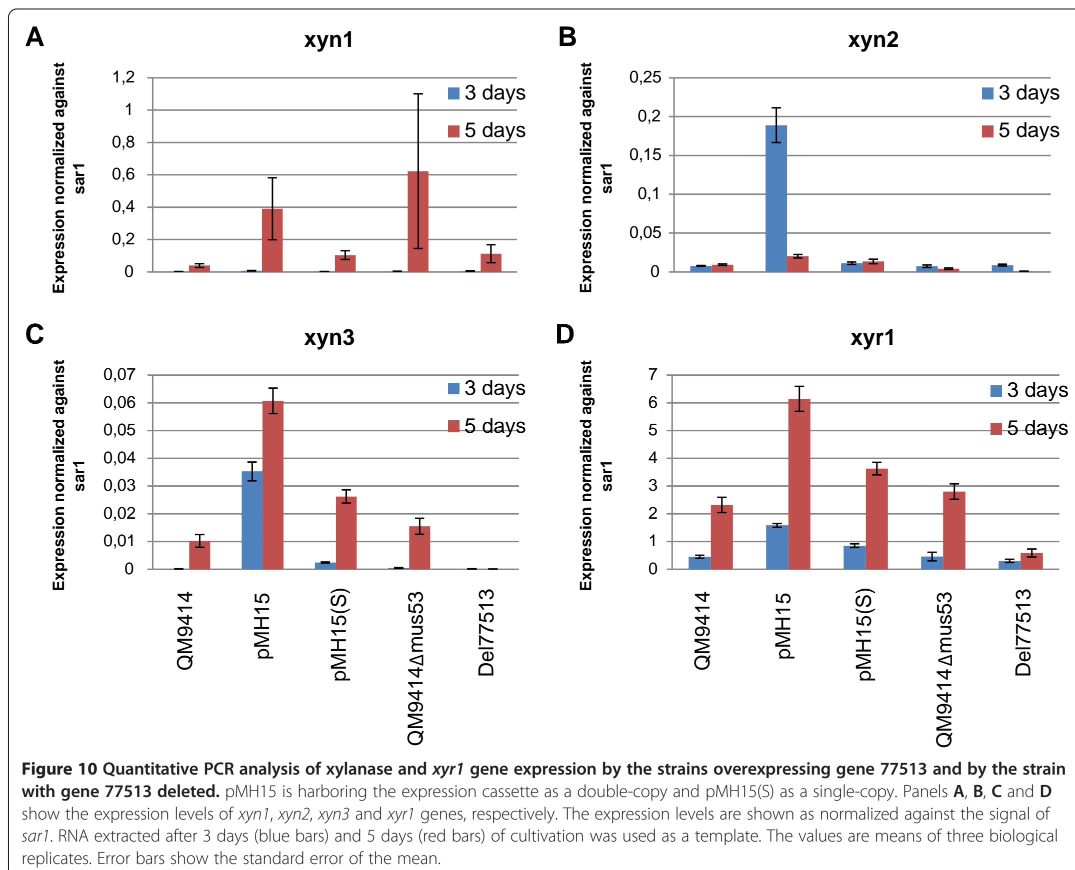


Figure 10 Quantitative PCR analysis of xylanase and *xyr1* gene expression by the strains overexpressing gene 77513 and by the strain with gene 77513 deleted. pMH15 is harboring the expression cassette as a double-copy and pMH15(S) as a single-copy. Panels **A**, **B**, **C** and **D** show the expression levels of *xyn1*, *xyn2*, *xyn3* and *xyr1* genes, respectively. The expression levels are shown as normalized against the signal of *sar1*. RNA extracted after 3 days (blue bars) and 5 days (red bars) of cultivation was used as a template. The values are means of three biological replicates. Error bars show the standard error of the mean.

Co-location of a putative regulatory gene with a β-glucosidase gene and a transporter gene is not a unique feature of the *T. reesei* genome. For example, the homologues of 77513/*ace3* (pMH15) in *A. fumigatus* (AFUA_016410) and in *A. clavatus* (ACLA_01970) are accompanied by a candidate β-glucosidase gene (AFUA_1G16400/ACLA_01980) and a candidate hexose transporter gene (AFUA_1G16390/ACLA_019190) next to it in the genome. Similarly, the homologues of gene 121121 (pMH10) in *A. fumigatus* (AFUA_7G00210) and in *A. nidulans* (ANIA_02615) are located next to a candidate hexose transporter gene (AFUA_7G00220/ANIA_02614), a candidate major facilitator superfamily multidrug transporter gene (AFUA_7G2613/ANIA_02614), and a β-glucosidase gene (AFUA_7G00240/ANIA_026142) [43].

In a recent study, it was suggested that, in *N. crassa*, the cellulase/hemicellulase regulator CLR-1 would promote the expression of cellodextrin transporters and β-glucosidase genes as well as a second regulatory gene, *clr-2,* which in turn activates cellulase genes [25]. In *N. crassa*, *clr-2* is essential for cellulase production in the presence of Avicel cellulose [25]. In *T. reesei*, the homologue of *clr-2,* gene 26163 (construct pMH9), is located next to a co-regulated sugar transporter gene that has recently been described as a lactose permease essential for the induction of *cbh1* and *cbh2* [42]. Overexpression of gene 26163 alone resulted only in a minute enhancement in production of cellulase and xylanase activity. However, no close homologue for *clr-1* can be identified from *T. reesei*, suggesting an important difference in the activation mechanisms of *clr-2*/26163 and/or the accompanying transporter genes in *N. crassa* and in *T. reesei.*

Overexpression of genes 105263 (pMH16) and 121121 (pMH10) did not have a significant effect on protein production under the conditions studied. However, overexpression of *ace3*, which is located next to a co-regulated β-glucosidase gene (*cel1b)* and a candidate sugar transporter gene in its original locus, resulted in a significantly increased production of cellulase and xylanase

activity as compared to the parental strain. Deletion of the gene was detrimental to the production of cellulase activity and decreased the production of xylanase activity. Quantitative PCR analysis of transcript levels of cellulase and xylanase genes supported the enzymatic activity measurements. Therefore, *ace3* can be considered to code for a novel master regulator of cellulase expression and a modulator of xylan degrading enzyme expression. Thus its role appears to be different from that of XYR1/XlnR, which has a major role in both xylan and cellulose degradation [18,44]. Interestingly, the Mfuzz clustering of *ace3* reflects the quantitative PCR results to some extent. The gene clustered together with *egl1*, *cbh1*, *cbh2*, *bgl1* and *xyn3, which* were most affected by *ace3* modifications, whereas *axe1*, *bxl1*, *xyn1* and *xyn2* are in different clusters.

Transcription of *xyr1* was increased in the strains overexpressing *ace3* and decreased in the deletion strain, indicating that the effects on the target genes observed could be at least partly mediated via *xyr1*. However, the deletion of *ace3* did not totally abolish *xyr1* transcription. Therefore, the absence of XYR1 is not an explanation for the total lack of cellulase activity and gene expression exhibited by the deletion strain.

Conclusions

Combining genome-wide data on cultures with different protein production properties is a useful method for identifying novel regulatory genes relevant for cellulase and xylanase production in *T. reesei*. Altogether, overexpression of seven of the candidate regulatory genes resulted in improved (>1.5 fold) production of cellulase and/or xylanase activity as compared to the parental strain. Further studies are required to confirm the role of most of these genes in cellulase and hemicellulase gene regulation and to elucidate the actual regulatory mechanisms. However, our data show a positive effect of cellulase and/or xylanase gene expression for three of the candidate regulatory genes. The deletion of one of these genes, *ace3*, totally abolished cellulase expression and reduced xylan degrading enzyme expression, thus identifying it as a novel master regulator of lignocellulose degradation. Furthermore, our data reveal genomic regions enriched in co-regulated CAZy genes and candidate regulatory genes, therefore supporting the hypothesis that chromatin-level regional regulation plays a role, at least in part, in the expression of CAZy genes in *T. reesei*.

Methods

Strains, media and culture conditions

Escherichia coli DH5α (*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) was used for propagation of the plasmids. *T. reesei* Rut-C30 (ATCC 56765, VTT-D-86271), QM6a (ATCC13631, VTT-D-071262 T) and QM9414 (ATCC 26921, VTT-D-74075) were obtained from VTT Culture Collection (Espoo, Finland). Spore suspensions were prepared by cultivating the fungus on potato-dextrose plates (BD, Sparks, Maryland, USA) for 5 days, after which the spores were harvested, suspended in a buffer containing 0.8% NaCl, 0.025% Tween20 and 20% glycerol, filtered through cotton, and stored at −80°C. For DNA isolation, the fungus was grown in a medium containing 0.2% proteose peptone (BD), 2% glucose, 7.6 g/l $(NH_4)_2SO_4$, 15.0 g/l KH_2PO_4, 2.4 mM $MgSO_4.7H_2O$, 4.1 mM $CaCI_2.H_2O$, 3.7 mg/l $CoCI_2$, 5 mg/l $FeSO_4.7H_2O$, 1.4 mg/l $ZnSO_4.7H_2O$ and 1.6 mg/l $MnSO_4.7H_2O$, pH 4.8.

Transcriptional profiling data

Transcriptional profiling data used in the study have been described elsewhere [8]. In short, pre-cultures of *T. reesei* Rut-C30 were first cultivated on a minimal medium containing sorbitol as a carbon source. Cellulase and hemicellulase gene expression was induced by addition of different lignocellulose material, purified lignocellulose-derived polymers or specific disaccharides (Cultivation set 1: addition of Avicel cellulose, pretreated wheat straw, pretreated spruce or sophorose; Cultivation set 2: addition of Avicel cellulose, birch xylan, oat spelt xylan, or differentially pretreated bagasse). Wheat straw and spruce were pretreated using steam explosion. Three different pretreatment methods were applied to bagasse, including grinding of the untreated bagasse material, steam explosion, or steam explosion followed by enzymatic treatment. Enzymatic pretreatment was done with a commercial cellulase and hemicellulase mixture followed by a protease treatment. Samples for transcriptional profiling were collected at different time points of induction (0, 6 or 17 h).

Custom-made microarray slides from RocheNimbleGen were used for transcriptional profiling. Sample preparation, hybridization onto microarray slides and collection of raw data was carried out as instructed by Roche. The microarray data were analyzed using the R package Oligo for preprocessing of the data and the package Limma for identifying differentially expressed genes [28,32]. In the analysis of the differentially expressed genes, the signals in the samples of the induced cultures were compared to the ones in the uninduced control cultures at the corresponding time point as described in [8]. Four biological replicates of each condition and time point were analyzed. The cut-off used for statistical significance was $P < 0.01$, and an additional cut-off for the log2 scale fold change was set as 0.4. In addition, the expression array datasets were clustered using the R package Mfuzz [29]. Co-expressed genomic clusters were determined by enrichment of Mfuzz cluster members in the genomic regions. Three or more

gene members of the expression cluster within a window of nine neighboring genes and with the maximal distance of five genes were considered as a genomic region enriched with co-regulated genes. In addition, genomic regions with multiple adjacent genes belonging to the same expression cluster were searched for.

The expression of the selected candidate regulatory genes was compared to the transcriptome and proteome data described in [14].

Construction of *T. reesei* strains overexpressing candidate regulatory genes

The regulatory genes were amplified by PCR using Gateway compatible primers (Table 3) and the genomic DNA of *T. reesei* QM6a as a template. For the majority of the genes, the open reading frame (ORF) predictions used were as in the genome version 2.0 [45] with the following exceptions: the primers for genes 26163 and 64608 and the N-terminal primer for gene 47317were designed according to the ORF prediction in archived genome version 1.0 [46], and the ORF prediction for gene 64608 was modified by taking into account expressed sequence tag sequence data. In order to construct the plasmid vectors for overexpression of the genes in *T. reesei*, the PCR fragments were inserted in the expression vector pMS204 using the Gateway recombination system (One-Tube protocol) according to the manufacturer's instructions (Invitrogen, Carlsbad, California, USA). The expression vector contains the hygromycin resistance gene (ZP_12918108) under the *A. nidulans gpdA* promoter [47] and *trpC* terminator [48], as well as an additional copy of the *gpdA* promoter and *trpC* terminator for expression of the gene of interest (the vector map is illustrated in Additional file 5). The plasmids were linearized using *HindIII*, *PciI* or *SpeI* enzyme (New England BioLabs, Ipswich, Massachusetts, USA) and transformed to *T. reesei* QM9414 by polyethylene glycol-mediated protoplast transformation [49]. The transformants were selected for hygromycin resistance on plates containing 150 μg ml^{-1} of hygromycin B (Calbiochem, San Diego, California, USA). Stable transformants were obtained by streaking on plates containing 125 μg ml^{-1} of hygromycin B for two successive rounds, after which single colonies were obtained by plating dilutions of spore suspensions. Integration was verified by PCR with one primer binding the *gpdA* promoter and one binding the ORF of the overexpressed gene (the primers used are listed in Table 4). The cellulase production levels of transformants from each construct were assayed on β-glucan plates (see below). Southern blot analysis was carried out for additional confirmation of the transformants showing improved protein production as compared to the parental strain. Genomic DNA was isolated using an Easy-DNA Kit (Invitrogen) according to manufacturer's instructions. Southern blotting and hybridization on nitrocellulose filters (Hybond N, GE Healthcare, Little Chalfont, UK) were carried out according to standard procedures [50]. Probe fragments were PCR-amplified from the genomic DNA. The signals were detected using a phosphorimager (Typhoon imager, GE Healthcare).

Plate assay for β-glucan hydrolysis using Congo red staining

For detection of enzymatic activity against the β-glucan produced by fungal colonies, spores were mixed with 50°C top agar containing 0.1% β-glucan (Megazyme, Bray, Wicklow, Ireland), 2% lactose (Fagron, Rotterdam, the Netherlands), 0.05% proteose peptone (BD), 7.6 g/l $(NH_4)_2SO_4$, 15.0 g/l KH_2PO_4, 2.4 mM $MgSO_4.7H_20$, 4.1 mM $CaCI_2.H_2O$, 3.7 mg/l $CoCI_2$, 5 mg/l $FeSO_4.7H_2O$, 1.4 mg/l $ZnSO_4.7H_2O$, 1.6 mg/l $MnSO_4.7H_2O$, 0.1% Triton TX-100 (Fluka, St Louis, Missouri, USA) and 3% agar Noble (BD), pH 5.5, and plated on solid medium (composition of the medium was the same as that of the top agar except that β-glucan was omitted and the concentration of agar Noble was 1.8% (w/v)). After 4 days of cultivation at 28°C, the plates were rinsed with 0.9% NaCl, submerged in 0.1% Congo red (Merck, Darmstadt, Germany) in 1 M Tris-HCl (pH 9.5), and incubated for 30 min with shaking at 100 rpm. After the incubation, the plates were washed with 0.9% (w/v) NaCl, and the diameter of the colonies and the halo around them were measured. The size of the halo compared to the colony size was calculated and compared to the corresponding size of the parental strain QM9414.

Construction of a deletion strain

The deletion cassette for the deletion of gene 77513 was constructed by Golden Gate cloning [51]. The construct contained the hygromycin resistance cassette (gpdA promoter, hygromycin resistance gene, trpC terminator) flanked by 1.523 kb and 1.024 kb fragments from the 5′ and 3′ sides of the ORF of 77513, respectively. The 5′-flanking region fragment was amplified by PCR with oligos 5′-GCGCGGTCTCCGGGTGGCGAGGTGGGAGA AGGGGA-3′ and 5′-GCGCGGTCTCGCATGGGAAG ACGAGGTCGGTGTTG-3′. The 3′-flanking region was amplified by PCR with oligos 5′-GCGCGGTCTCCGA GAAAGCGGTCGGGGAAATGGCG-3′ and 5′-GCGC GGTCTCGGCGGTTGCGTGGGCGTT GCTCGAT-3′. The fragments of the marker cassette and the flanks were first ligated to a pBsV2 vector [52] and subsequently cloned to a modified pBluescript vector (lacking the *BsaI* site). The deletion cassette was digested from the vector with *PmeI* enzyme and transformed to *T. reesei* QM9414Δmus53 strain (QM9414 strain from which gene 58509 had been deleted) with high targeted integration frequency.

Table 3 Gateway compatible primers for the cloning of the putative regulatory genes

Construct	5′primer	3′primer
pMH8	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGGCGCTCTTTGTCTGCTTGG	3′GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATTGTTGCTGCCCGCCCCA
pMH9	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTACATCATGTTCTACACATGTGG	3′GGGGACCACTTTGTACAAGAAAGCTGGGTTTCACGACGGCGGTAGAGC
pMH10	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTCGCACTAGAGCACAATGGAGAC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTGCTACTTCTGTATACACTTAATCAC
pMH11	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGGCCTCCAATGCCAAC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTTTCATAATCAGACCAGCTCTTTC
pMH12	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGGGGAGCAGCGCCA	3′GGGGACCACTTTGTACAAGAAAGCTGGGTTCTAGCCGTAAATCTATGTAGTTGA
pMH13	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGCCACGCCCAAAAGTCCACC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGAACCCAAACGCCCGCGG
pMH14	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGGCGTCCTCTTACGGCACTC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTGTTAGAATACTAAACTCTTCGC
pMH15	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGCTGCGCTACTCCCCCGTCT	3′GGGGACCACTTTGTACAAGAAAGCTGGGTTTTAGCCAACAACGGTAGTGGA
pMH16	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGACCAGCTCGGACGATTCCA	3′GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGTGAAGGAGGGCGGTAT
pMH17	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGGCTGGATCGCCTGCTGCTG	3′GGGGACCACTTTGTACAAGAAAGCTGGGTACACATTCATCCCTGCGCCCAG
pMH18	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGCCTCTCGTTGTCGTCCCAG	3′GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAATTGAGCAGCGGCTCGCG
pMH19	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGGATCTGCAATCCTTTGACA	3′GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACAGACGCTTTCCGAAAAAG
pMH20	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGGGCCGGCAACCGAGACAAC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTATTATATAAACGGGGCATCAAT
pMH21	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGGTTCGAGGCACCGGATC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTTCTAAGAAACATCTTCCGACCTGA
pMH22	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGTCCCGCCAAATCTCC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTCTTACTCGGTGCTGATACTTCT
pMH24	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGTCGTCAAACGCTTCACCG	3′GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAGCCCAAATGGCCCATATTG
pMH25	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGACTTCTGAAGCCCCCTCTC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTCGCCCTCTTCGCCTC
pMH26	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGGCCGACACCCCGACTC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTTTTAGAAGCCCGCCTGCTCTGC
pMH27	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCACAATGCCTCGCCGCGCCT	3′GGGGACCACTTTGTACAAGAAAGCTGGGTTTCATTCATCGCCCCAGAACAA
pMH28	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGAACATGACGACAACGCT	3′GGGGACCACTTTGTACAAGAAAGCTGGGTTCTATCTATAACTTGGTATTTTGC
pMH29	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGGTAGCACATAGTCTACCCTCT	3′GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATATCGGCACCATGTCG
pMH30	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGCTCATCAACAACCTCGATCC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTGTCAGACGAAACGCCGCCAG
pMH32	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGTTCCCGTACGGTGCC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACATACCCATAATCATTCCTC
pMH33	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGACATCAATAACGCATCCCTC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTTGCCTTCATCTCCTGGTGGAAT
pMH34	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGTTCGGACAGTACTCACTCG	3′GGGGACCACTTTGTACAAGAAAGCTGGGTATTACGACAATGGCAAGATCCT
pMH35	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGGCCAAGAAGGCGCGTC	3′GGGGACCACTTTGTACAAGAAAGCTGGGTGCTAGGCGCCGTTGACGACTC
pMH36	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGTCCAGATTCTGTCCGCT	3′GGGGACCACTTTGTACAAGAAAGCTGGGTGCATCAATAGGCCGTATCAGAG
pMH37	5′GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCGCATCATGACGGCCGAGTACGAAG	3′GGGGACCACTTTGTACAAGAAAGCTGGGTCTCGTTCATGCAGCAGCAG

Table 4 Primers for the PCR screening of the overexpression strains

Primer	Sequence 5′- 3′
pgpdA	GGCAGTAAGCGAAGGAGAATG
pMH8R1	ACACGGCTTCTTATATCTCGACC
pMH9R3	ATGGTCTCGATGTGGCTGCT
pMH10R1	CTGCGAGAGCAGCTAGGAGC
pMH11R2	CGTCGATTCGCGCTTGAACA
pMH12R2	GATGCACGCCGCCATCGAGT
pMH13R2	TCGTTCTCCTCGTAGATTCAG
pMH14R1	GCTGGCTCTTCTCCCTCACAC
pMH15R3	TGAGTATAGCGGCTGACTTGTCG
pMH16R3	CTCGTTGACTTGCAGGCCTTG
pMH17R1	CTGAGGGCTGTAGACGCACTC
pMH18R1	TTACAGAGGTGAGACTTTCCCT
pMH19R1	TTGCGTTGCGCCTTTACC
pMH20R3	TCGAGACGATGCAGCGATAG
pMH21R1	TGGTTCTGGATCACTCGTCA
pMH22R1	TTCGTCCTCCGTCTTGAGCA
pMH24R2	CTCACCTCGTCGTACACACTA
pMH25R1	ATGCGGTTGACTTGACAGAT
pMH26R2	GGTTGACTCTGGATGTTGGA
pMH27R1	ATCTTGACGTCCTTGTCGAT
pMH28R1	GCGAATCGACCAGATCGTGT
pMH29R1	GTCCTTGCACCGCTTACACG
pMH30R2	GTAGAAGCGCAATGCGGTGG
pMH32R2	CAGATGCACGTCTTCCAGAT
pMH33R1	TCTGGTCTCGATTGCTCGTG
pMH34R1	CATCAGCCTCGTCTCCAGCA
pMH35R3	CATCATCAATGTCCTCGAAG
pMH36R1	GTCAGGATAGCGCCTGTCTG
pMH37R1	GTCCGGTACAGCGTGTCAAT

Primer named pgpdA was used in combination with the gene specific primers.

Cultivation of *T. reesei* in shake flasks

T. reesei QM9414 and representative clones from transformations of each of the regulatory factor constructs were cultivated on medium containing 4% lactose (Fagron), 2% spent grain extract, 7.6 g/l $(NH_4)_2SO_4$, 15.0 g/l KH_2PO_4, 2.4 mM $MgSO_4.7H_2O$, 4.1 mM $CaCl_2.H_2O$, 3.7 mg/l $CoCl_2$, 5 mg/l $FeSO_4.7H_2O$, 1.4 mg/l $ZnSO_4.7H_2O$ and 1.6 mg/l $MnSO_4.7H_2O$, pH adjusted to 5.2 with KOH. The culture medium was inoculated with 2×10^7 spores per 200 ml of the medium, and grown at 28°C in conical flasks with shaking at 250 rpm for 10 days. The strains were cultivated in triplicate. Samples were collected after 3, 5, 7 and 9 or 10 days of cultivation. For RNA isolation, mycelium was collected by filtering the samples, and the mycelium was washed with equal volume of 0.7% NaCl, frozen immediately in liquid nitrogen and stored at −80°C. For measurement of the biomass dry weight, the filtered and washed mycelium samples were dried at 105°C to constant weight (24 h). Filtered culture media was used for enzymatic assays and for measuring pH.

Enzyme assays

Cellulase activity against the MUL substrate, CBHI and EGI activity was determined by detecting the fluorescent hydrolysis product methylumbelliferone released from the substrate MUL (Sigma-Aldrich, Steinheim, Germany) as described in [53]. The combined activity of EGI and CBHI was measured by inhibiting β-glucosidase activity with glucose. EGI activity was measured by adding cellobiose to inhibit CBHI and glucose to inhibit β-glucosidase. CBHI activity was deduced by subtracting EGI activity from the combined CBHI and EGI activity. Endo-β-1.4-xylanase activity was assayed using 1.0% birch glucuronoxylan as a substrate [54]. The released reducing sugars were detected with 2-hydroxy-3,5-dinitrobenzoic acid. Pure xylose (Sigma-Aldrich) was used as a standard.

Northern analysis

Total RNA was isolated from the mycelium samples using the Trizol™ Reagent (Gibco BRL, Carlsbad, California, USA), essentially according to manufacturer's instructions. Northern blotting and hybridization on nitrocellulose filters (Hybond N, GE Healthcare) were carried out according to standard procedures [50]. Fragments of the genes to be analyzed were PCR amplified from the genomic DNA and used as probes in the Northern analysis. The signals in the northern blots were quantified using a phosphorimager (Typhoon imager, GE Healthcare), and the signals were normalized with those of actin.

Quantitative PCR

Total RNA was isolated from the mycelial samples of three parallel cultivations collected at the cultivation time

Table 5 Primers for the quantitative PCR analysis

Gene	5′ primer	3′ primer
cbh1	GCGGATCCTCTTTCTCAG	ATGTTGGCGTAGTAATCATCC
cbh2	TCCTGGTTATTGAGCCTGAC	GCAACATTTGGAAGGTTCAG
egl1	GTCTACTACGAACTCGAC	GTAGTAGTCGTTGCTATACTG
bgl1	GCCTCCAAGATCAGCTATCC	ACCTCCTCACCGATGAACTG
xyn1	AAACTACCAAACTGGCGG	TTGATGGGAGCAGAAGATCC
xyn2	CGGCTACTTCTACTCGTACTG	TTGATGACCTTGTTCTTGGTG
xyn3	TACAAGGGCAAGATTCGTG	ACTGGCTTCCAATACCGT
axe1	TAAAGCAGCAATCTTCATGG	GCAGTAAGACTTGATCTTGG
bxl1	GTCACTCTTCCAAGCTCAG	ATCGTTACCTCTTCTCCCA
xyr1	GAGTATCAGCGCAACTTTAGCA	CATCGGTATAGTGCAAGAAGCTC
sar1	TCTCCACCCTACTTCTGAG	CTTGTTGCCCAGGATGAC

points 3 and 5 days using Trizol™ Reagent (Gibco BRL), essentially according to manufacturer's instructions. Single stranded cDNA was synthesized using a QuantiTect Reverse Transcription Kit (Qiagen, Hilden, Germany), with 1.5 μg of total RNA as a template. The cDNA samples were diluted 1:10 to 1:50 and quantitative PCR reactions of two technical replicates were performed using a LightCycler 480 SYBR Green I Master Kit (Roche, Mannheim, Germany) according to the instructions of the manufacturer. The instrument used for quantitative PCR was Light Cycler 480 II and the results were analyzed with LightCycler 480 Software release 1.5.0. (version 1.5.0.39) using *sar1* signal for normalization. The primers used in the quantitative PCR are listed in Table 5.

Additional files

Additional file 1: Transcriptional profiling data of the putative regulatory genes. Gene identifiers are as in *T. reesei* database version 2.0. Functional Interpro domain identifiers are as in InterPro database. Fold changes (log2 scale compared to uninduced control culture at a corresponding time point), signal intensities (log2 scale) and significance test (R package limma, $P < 0.01$, log2 fold change >0.4) are shown for the genes. 1 indicates induction and −1 repression. The intensity of the red color and blue color indicates the strength of positive and negative fold changes, respectively. Color scales of yellow, red and green indicate different intensities of signals, red represents the strongest signals and green the weakest signals. AV1, 1% Avicel cellulose; AV0.75, 0.75% Avicel cellulose; BE, enzymatically hydrolyzed steam-exploded bagasse; BO, ground bagasse; BS, steam-exploded bagasse; SO, sophorose; SP, steam-exploded spruce; WH, steam-exploded wheat straw; XB, birch xylan; XO, oat spelt xylan.

Additional file 2: Production of total proteins and cellulase and xylanase activity by the recombinant strains at different time points of the cultivation. Results are shown for each strain volumetrically (nkat/l) and per biomass dry weight (nkat/g). The values are means of three biological replicates. Error bars show the standard error of the mean. BGL, β-glucosidase activity; CBHI, cellobiohydrolase activity; EGI, endoglucanase activity; MUL, total cellulase activity measured against the substrate 4-methylumbelliferyl-β-D-lactoside; XYN, xylanase activity.

Additional file 3: Results of Southern hybridizations. Position of the molecular weight size marker is shown as kb on the left. The restriction enzymes used for the digestion in the analysis are indicated by the letters: A, *NcoI + BstXI*; B, *BglII*; C, *SpeI + BclI*; D, *ClaI + BamHI*; E, *SacI*; F, *NaeI*; G, *ClaI + XbaI*; H, *SnaBI + XbaI*; I, *StuI*; J, *SacI*; K, *StuI*; L, *XmnI*; M, *BstEII*; N, *SspI*; O, *StuI*; P, *SspI*; Q, *StuI*. For Del77513 strain two different probes were used: hygromycin selection marker (hph) open reading frame (N and O) and fragment of the gene 77513 open reading frame (P and Q).

Additional file 4: Results of Northern hybridizations. Northern blot analysis of the expression of the candidate regulatory genes in the recombinant strains. **(A)** mRNA signals of genes 123668, 80291, 74765, 122523, 66966 and 64608 in cultures of the strains harboring the corresponding overexpression cassettes pMH18, pMH20, pMH25, pMH29, pMH35 and pMH36, respectively, are shown on the top. The mobility of the transcript encoded by the overexpression construct is indicated by an arrow in the blot. Samples collected after 3 days of cultivation (two biological replicates) were analyzed. The northern hybridization signal of actin and staining of total RNA with the SYBR Green II in the same gels are shown below each of the northern blots, as indicated. **(B)** mRNA signals of gene 77513 in cultures of overexpression strains pMH15 and pMH15(S), and in the Del77513 strain. Samples collected after 3 and 5 days of cultivation (two biological replicates) were analyzed. The northern hybridization signal of actin and staining of total RNA with the SYBR Green II in the same gel are shown below, as indicated. **(C)** Signal fold change of the northern signals in the recombinant strain versus the control strain. Signal intensities were normalized using the actin signal.

Additional file 5: pMS204 vector with hygromycin resistance gene and gateway cloning cassette under gpdA promoter and trpC terminator. AmpR, ampicillin resistance gene; attR1/attR2, att sites for recombination; ccdB, ccdB gene for negative selection; CmR, chloramphenicol resistance gene; hph, hygromycin resistance gene; MCS, multiple cloning site; ORI, origin of replication.

Abbreviations
CAZy: carbohydrate active enzyme; CBHI: cellobiohydrolase 1; EGI: endoglucanase 1; MUL: 4-methylumbelliferyl-β-D-lactoside; ORF: open reading frame; PCR: polymerase chain reaction.

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

Authors' contributions
MH carried out cloning of the genes, participated in the construction and cultivation of the recombinant strains, enzymatic activity measurements and quantitative PCR analysis, carried out fungal cultivations and microarray detection of the expression signals for the second cultivation set, and drafted the manuscript. MJV carried out the Southern and northern hybridizations. AWP participated in the cloning of the genes, construction and cultivation of the recombinant strains and enzymatic activity measurements. NA carried out quantitative PCR analysis of transcript levels. MA analyzed the transcriptome data from the chemostat cultivations. MV participated in the data analysis and constructed the deletion strain. MP and MS conceived of the study, and participated in its design and coordination. TMP participated in the design and coordination of the study, carried out microarray data analysis including the selection of the candidate genes, and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements
Aili Grundström, Riitta Nurmi, Hanna Kuusinen, Seija Nordberg, Riitta Lampinen and Tuuli Teikari are acknowledged for extremely skillful technical assistance. The work was funded by the VTT Biorefinery theme, the European Commission (the 6th Framework Programme, contract No. 019882), the Finnish Funding Agency for Technology and Innovation, Tekes (decision 40282/08), and Academy of Finland (decision no. 133455).

Received: 5 July 2013 Accepted: 14 January 2014
Published: 28 January 2014

References

1. Schuster A, Schmoll M: **Biology and biotechnology of *Trichoderma*.** *Appl Microbiol Biotechnol* 2010, **87**(3):787–799.
2. Penttilä M, Limón C, Nevalainen H: **Molecular biology of *Trichoderma* and biotechnological applications.** In *Mycology, Handbook of Fungal Biotechnology. Volume 20.* 2nd edition. Edited by Arora DK. New York: Marcel Dekker; 2004:413–427.
3. Saloheimo M, Pakula TM: **The cargo and the transport system: secreted proteins and protein secretion in *Trichoderma reesei* (*Hypocrea jecorina*).** *Microbiology* 2012, **158**(1):46–57.
4. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics.** *Nucleic Acids Res* 2009, **37**(suppl 1):D233–D238.
5. **Carbohydrate Active Enzymes database.** [http://www.cazy.org/]
6. Foreman PK, Brown D, Dankmeyer L, Dean R, Diener S, Dunn-Coleman NS, Goedegebuur F, Houfek TD, England GJ, Kelley AS, Meerman HJ, Mitchell T, Mitchinson C, Olivares HA, Teunissen PJM, Yao J, Ward M: **Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*.** *J Biol Chem* 2003, **278**(34):31988–31997.
7. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EGJ, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, *et al*: **Genome**

sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotech* 2008, **26**(5):553–560.
8. Häkkinen M, Arvas M, Oja M, Aro N, Penttilä M, Saloheimo M, Pakula T: **Re-annotation of the CAZy genes of *Trichoderma reesei* and transcription in the presence of lignocellulosic substrates.** *Microb Cell Fact* 2012, **11**(1):134.
9. Aro N, Pakula T, Penttilä M: **Transcriptional regulation of plant cell wall degradation by filamentous fungi.** *FEMS Microbiol Rev* 2005, **29**(4):719–739.
10. Kubicek C, Mikus M, Schuster A, Schmoll M, Seiboth B: **Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*.** *Biotechnol Biofuels* 2009, **2**(1):19.
11. Adav SS, Ravindran A, Chao LT, Tan L, Singh S, Sze SK: **Proteomic analysis of pH and strains dependent protein secretion of *Trichoderma reesei*.** *J Proteome Res* 2011, **10**(10):4579–4596.
12. Schmoll M, Esquivel-Naranjo EU, Herrera-Estrella A: ***Trichoderma* in the light of day – physiology and development.** *Fungal Genet Biol* 2010, **47**(11):909–916.
13. Pakula TM, Salonen K, Uusitalo J, Penttilä M: **The effect of specific growth rate on protein synthesis and secretion in the filamentous fungus *Trichoderma reesei*.** *Microbiology* 2005, **151**(1):135–143.
14. Arvas M, Pakula T, Smit B, Rautio J, Koivistoinen H, Jouhten P, Lindfors E, Wiebe M, Penttilä M, Saloheimo M: **Correlation of gene expression and protein production rate - a system wide study.** *BMC Genomics* 2011, **12:**616.
15. Abrahao Neto J, Rossini CHB, El-Gogary S, Henrique-Silva F, Crivellaro O, El-Dorry H: **Mitochondrial functions mediate cellulase gene expression in *Trichoderma reesei*.** *Biochemistry* 1995, **34**(33):10456–10462.
16. Pakula TM, Laxell M, Huuskonen A, Uusitalo J, Saloheimo M, Penttilä M: **The effects of drugs inhibiting protein secretion in the filamentous fungus *Trichoderma reesei*.** *J Biol Chem* 2003, **278**(45):45011–45020.
17. Ilmén M, Thrane C, Penttilä M: **The glucose repressor gene *cre1* of *Trichoderma*: isolation and expression of a full-length and a truncated mutant form.** *Mol Gen Genet MGG* 1996, **251**(4):451–460.
18. Stricker AR, Grosstessner-Hain K, Wurleitner E, Mach RL: **Xyr1 (Xylanase Regulator 1) regulates both the hydrolytic enzyme system and D-Xylose metabolism in *Hypocrea jecorina*.** *Eukaryot Cell* 2006, **5**(12):2128–2137.
19. Aro N, Saloheimo A, Ilmén M, Penttilä M: **ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of *Trichoderma reesei*.** *J Biol Chem* 2001, **276**(26):24309–24314.
20. Zeilinger SZ, Ebner AE, Marosits TM, Mach RM, Kubicek CK: **The *Hypocrea jecorina* HAP 2/3/5 protein complex binds to the inverted CCAAT-box (ATTGG) within the *cbh2* (cellobiohydrolase II-gene) activating element.** *Mol Genet Genomics* 2001, **266**(1):56–63.
21. Saloheimo A, Aro N, Ilmén M, Penttilä M: **Isolation of the *ace1* gene encoding a Cys2-His2 transcription factor involved in regulation of activity of the cellulase promoter *cbh1* of *Trichoderma reesei*.** *J Biol Chem* 2000, **275**(8):5817–5825.
22. Aro N, Ilmén M, Saloheimo A, Penttilä M: **ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression.** *Appl Environ Microbiol* 2003, **69**(1):56–65.
23. Colabardini AC, Humanes AC, Gouvea PF, Savoldi M, Goldman MHS, Kress MRZ, Bayram Ö, Oliveira JVC, Gomes MD, Braus GH, Goldman GH: **Molecular characterization of the *Aspergillus nidulans fbxA* encoding an F-box protein involved in xylanase induction.** *Fungal Genet Biol* 2012, **49**(2):130–140.
24. Jonkers W, Rep M: **Mutation of CRE1 in *Fusarium oxysporum* reverts the pathogenicity defects of the FRP1 deletion mutant.** *Mol Microbiol* 2009, **74**(5):1100–1113.
25. Coradetti ST, Craig JP, Xiong Y, Shock T, Tian C, Glass NL: **Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi.** *Proc Natl Acad Sci* 2012, **109**(19):7397–7402.
26. Nitta M, Furukawa T, Shida Y, Mori K, Kuhara S, Morikawa Y, Ogasawara W: **A new Zn(II)2Cys6-type transcription factor BglR regulates β-glucosidase expression in *Trichoderma reesei*.** *Fungal Genet Biol* 2012, **49**(5):388–397.
27. Seiboth B, Karimi RA, Phatale PA, Linke R, Hartl L, Sauer DG, Smith KM, Baker SE, Freitag M, Kubicek CP: **The putative protein methyltransferase LAE1 controls cellulase gene expression in *Trichoderma reesei*.** *Mol Microbiol* 2012, **84**(6):1150–1164.
28. **Bioconductor, open source software for bioinformatics.** [http://www.bioconductor.org/]
29. Kumar L, Futschik ME: **Mfuzz: a software package for soft clustering of microarray data.** *Bioinformation* 2007, **2**(1):5–6,7.
30. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duquenne L, Finn RD, Fraser M, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, *et al*: **InterPro in 2011: new developments in the family and domain prediction database.** *Nucleic Acids Res* 2012, **40**(D1):D306–D312.
31. Todd RB, Lockington RA, Kelly JM: **The *Aspergillus nidulans creC* gene involved in carbon catabolite repression encodes a WD40 repeat protein.** *Mol Gen Genet* 2000, **263**(4):561–570.
32. Smyth GK, Michaud J, Scott HS: **Use of within-array replicate spots for assessing differential expression in microarray experiments.** *Bioinformatics* 2005, **21**(9):2067–2075.
33. Nakazawa H, Okada K, Kobayashi R, Kubota T, Onodera T, Ochiai N, Omata N, Ogasawara W, Okada H, Morikawa Y: **Characterization of the catalytic domains of *Trichoderma reesei* endoglucanase I, II, and III, expressed in *Escherichia coli*.** *Appl Microbiol Biotechnol* 2008, **81**(4):681–689.
34. Kulmburg P, Mathieu M, Dowzer C, Kelly J, Felenbok B: **Specific binding sites in the *alcR* and *alcA* promoters of the ethanol regulon for the CREA repressor mediating carbon cataboiite repression in *Aspergillus nidulans*.** *Mol Microbiol* 1993, **7**(6):847–857.
35. Tamayo EN, Villanueva A, Hasper AA, de Graaff LH, Ramón D, Orejas M: **CreA mediates repression of the regulatory gene *xlnR* which controls the production of xylanolytic enzymes in *Aspergillus nidulans*.** *Fungal Genet Biol* 2008, **45**(6):984–99327.
36. Mach-Aigner AR, Pucher ME, Steiger MG, Bauer GE, Preis SJ, Mach RL: **Transcriptional regulation of *xyr1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*.** *Appl Environ Microbiol* 2008, **74:**6554–6562.
37. Brakhage AA: **Regulation of fungal secondary metabolism.** *Nat Rev Micro* 2013, **11**(1):21–32.
38. Karimi-Aghcheh R, Bok JW, Phatale PA, Smith KM, Baker SE, Lichius A, Omann M, Zeilinger S, Seiboth B, Rhee C, Keller NP, Freitag M, Kubicek CP: **Functional analyses of *Trichoderma reesei* LAE1 reveal conserved and contrasting roles of this regulator.** *G3 (Bethesda)* 2013, **3**(2):369–378.
39. Metz B, Seidl-Seiboth V, Haarmann T, Kopchinskiy A, Lorenz P, Seiboth B, Kubicek CP: **Expression of biomass-degrading enzymes is a major event during conidium development in *Trichoderma reesei*.** *Eukaryot Cell* 2011, **10**(11):1527–1535.
40. Arvas M, Kivioja T, Mitchell A, Saloheimo M, Ussery D, Penttilä M, Oliver S: **Comparison of protein coding gene contents of the fungal phyla Pezizomycotina and Saccharomycotina.** *BMC Genomics* 2007, **8**:325.
41. Porciuncula Jde O, Furukawa T, Shida Y, Mori K, Kuhara S, Morikawa Y, Ogasawara W: **Identification of major facilitator transporters involved in cellulase production during lactose culture of *Trichoderma reesei* PC-3-7.** *Biosci Biotechnol Biochem* 2013, **77**(5):1014–1022.
42. Ivanova C, Bååth JA, Seiboth B, Kubicek CP: **Systems analysis of lactose metabolism in *Trichoderma reesei* identifies a lactose permease that is essential for cellulase induction.** *PLoS One* 2013, **8**(5):e62631.
43. Ensembl database: [http://fungi.ensembl.org/index.html]
44. Stricker A, Mach R, de Graaff L: **Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina (Trichoderma reesei)*.** *Appl Microbiol Biotechnol* 2008, **78**(2):211–220.
45. ***Trichoderma reesei* database v2.0.** [http://genome.jgi-psf.org/Trire2/Trire2.home.html]
46. ***Trichoderma reesei* database v1.0.** [http://genome.jgi-psf.org/trire1/trire1.home.html]
47. Punt PJ, Zegers ND, Busscher M, Pouwels PH, van den Hondel CA: **Intracellular and extracellular production of proteins in Aspergillus under the control of expression signals of the highly expressed *Aspergillus nidulans gpdA* gene.** *J Biotechnol* 1991, **17**(1):19–33.
48. Mullaney E, Hamer J, Roberti K, Yelton MM, Timberlake W: **Primary structure of the *trpC* gene from *Aspergillus nidulans*.** *Mol Gen Genet* 1985, **199**(1):37–45.
49. Penttilä M, Nevalainen H, Rättö M, Salminen E, Knowles J: **A versatile transformation system for the cellulolytic filamentous fungus *Trichoderma reesei*.** *Gene* 1987, **61**(2):155–164.
50. Sambrook J, Fritsch E, Maniatis T: *Molecular Cloning: A Laboratory Manual.* 2nd edition. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory; 1989.
51. Engler C, Kandzia R, Marillonnet S: **A one pot, one step, precision cloning method with high throughput capability.** *PLoS One* 2008, **3**(11):e3647.

52. Stewart PE, Thalken R, Bono JL, Rosa P: **Isolation of a circular plasmid region sufficient for autonomous replication and transformation of infectious *Borrelia burgdorferi*.** *Mol Microbiol* 2001, **39**(3):714–721.
53. Bailey MJ, Tähtiharju J: **Efficient cellulase production by *Trichoderma reesei* in continuous cultivation on lactose medium with a computer-controlled feeding strategy.** *Appl Microbiol Biotechnol* 2003, **62**(2):156–162.
54. Bailey MJ, Biely P, Poutanen K: **Interlaboratory testing of methods for assay of xylanase activity.** *J Biotechnol* 1992, **23**(3):257–270.

doi:10.1186/1754-6834-7-14
Cite this article as: Häkkinen *et al.*: Screening of candidate regulators for cellulase and hemicellulase production in *Trichoderma reesei* and identification of a factor essential for cellulase production. *Biotechnology for Biofuels* 2014 **7**:14.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

PUBLICATION III

The effects of extracellular pH and of the transcriptional regulator PACI on the transcriptome of *Trichoderma reesei*

Submitted to: Microbial Cell Factories.
Copyright 2014 Authors.

The effects of extracellular pH and of the transcriptional regulator PACI on the transcriptome of *Trichoderma reesei*

Mari Häkkinen, Dhinakaran Sivasiddarthan, Nina Aro, Markku Saloheimo and Tiina M. Pakula[§]

VTT Technical Research Centre of Finland, P.O.Box 1000 (Tietotie 2, Espoo), FI-02044 VTT, Finland

[§]Corresponding author

Email addresses:

MH: mari.hakkinen@vtt.fi

DS: dhinakaran.sivasiddarthan@gmail.com

NA: nina.aro@vtt.fi

MS: markku.saloheimo@vtt.fi

TMP: tiina.pakula@vtt.fi

Abstract

Background

Extracellular pH is one of the several environmental factors affecting protein production of filamentous fungi. Regulation methods have developed to ensure that for example extracellular enzymes are produced under pH-conditions suitable for the activity of the enzymes. In filamentous fungi, ambient pH-dependent regulation of extracellular protein production is governed by the transcriptional regulator PacC. The effects of pH and of PacC on protein production have been studied especially in *Aspergilli*, whereas the effects of pH in the industrial producer of hydrolytic enzymes, *Trichoderma reesei*, have mainly been studied at the protein level. In this study, the pH-dependent expression of *T. reesei* genes was investigated by genome-wide transcriptional profiling and by constructing a deletion strain from an orthologue of the *Aspergillus nidulans pacC* gene (designated as *pac1*). In addition, the effect of *pac1* modulation on enzyme production was studied.

Results

Transcriptional analysis revealed the pH-responsive genes of *T. reesei* and classification of these genes identified the activities most affected by changing pH. A large number of especially transporter genes, signalling-related genes, protease genes, glycoside hydrolase genes and genes involved in different metabolism-related functions were found to be pH-responsive. The glycoside hydrolases included several pH-responsive cellulase and hemicellulase genes. However, PACI-mediated regulation was not common for these genes, indicating the presence of additional regulatory mechanisms affecting the pH-dependent expression of cellulase and hemicellulase genes. Especially hemicellulase genes encoding the same activity were shown to include both genes up-regulated at low pH and genes preferring higher pH. Deletion of the *pac1* gene was shown to have a

negative effect on the production of cellulase and xylanase activity at pH6. New information was gained on the effects of pH on the genes involved in ambient pH-signalling and on the known and candidate regulatory genes involved in regulation of cellulase and hemicellulase genes. In addition, co-regulated genomic clusters responding to change of ambient pH were identified.

Conclusions

Ambient pH was shown to be an important determinant of *T. reesei* gene expression. The pH-responsive genes, including those affected by the regulator of ambient pH sensing, were identified and novel information was gained from the pH-responsiveness of the carbohydrate active enzyme genes.

Keywords: *Trichoderma reesei*, ambient pH, pac1, carbohydrate active enzymes, cellulase, hemicellulase, transcriptional profiling

Background

Filamentous fungi can adjust to changing ambient pH of their habitat by an intracellular pH homeostatic system and by controlling the synthesis of gene products directly exposed to the surrounding environment (for example cell surface proteins, secreted proteins). Ambient pH regulation of filamentous fungi has been most extensively studied in *Aspergillus nidulans* and was shown to be mediated by the wide-domain zinc finger transcription factor PacC together with a signal transduction cascade composed of the products of six pal genes (*palA*, *palB*, *palC*, *palF*, *palH* and *palI*) (reviewed by [1]).

64

65 Ambient pH is believed to be signalled through a plasma membrane complex [2–4]. The
66 components of the pal pathway and of an ESCRT complex are subsequently recruited to the plasma
67 membrane-associated structure. The proposed model for pH signalling can be found from [5]. There
68 are three different forms of PacC. In acidic conditions the full-length form predominates. At
69 alkaline pH, activation of the pal pathway leads to two subsequent cleavage steps resulting in two
70 shorter forms of PacC [6]. The product of the PalB-mediated cleavage step, PacC53, is further
71 cleaved by the proteasome to create the active form of PacC (PacC27) [7]. The active PacC binds
72 the core target sequence 5´-GCCARG-3´ in the promoters of its target genes[8].

73

74 PacC of *A. nidulans* acts as a repressor of acid-expressed genes and as an activator of alkaline-
75 expressed genes [9]. In accordance, null mutations in *pacC* or in the pal genes [10, 11] lead to a
76 phenotype mimicking that displayed in acidic conditions, whereas a constitutively active PacC
77 results in activation of alkaline-expressed genes and repression of acid-expressed genes regardless
78 of pH [9, 10]. Genes controlled by PacC include e.g. acid and alkaline phosphatases [10], permeases
79 [12] and genes encoding extracellular enzymes such as the *A. nidulans* xylanase genes (*xlnA* and
80 *xlnB*) [13] and an α-L-arabinofuranosidase gene (*abfB*) [14]. PacC has also been shown to regulate
81 the expression of the *pacC* gene itself, resulting in an abundance of *pacC* mRNA at alkaline pH [9].
82 In a recent study, a negative feedback loop attenuating the first cleavage step of PacC during longer
83 exposure to alkaline conditions was proposed to exist [15].

84

85 *Trichoderma reesei* is an industrial producer of especially cellulases and hemicellulases for various
86 applications and is also a widely used host for the production of heterologous proteins [16–18]. In
87 order to create better production strains with enhanced enzyme production properties, the

88 physiological and environmental factors and regulatory mechanisms affecting enzyme production
89 by *T. reesei* have been widely studied. Several transcription factors regulating the expression of
90 cellulase and hemicellulase genes have been identified and characterized in detail (for reviews see
91 [18, 19]). The availability of a complete genome sequence of *T. reesei* has made it possible to use
92 genome-wide methods to study the various factors affecting protein production [20]. However, the
93 effect of ambient pH on the production of cellulases and hemicellulases has received less attention,
94 although previous studies indicate that high pH favours xylanase and low pH cellulase production
95 [21]. Recently, endoglucanase production by *T. reesei* was suggested to be highest at pH4.5,
96 whereas exoglucanase and β-glucanase production reached their highest values at pH5 and 5.5,
97 respectively [22]. However, another recent study suggests that there are differences in the optimal
98 pH for cellulase and hemicellulase secretion between different *T. reesei* strains [23].

99

100 In this study, effects of the changing ambient pH on the transcriptome of *Trichoderma reesei* were
101 studied by transcriptional profiling. The pH-responsive genes were screened from the genome and
102 classified in order to identify the major gene groups affected by the extracellular pH. In addition,
103 the effects of the regulator PACI were studied by constructing a deletion strain. The effects of the
104 changing pH conditions on the members of specific gene groups were investigated in more detail.
105 These genes included the carbohydrate active enzyme genes, genes encoding the components of the
106 pH signal transduction pathway and genes encoding known and candidate regulators of cellulase
107 and hemicellulase genes. In addition, cellulase and xylanase activities produced by the parental
108 strain and the recombinant strains were determined and compared to the results of the
109 transcriptional analysis. The transcriptomics data also enabled identification of co-regulated
110 genomic clusters containing especially secondary metabolism genes.

111

Results

pH responsive genes of *T. reesei* according to microarray analysis

A recombinant strain from which the *pac1* open reading frame had been replaced by a hygromycin resistance cassette was constructed. This *pac1* deletion strain (designated as Δpac1) was cultivated in a bioreactor on a medium containing Avicel cellulose at pH6 in parallel with the parental strain QM9414. In addition, the parental strain QM9414 was cultivated in a bioreactor on the same medium in different pH conditions (pH3, pH4.5 and pH6). Three biological replicates of the strain QM9414 and Δpac1 were cultivated in each case. Samples collected from two different time points of cultivation were subjected to transcriptional analysis using the microarray method.

Data from the three biological replicates of the strains QM9414 and Δpac1 of the two cultivations was statistically analysed using Limma analysis (R, bioconductor [24, 25]) with a fold change log2 >0.4 and a p-value <0.01 as a threshold. The statistical analysis revealed that in the strain QM9414 the expression of approximately 940 genes responded significantly to the change of pH in pair-wise comparisons between pH6 and pH3 and/or pH6 and pH4.5 and/or pH4.5 and pH3 (Figure 1A, Additional file 1). Of these genes, 354 were high pH up-regulated genes (expression increased significantly in a comparison between a higher and a lower pH) and 586 low pH up-regulated genes (expression decreased significantly in a comparison between a higher and a lower pH) (Figure 1A). These numbers include two genes that had a positive response to the change of pH from 3 to 4.5 but a negative response to the change of pH from 4.5 to 6.

In total 189 genes were differentially expressed in the cultures of the Δpac1 strain as compared to the parallel parental strain cultures at pH6 (Figure 1A, Additional file 2). Of these genes 77 were

135 up-regulated and 112 down-regulated in the deletion strain as compared to the parental strain. The
136 microarray platform used in this study includes probes for the transcripts of approximately 10 000
137 genes, including genes from the genome versions 2.0 and 1.2 [26, 27], and for putative novel genes
138 detected in a previous study [28]. Therefore, ~ 9 % of the predicted transcripts of the microarray
139 analysis were regulated by pH and ~ 2 % were affected by the PACI transcription factor directly or
140 indirectly. Among these 189 genes, the ones most likely to be under direct PACI regulation were
141 searched for according to the Limma analysis. The PacC transcription factor of *A. nidulans* is
142 activated at alkaline pH. The active regulator represses the transcription of acid-expressed genes
143 and activates the transcription of alkaline-expressed genes [9]. Deletion of the *pacC* gene leads to a
144 phenotype mimicking acidic conditions. Therefore, the hypothesis was that the *T. reesei* genes
145 directly repressed by PACI are only active at acidic pH and thus are more highly expressed in the
146 parental strain at pH3 as compared to pH6 (low pH up-regulated genes). In the deletion strain, gene
147 expression mimics that of the parental strain in acidic conditions and therefore the expression of the
148 PACI repressed genes is higher as compared to the parental strain when grown at pH6. Similarly,
149 the *T. reesei* genes induced by PACI are active at alkaline pH and thus are more highly expressed at
150 pH6 as compared to pH3 in the parental strain (high pH up-regulated genes). In the deletion strain,
151 expression of the PACI-induced genes is lower as compared to the parental strain grown at pH6.
152 According to these criteria, 30 and 38 genes were found to be repressed and induced by PACI,
153 respectively (Additional file 2). When the expression data was clustered using the Mfuzz clustering
154 method [29], the majority of the genes activated by PACI were assigned to a common Mfuzz-
155 cluster whereas the PACI repressed genes were more evenly divided to several clusters.

156

157 Classification of the pH-responsive genes

158 The pH-responsive genes were divided into different functional classes according to the Eukaryotic
159 orthologous groups (KOG) classification (Figure 1B). The glycoside hydrolase, carbohydrate
160 esterase and polysaccharide lyase genes were classified according to the updated annotations [30].
161 In addition to the glycoside hydrolases, which will be discussed further below, especially genes
162 with functions related to energy production and conversion; posttranslational modification, protein
163 turnover and chaperones; signal transduction mechanisms; carbohydrate, inorganic ion, lipid and
164 amino acid transport and metabolism and secondary metabolite biosynthesis, transport and
165 catabolism were abundant among the pH-responsive genes.

166

167 A substantial proportion of the pH-responsive genes was classified as having a general function
168 prediction only, an unknown function or did not have a KOG classification. For 47 of these genes, a
169 putative function could be assigned according to the manual annotations in the Joint Genome
170 Institute (JGI) *T. reesei* v2.0 databank [26]. These genes included for example the *pac1* gene itself
171 together with several other regulatory genes, transferase genes, hydrophobin genes, signalling-
172 related genes and genes with transport-related functions, among others. For 120 genes without a
173 manual JGI annotation and classified as having a general function prediction only or an unknown
174 function, an approximate annotation was assigned according to the KOG definition. These included
175 for example candidate regulatory genes, candidate transporter genes, candidate peptidase genes and
176 several candidate reductase, oxidoreductase, transferase and dehydrogenase genes. Of the genes
177 without a KOG classification or a JGI annotation, a function prediction could be assigned to 114
178 according to the functional Interpro domains, revealing for example additional peptidase, kinase,
179 transporter and regulatory genes. Classification of all the pH responsive genes including those
180 affected by the *pac1* deletion can be found in Additional files 1 and 2.

181

182 **pH regulation of the components of the pH-signalling pathway**

183 The pH-signalling pathway leading to the activation of PacC consists of the six pal genes (*palA*,
184 *palB*, *palC*, *palF*, *palH* and *palI*) together with the components of the ESCRT-complex (for
185 example vps20, vps23, vps24 and vps32). The *T. reesei* homologues of the *A. nidulans* pal-genes
186 and vps-genes were searched from the genome and the behaviour of these genes in different pH
187 conditions was studied. When the three pH conditions were compared, a Limma analysis using a
188 fold change log2 > 0.4 and a p-value < 0.01 as a threshold did not detect any significant changes in
189 the expression of the six pal gene homologues of *T. reesei*, suggesting a rather steady expression of
190 these genes regardless of pH. In addition, the vps gene homologues of *T. reesei* were not pH-
191 regulated according to the Limma analysis.

192

193 Of the pal-genes, the *palC* homologue had the most similar expression pattern to that of *pac1*
194 (Figure 3). Expression level of the gene increased steadily together with the pH. In addition, the
195 statistically significant down-regulation of the gene in the deletion strain as compared to the
196 parental strain at pH6 suggests that the *palC* homologue is at least partially under positive PACI-
197 mediated regulation (Additional file 2). By contrast, the trend in the expression of the *palF*
198 homologue indicates PACI-mediated repression (Figure 3), which is further supported by the fact
199 that the *palF* homologue clustered together with the majority of the PACI-repressed genes.

200

201 **The effect of pH on CAZy genes and on the regulators of cellulase and**
202 **hemicellulase genes**

203 Due to the fact that *T. reesei* is an exceptionally efficient producer of cellulases and hemicellulases,
204 the effect of pH on the expression of genes encoding carbohydrate active enzymes (CAZy, [31, 32])

205 was studied in more detail. The majority of the enzymes involved in the degradation of plant cell
206 wall material belong to the classes of glycoside hydrolases (GH), carbohydrate esterases (CE) and
207 polysaccharide lyases (PL). Of the pH-responsive genes, 60 are classified as glycoside hydrolases
208 and carbohydrate esterases and one as a polysaccharide lyase (Figure 1B). Of these genes, 23 and
209 38 were determined to be high pH up-regulated and low pH up-regulated genes, respectively. Seven
210 of the glycoside hydrolase genes were down-regulated in the Δpac1 strain and one was up-regulated
211 (Figure 2A). One carbohydrate esterase gene was down-regulated in the Δpac1 strain. One
212 carbohydrate esterase gene and two glycoside hydrolase genes are under PACI-mediated regulation
213 (Figure 2B). The microarray data of the CAZy genes can be found from Additional file 3. A major
214 proportion of the pH-responsive CAZy genes belong to families GH16, GH27, GH55 and GH18
215 (Figure 4A). All of these families contain genes coding for activities that are most probably
216 involved in cell wall rearrangements during growth of the fungus.

217

218 The group of the pH-responsive CAZy genes included several genes encoding enzymes with
219 activities against the cellulose and hemicellulose polymers (Figure 4B). The group of high pH up-
220 regulated genes included GH10, GH11 and GH30 endo-β-1,4-xylanase genes (*xyn3*, *xyn2* and
221 69276), a GH12 endo-β-1,4-glucanase gene *egl3*, a CE15 glucuronoyl esterase gene *cip2*, a
222 candidate GH3 β-glucosidase gene *bgl3i*, two candidate GH61 copper-dependent polysaccharide
223 mono-oxygenase genes (22129 and 31447) and candidate acetyl xylan esterase genes from families
224 CE3 (70021) and CE5 (54219). Similarly, low pH up-regulated genes included two GH11 endo-β-
225 1,4-xylanase genes (*xyn1*and *xyn5*), a CE5 acetyl xylan esterase gene *axe2*, four GH27 α-
226 galactosidase genes (*agl1*, *agl3*, 27259 and 27219), a candidate CE16 acetyl esterase gene
227 (103825), candidate GH5 membrane bound endoglucanase and β-1,3-mannanase/endo-β-1,4-
228 mannosidase genes (*cel5b* and 71554), candidate GH3 β-glucosidase and β-xylosidase genes (*bgl3j*
229 and *xyl3b*), a candidate GH79 β-glucuronidase gene (71394) and two candidate GH95 α-fucosidase

230 genes (72488 and 111138). Genes down-regulated in the Δpac1 strain included two candidate β-
231 glucosidase genes *bgl3i* and *cel3e*, of which *bgl3i* is PACI-induced according to the Limma analysis
232 (Figure 4B, Additional file 2).

233

234 According to the criteria used, in addition to *bgl3i*, GH/CE genes under positive PACI regulation
235 also include a candidate GH76 α-1,6-mannanase gene (122495) and a candidate CE9 N-acetyl-
236 glucosamine-6-phosphate deacetylase gene (79671). These three genes are all assigned to the Mfuzz
237 cluster containing the majority of the PACI-induced genes. Other CAZy genes belonging to this
238 cluster include a xylanase gene *xyn2*, a candidate CE3 esterase gene (44366), a candidate GH92 α-
239 1,2-mannosidase gene (79921), a candidate GH72 β-1,3-glucanosyltransferase gene (77942), a
240 candidate GH16 cell wall glucanosyl transferase gene (65406), a candidate GH24 lysozyme gene
241 (109278), a candidate CE1 esterase gene (72072) and a glycoside hydrolase gene of unknown
242 function (108348). Although these genes are not under PACI regulation according to the statistical
243 test they have similar expression patterns as the PACI-induced genes.

244

245 No GH or CE genes under negative PACI regulation were detected with the used parameters
246 (Figure 2B). The vast majority of the PACI-repressed pH-responsive genes were assigned to three
247 different Mfuzz clusters (Additional file 2.). In total 23 CAZy genes were found from these clusters.
248 These genes included for example *agl1*, two candidate GH27 α-galactosidase genes (27219 and
249 27259), candidate GH3 (*xyl3b*) and GH39 (73102) β-xylosidase genes, a candidate GH95 α-L-
250 fucosidase gene (72488), *bga1*, a candidate GH2 β-galactosidase/β-glucuronidase gene (76852), a
251 candidate CE3 acetyl xylan esterase gene (41248), a candidate GH79 β-glucuronidase gene (71394)
252 and a candidate CE16 acetyl esterase gene (103825) (Additional file 3).

253

254 The expression patterns of the genes encoding for the candidate or characterized (*xyr1*, *ace1*, *ace2*
255 and *cre1*) regulators of cellulase and hemicellulase genes were also studied and compared to those
256 of the cellulase and hemicellulase genes. The characterized regulators were not pH-regulated
257 according to the Limma test but some of the novel candidate regulatory genes identified during a
258 previous study [33] (genes 111742 (pMH14), 120120 (pMH22) and 123019 (pMH30)) were low
259 pH up-regulated genes (Additional file 1). A heatmap representation from the fold changes of the
260 genes for known and candidate cellulolytic and hemicellulolytic enzymes and of the characterized
261 or candidate regulatory genes for these enzyme genes is shown in Figure 5. Genes encoding the
262 components of the pH-signalling pathway were also included. Genes forming the heat map are
263 listed in Additional file 4. The first two branches of the heat map contain the high pH up-regulated
264 genes identified with Limma analysis. Among these genes are several other genes the expression
265 profiles of which indicate expression preferably at high pH (*cel74a*, *cel61b*, *abf2*, *abf3*, *cel3d* and
266 *axe1*). Most of the low pH up-regulated genes are divided into two branches at the bottom of the
267 heat map. Genes such as a candidate GH27 α-galactosidase gene (59391), a candidate GH39 β-
268 xylosidase gene (73102) and *abf1* group together with the low pH up-regulated genes. A small
269 branch of genes preferably expressed at low pH according to the fold changes is found from the top
270 of the heat map. These include *bgl3f*, *agl2* and a candidate GH5 endoglucanase gene (53731). This
271 branch also contains a candidate acetyl esterase gene (103825) shown to be a low pH up-regulated
272 gene according to the Limma analysis. The candidate regulatory genes 111742 (pMH14), 120120
273 (pMH22), 123019 (pMH30) and 122523 (pMH29) [33] group together with the low pH up-
274 regulated genes. Interestingly, the *ace3* gene encoding a novel regulator of especially cellulase
275 genes [33] is located in the same branch as *cel3c*, a candidate β-xylosidase/α-L-arabinofuranosidase
276 gene (3739), *egl2*, *cbh2*, *swo1*, *egl4* and two candidate regulatory genes (121121 and 56077),
277 indicating similar expression of these genes under the prevailing conditions.

278

279 As a conclusion, especially the xylanase, β-glucosidase, acetyl xylan esterase and α-L-
280 arabinofuranosidase genes (according to the heat map) of *T. reesei* include both high pH up-
281 regulated and low pH up-regulated genes. However, the α-galactosidase genes of *T. reesei* were
282 preferably expressed in acidic conditions. Furthermore, the pH-responsive GH61 family members
283 were all high pH up-regulated genes.

284

285 **Cellulase and xylanase activities produced in different pH conditions**

286 For the transcriptional analysis, the parental strain QM9414 and the *pac1*deletion strain were
287 cultivated in bioreactors in a medium containing Avicel cellulose at pH6. Supernatant samples were
288 collected from different time points of the bioreactor cultivations, enabling the study of enzyme
289 production in the different pH conditions and between the two different strains. Xylanase activity,
290 cellulase activity against the 4-methyl umbelliferyl-β-D-lactoside (MUL) substrate and total protein
291 production were measured from the supernatant samples (Figures 6-7). In addition, specific
292 activities of the cellobiohydrolase 1 (CBHI) and endoglucanase 1 (EGI) enzymes were determined
293 (Figure 6). After the third time point of cultivation (41 h), the cellulase activity produced by the
294 Δpac1 strain was lower compared to the parental strain (Figure 6A). At the final time point (139.5
295 h) parental strain had reached total production of MUL activity of almost12 nkat/ml, whereas the
296 deletion strain produced less than 7 nkat/ml. A similar trend was also seen in the production of EGI
297 and CBHI activities (Figures 6B and C). The effect of *pac1* deletion on the production of xylanase
298 activity was even more pronounced (Figure 7A). At the final time point of cultivation, the parental
299 strain produced approximately 727 nkat/ml of xylanase activity whereas the deletion strain
300 produced only ~ 271 nkat/ml. The production of total proteins followed the same trend as the
301 cellulase activities (Figure 7B). Therefore, it can be concluded that in the strain QM9414, deletion

302 of the *pac1* gene results in decreased production of cellulase and xylanase activity as well as of total
303 secreted proteins at pH6.

304

305 In addition, QM9414 was cultivated in bioreactors at pH3, 4.5 and 6. Especially the xylanase
306 activity produced was highest at pH6 and lowest at pH3 (Figure 8B). The difference in total
307 cellulase activity produced between the different pH values was less pronounced (Figure 8A). On
308 the basis of production of CO_2, growth of the fungus was slowest at pH3 and therefore especially
309 MUL activity started to increase at a later time during the cultivation at pH3 (Figure 8D).

310

311 Co-regulated genomic regions

312 Mfuzz clustering of the expression profiles was performed in order to study the expression patterns
313 of the genes in more detail and to identify co-regulated groups of genes. Co-regulated genes were
314 also searched for in a genomic scale, identifying genes located close to each other in the scaffolds
315 and being co-regulated. Co-expressed genomic regions were found by searching the genome for
316 neighbouring genes that were assigned to the same Mfuzz cluster (at least 3 genes from 4
317 consecutive genes belonging to the same Mfuzz cluster) and of which at least one was pH-
318 responsive according to the Limma analysis (in the strain QM9414 and/or in the deletion strain). In
319 total 40 such clusters were found. The pH-responsive members of the clusters are marked in the
320 Additional files 1 and 2.

321

322 Genomic clusters of co-regulated pH-responsive genes most probably involved in siderophore
323 biosynthesis were identified. There is a tightly co-regulated genomic region in scaffold 46
324 containing six high pH up-regulated genes. The genes are not under PACI regulation according to

325 the criteria used. According to manually curated annotations, these genes putatively encode a non-
326 ribosomal peptide synthase, a siderophore biosynthesis lipase/esterase, an ABC transporter, a
327 siderophore biosynthesis acetylase, an enoyl-CoA hydratase/isomerase and a siderophore iron
328 transporter. A very similar co-regulated cluster of six high pH up-regulated genes located in
329 scaffold 20 includes genes annotated as an ABC transporter, an oxidoreductase, an MFS transporter,
330 a siderophore biosynthesis acetylase, a long chain fatty acid acyl-CoA ligase and a non-ribosomal
331 peptide synthase. Genes 67026 and 67189 belonging to this cluster are not pH-responsive according
332 to the Limma analysis, although the *T. virens* homologues for these genes have been shown to be
333 up-regulated at alkaline pH [34].

334

335 An additional gene cluster most probably involved in siderophore biosynthesis but not passing the
336 criteria set for the co-regulated genomic regions is located in scaffold 31. The cluster includes a
337 gene homologous with the *T. virens* siderophore biosynthesis gene *sidA* which is a high pH gene, as
338 is the *T. virens* counterpart. Similarly, the *T. virens* non-ribosomal peptide synthase gene (*tex10*)
339 homologue is clustered with the siderophore biosynthesis gene but is not co-regulated with it (and
340 therefore is not assigned to the same Mfuzz-cluster).

341

342 Three clusters possibly involved in cellulase signalling were found. In scaffold 17, a cluster of three
343 high pH up-regulated genes included the *ooc1* gene encoding a secreted protein suggested to be
344 involved in cellulase signal transduction [35] together with a gene homologous to *ooc1* and a gene
345 of unknown function. *ooc1* encodes a small acidic protein the expression of which is abundant on
346 cellulose in strain QM9414 [35]. A cluster of three genes from scaffold 13 includes a major
347 facilitator superfamily transporter gene 79202 possibly involved in cellulase signal transduction
348 [36]. The two other genes belonging to the cluster encode proteins of unknown function. Only one

349 of the unknown genes is pH-responsive (high pH up-regulated gene); expression of the other genes
350 is not significantly affected by pH. These three genes belong to the same Mfuzz cluster as the
351 majority of the genes induced by PACI. Another cluster of low pH up-regulated genes in scaffold
352 29 includes a gene encoding GRDI that is a putative dehydrogenase associated with cellulase signal
353 transduction [37]. The other members of the cluster include four genes with unknown functions and
354 a fungal transcriptional regulatory protein (pMH14 [33]).

355

356 Discussion

357 The transcriptome of *T. reesei* at different extracellular pH values was studied in order to reveal the
358 role of pH regulation in a saprotrophic cellulolytic fungus. Ambient pH was found to be an
359 important determinant of *T. reesei* gene expression. Gene groups abundant among the pH-
360 responsive genes included for example genes encoding for different transporters, exported enzymes
361 such as proteases, signalling-related genes, glycoside hydrolase genes, regulatory genes, and genes
362 involved in various metabolic reactions (oxidases, reductases, dehydrogenases, transferases etc.)
363 and in secondary metabolism. Most of the genes encoding activities against plant cell wall material
364 belong to the glycoside hydrolase family. Therefore, the abundance of glycoside hydrolase genes
365 among the pH-responsive genes might indicate that adaptation to changing ambient pH conditions
366 is an important determinant of the success of a saprotrophic fungus. However, it is likely that the
367 group of pH-responsive genes also includes many that are not directly regulated by pH but are
368 affected by other factors such as growth or stress reactions.

369

370 Of the *T. reesei* genes repressed by PACI, the aspartic protease encoded by gene 74156 has been
371 suggested to be the major protease functioning at and below pH5 [38]. Interestingly, an ammonium

372 transporter gene 43671 was found to be a low pH up-regulated gene repressed by PACI. Therefore,
373 this transporter might be involved in increasing the ambient pH in low pH environments. Especially
374 plant pathogenic fungi are known to secrete ammonia to increase the pH of the environment in
375 order to facilitate infection [39]. The transport of ammonia could also be useful under nutrient-
376 limiting conditions when the fungus needs to modify the ambient pH to facilitate the performance
377 of the extracellular metabolites and enzymes [40]. γ-Amino-*n*-butyrate (GABA) transporters have
378 been shown to be preferentially expressed at acidic pH [41, 42]. Accordingly, three candidate *T.*
379 *reesei* GABA transporters were identified among the pH-responsive genes. These genes were all
380 low pH up-regulated, and one gene was under PACI-mediated repression.

381

382 Of the genes induced by PACI, the candidate potassium/sodium efflux P-type ATPase gene 122972
383 is a homologue of *T. virens* [34] and *Fusarium oxysporum* [43] genes shown to be under Pac1-
384 dependent expression. The *F. oxysporum ena1* gene was shown to be involved in maintaining ion
385 homeostasis at alkaline pH. Interestingly, the *T. reesei* genome encodes three cation-transporting
386 ATPase genes that have homology with the *T. virens* gene and with each other. Genes 122972 and
387 81536 are high pH up-regulated genes and are induced by PACI, whereas the expression of gene
388 81430 is down-regulated in the deletion strain at pH6 but induction at higher pH is not statistically
389 significant. Therefore, it is possible that these three cation-transporting ATPase genes of *T. reesei*
390 are also involved in maintaining intracellular homeostasis.

391

392 In the fungus *Magnaporthe oryzae*, the only pal gene influenced by pH and by the deletion of the
393 *pacC* paralogue was MOPALF [44]. The expression of the MOPALF gene decreased significantly
394 with increasing pH from 5 to 8. When MOPACC was deleted, the expression of MOPALF
395 increased slightly at pH5 and did not decrease at pH8, indicating MOPACC-mediated repression in

396 the parental strain. In *T. reesei*, the effect of pH and of the *pac1* deletion on the *palF* homologue
397 was not very significant, but the expression patterns of the parental and deletion strains indicate that
398 at least partial PACI-mediated regulation might take place. The PacC-mediated repression of palF
399 has been suggested to be involved in cell desensitization [44]. PalF is the first component after the
400 plasma-membrane attached complex to receive the pH signal [5]. However, the PalC homologue of
401 *T. reesei* that is recruited to the complex after PalF was suggested to be at least partially induced by
402 PACI, indicating that the differential regulation of these two genes might be involved in modulating
403 the response to ambient pH. In previous studies, it was concluded that expression of the genes of the
404 pal-signalling pathway is not regulated by ambient pH in the fungi *A. nidulans* and *A. niger* [45,
405 46]. These contradictory results suggest that either these fungi use partially different mechanisms
406 for the regulation of pH signalling or that as in the case of *T. reesei*, the regulation of pal genes is
407 difficult to detect using traditional significance tests.

408

409 The expression patterns of the genes encoding carbohydrate active enzymes and especially those
410 involved in the degradation of cellulose and hemicellulose polymers were studied in more detail.
411 The genome of *T. reesei* encodes four characterized and two candidate xylanases involved in the
412 degradation of the xylan chain of hemicellulose. The expression of *xyn2*, *xyn3* and a candidate
413 GH30 xylanase gene (69276) favoured high pH, whereas *xyn1* and *xyn5* were more highly
414 expressed at lower pH. In addition, PACI was suggested to have a slight inducing effect on *xyn2*
415 according to the Mfuzz clustering. There is also a putative PACI binding site on the promoter of
416 *xyn2*. The *Aspergillus nidulans* xylanase genes *xlnA* and *xlnB* have been shown to have opposite
417 patterns of expression when cultivated in different pH conditions in the presence of D-xylose [13].
418 According to a Northern analysis, *xlnA* was preferably expressed in alkaline (pH 7.5) conditions and
419 *xlnB* in acidic (pH 4.5) conditions. The pH optima for the characterized *T. reesei* xylanases have
420 been determined and they support the results of the transcriptional analysis. XYNIII has a rather

421 high pH optimum (6-6.5), whereas the pH optima of XYNI and XYNII are 4.0-4.5 and 4.0-6.0,
422 respectively [47–50]. In lactose medium, pH4 and pH6 have been shown to favour the production of
423 XYNI and XYNIII, respectively [51]. However, XYNII was produced in both pH conditions. The
424 strain QM9414 produced highest xylanase activity in cellulose-containing medium at pH6, which is
425 in accordance with the results of previous studies [21, 51]. According to the transcriptional profiling
426 data, the high pH up-regulated genes (*xyn2* and *xyn3*) have a higher expression level as compared to
427 the low pH up-regulated genes (*xyn1* and *xyn5*), indicating that the XYNII and XYNIII enzymes are
428 mainly responsible for the high xylanase activity produced at pH6. The effect of *pac1* deletion at
429 pH6 was more severe for xylanase production than for cellulase production. The decrease in
430 xylanase activity observed for the Δpac1 strain could be at least partially the result of *xyn2* down-
431 regulation in the absence of a functional PACI.

432

433 The production of cellulase activity was less affected by pH than the xylanase activity, although a
434 lower amount of activity was produced at pH3 at least partially due to slower growth of the fungus.
435 The results of a previous proteomics study indicated that for strain QM9414 the production of
436 CBHI increases with pH and reaches a maximum at pH6, whereas CBHII abundance was the
437 highest at pH3 [23]. EGI production was also suggested to be highest at pH6 for the strain QM9414.
438 In another study, production of cellulase activity by *T. reesei* has been suggested to be more
439 efficient at low pH [21]. However, this study utilised the strain Rut-C30, for which the optimal pH
440 for production of CBHI was later suggested to be 4 [23]. In accordance, *T. reesei* Rut-C30 has been
441 shown to produce the highest cellulase activity at pH4-4.5 and xylanase activity at pH6 in lactose
442 culture [51]. Therefore, as already shown by [23], the differences between different *T. reesei* mutant
443 strains must be taken into account when designing optimal pH conditions for enzyme production.
444 Results from the bioreactor cultivations indicate that at pH6 the production of cellulase activity is
445 negatively affected by deletion of the *pac1* gene. The transcriptional analysis did not indicate PACI-

446 mediated regulation of the main cellulase genes. Therefore the effect could be mediated via an
447 indirect mechanism. However, putative PACI binding sites are found from the promoters of the
448 main cellulase genes, indicating that some level of regulation might occur although more studies are
449 needed to confirm this.

450

451 Although several CAZy genes respond to change of pH, only a few are clearly under PACI
452 regulation or at least other regulation mechanisms are functioning simultaneously, thus masking the
453 effect of PACI. For example, the promoter of the *Aspergillus tubingensis* xylanase gene *xlnA*
454 contains overlapping binding sites for XlnR and PacC, suggesting that PacC could be competing
455 with the activator [52]. Several other activators and repressors besides XYRI [53] are involved in
456 the regulation of cellulase and hemicellulase genes of *T. reesei*. These include for example ACEI,
457 ACEII and CREI [54–57]. pH-dependent expression was rare among the characterized and
458 candidate regulatory genes of cellulase and hemicellulase genes. Of the novel candidate regulatory
459 genes identified during a previous study [33], genes 111742 (pMH14), 120120 (pMH22) and
460 123019 (pMH30) were low pH up-regulated genes. Over-expression of these genes has been shown
461 to result in slightly declined production of cellulase activity [33]. Interestingly, the novel regulator
462 of especially cellulase genes, *ace3* [33], was assigned to the same Mfuzz cluster and/or heat map
463 branch as several genes that were already previously shown to cluster together with *ace3* in a
464 dataset from a transcriptional analysis with several different inducing carbon sources [33]. These
465 genes included *egl2*, *egl3*, *cbh2*, *swo1*, *cip2*, *xyn3*, *cel1b* and *cel3c*. These results indicate that
466 regardless of the environmental conditions (different inducing substrates or different ambient pH),
467 the genes most probably under *ace3* regulation are co-expressed with *ace3*.

468

469 Genomic secondary metabolism gene clusters of various fungi are believed to be under pH-
470 dependent regulation [58]. Clusters most likely to be involved in the biosynthesis of siderophores
471 were detected during this study. A common characteristic for these clusters was the presence of
472 genes preferring higher pH. In various fungi, the siderophore biosynthesis genes are activated by the
473 global regulator CCAAT-binding complex (CBC) during iron deprivation [58]. CBC complex itself
474 is activated according to the redox status and subsequently interacts with HapX (a regulator sensing
475 iron starvation). The solubility of iron decreases as pH increases. Therefore at pH6, integration of
476 information about the redox status and iron deprivation by the CBC complex leads to activation of
477 the genes of the siderophore biosynthesis cluster.

478

479 The genomic cluster containing the *grd1* gene which is believed to be involved in cellulase signal
480 transduction also included the gene 111742 encoding a transcriptional regulator. Both genes were
481 more highly expressed at low pH compared to high pH. The over-expression of gene 111742
482 (pMH14) had a negative effect on the production of cellulase activity and especially on
483 endoglucanase activity [33]. GRDI, however, has been shown to have a positive effect on cellulase
484 activity [37]. These results make this genomic cluster an interesting target for the study of cellulase
485 signalling in *T. reesei*.

486

487 Conclusions

488 In this study, the pH-responsive genes from the genome of *T. reesei* were identified, including
489 genes affected by the regulator of ambient pH sensing, PACI. The effect of ambient pH on the
490 genes of the pH-signalling cascade leading to activation of PACI was investigated and novel
491 information was gained from the pH-regulation of the carbohydrate active enzyme genes and

492 especially genes encoding cellulose- and hemicellulose-degrading enzymes. In addition, several co-
493 regulated genomic clusters responding to changes in ambient pH were identified.

494

495 Materials and methods

496 Strains, media and culture conditions

497 Escherichia coli DH5α (*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1*
498 *relA1 endA1 thi-1 hsdR17*) was used for propagation of the plasmids and *Saccharomyces cerevisiae*
499 FY834 (*MAT*α *his3Δ200 ura3-52 leu2Δ1 lys2Δ202 trp1Δ63*) for yeast recombinational cloning.
500 *Trichoderma reesei* QM9414 (ATCC 26921, VTT-D-74075) was obtained from VTT Culture
501 Collection and maintained on potato-dextrose agar plates (Difco). Spore suspensions were prepared
502 by cultivating the fungus on potato-dextrose plates for 5 days, after which the spores were
503 dislodged, suspended in a buffer containing 0.8 % NaCl, 0.025 % Tween20 and 20 % glycerol,
504 filtered through cotton and stored at -80 °C. In order to collect mycelia for DNA isolation, the
505 fungus was grown in a medium containing 0.2 % proteose peptone (BD Biosciences), 2 % glucose,
506 7.6 g/l $(NH_4)_2SO_4$, 15.0 g/l KH_2PO_4, 2.4mM $MgSO_4.7H_2O$, 4.1 mM $CaCl_2.H_2O$, 3.7 mg/l $CoCl_2$, 5
507 mg/l $FeSO_4.7H_2O$, 1.4 mg/l $ZnSO_4.7H_2O$, 1.6 mg/l $MnSO_4.7H_2O$, pH 4.8. Flask inoculum for the
508 bioreactor cultivations was prepared by cultivating 8 x 10^7 spores for 3 days in 400 ml of medium
509 containing 25 g/l Avicel cellulose (Fluka BioChemika), 6 g/l proteose peptone (BD Biosciences),
510 7.6 g/l $(NH_4)_2SO_4$, 15.0 g/l KH_2PO_4, 2.4 mM $MgSO_4.7H_2O$, 4.1 mM $CaCl_2.H_2O$, 3.7 mg/l $CoCl_2$, 5
511 mg/l $FeSO_4.7H_2O$, 1.4 mg/l $ZnSO_4.7H_2O$, 1.6 mg/l $MnSO_4.7H_2O$, pH adjusted to 5.2 with KOH.
512 Bioreactor medium contained 25 g/l Avicel cellulose, 6 g/l proteose peptone, 13.6 g/l $(NH_4)_2SO_4$, 4
513 g/l KH_2PO_4, 0.2 ml/l Tween80, 0.6 g/l $MgSO_4.7H_2O$, 0.8 g/l $CaCl_2.H_2O$, 3.7 mg/l $CoCl_2$, 5 mg/l
514 $FeSO_4.7H_2O$ 1.4 mg/l $ZnSO_4.7H_2O$, 1.6 mg/l $MnSO_4.7H_2O$ and 1 ml/l of antifoam agent (Dow

515 corning 1500). pH of the medium was adjusted before inoculation to pH3, 4.5 or 6. The bioreactor
516 containing 0.9 l of medium was inoculated with 0.1 l of flask inoculum. Bioreactor cultivations
517 were performed in Sartorius Q plus reactors at 28 °C with a dissolved oxygen saturation level of
518 >30 %, agitation of 500-1200 rpm and a constant air flow of 0.6 l/min. pH was controlled with 15
519 % KOH and 15 % H_3PO_4. Samples from the cultivations of the QM9414 strain at different pH
520 values were collected at time points of approximately 8 h, 18 h, 28 h, 44 h, 68 h, 93 h and 119 h.
521 Samples from the cultivation of the QM9414 and Δpac1 strains at pH6 were collected at time points
522 17 h, 24 h, 40.5 h, 69 h, 95.3 h, 116.5 h and 139.3 h.

523

524 **Construction of the *pac1* deletion strain**

525 Deletion cassette for the knock-out of the *pac1* gene was constructed by yeast recombination
526 cloning [59] in the pRS426 plasmid [60]. The construct contained a hygromycin resistance cassette
527 flanked by 1890 bp fragments from 5´ and 3´ ends of the *pac1* gene. Fragment 5´ from the *pac1*
528 open reading frame was PCR amplified with the oligos 5´-
529 GTAACGCCAGGGTTTTCCCAGTCACGACGGTTTAAACCCACCTCACCAGCCTTTGGTTT
530 GCA-3´ and 5´-
531 ACCGGGATCCACTTAACGTTACTGAAATCGGAAGGTCTTGGCGGTCTTGGCATT-3´.
532 Fragment 3´ from the *pac1* open reading frame was PCR amplified with the oligos 5´-
533 ATGCCAGAAAGAGTCACCGGTCACTGTACTGAATGAACATTCTTCAACATAACA-3´and
534 5´-
535 GCGGATAACAATTTCACACAGGAAACAGCGTTTAAACCGCAGCAGCAGCATTGCTTGG
536 GCGG-3´. Hygromycin resistance cassette was PCR amplified with the oligos 5´-
537 AGACAATGCCAAGACCGCCAAGACCTTCCGATTTCAGTAACGTTAAGTG-3´and 5´
538 TGGGTGTTATGTTGAAGAATGTTCATTCAGTACAGTGACCGGTGACTCT-3´. All the PCR

539 products were purified and combined with the vector backbone. The resulting plasmid was digested
540 with *PmeI* and *SspI* enzymes (New England biolabs) and the deletion cassette was transformed to *T.*
541 *reesei* QM9414 by polyethylene glycol mediated protoplast transformation [61]. The hygromycin
542 resistant transformants were initially selected from plates containing 150 µg ml^{-1} of hygromycin B
543 (Calbiochem). Stable transformants were obtained by streaking on plates containing 125 µg ml^{-1} of
544 hygromycin B for two successive rounds. Single colonies resulting from uninuclear transformants
545 were obtained by plating dilutions of spore suspensions. Transformants with the expression cassette
546 at the correct locus were identified by PCR and integration of only one copy of the cassette was
547 further verified by Southern blot analysis. Genomic DNA for Southern blot was isolated using
548 Easy-DNA Kit (Invitrogen) according to the manufacturer´s instructions. Southern blotting and
549 hybridisation on nitrocellulose filters (Hybond N, GE Healthcare) were carried out according to
550 standard procedures [62]. The signals were detected using a phosphorimager (Typhoon imager, GE
551 Healthcare). The deletion of the *pac1* gene was confirmed by Norhern blot analysis. Total RNA was
552 isolated from the mycelium samples using the Trizol™ Reagent (Invitrogen Life Technologies,
553 Carlsbad, CA, USA) essentially according to the manufacturer's instructions. Northern blotting and
554 hybridisation on nitrocellulose filters (Hybond N, GE Healthcare) were carried out according to
555 standard procedures [62].

556

557 **Sample preparation and microarray analytics**

558 Frozen mycelium was ground under liquid nitrogen, and total RNA was isolated with Trizol reagent
559 according to the manufacturer´s instructions. RNA was subsequently purified using RNeasy Mini
560 Kit (Qiagen, Hilden, Germany) and RNA concentration was measured using NanoDrop ND-1000
561 (NanoDrop Technologies Inc. Wilmington, DE, USA). Integrity of the isolated RNA was verified
562 using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA).

Processing of the RNA samples for the microarray analysis was performed essentially according to the instructions from RocheNimblegen. After synthesising the double-stranded cDNA using Superscript Double-Stranded cDNA synthesis Kit (Invitrogen), the integrity of the double-stranded cDNA was analysed using an Agilent 2100 Bioanalyzer. The double-stranded cDNA was labelled with Cy3 fluorescent dye, after which it was hybridized to microarray slides (Roche-NimbleGen, Inc., Madison, WI, USA) and the slides were scanned using a Roche NimbleGen Microarray scanner according to the instructions of the manufacturer. The probe design and manufacturing of the microarray slides was carried out by RocheNimbleGen using an array design based on the *T. reesei* genome version 2.0 [26]. The design includes six 60mer probes for each of the genes.

The data was pre-processed using the R package Oligo and differentially expressed genes were identified with the package Limma [24, 25, 63]. Differentially expressed genes were identified by comparing the signals from the different pH-conditions and between the different strains at the corresponding time point. pH6 samples were compared to pH3 and pH4.5 samples and pH4.5 samples were compared to pH3 samples. Signals from the cultivation of the Δpac1 strain were compared to those of the QM9414 strain from the same cultivation conditions. Three biological replicates were analysed for each condition and each time point. The threshold used for statistical significance was p-value <0.01 and log2-scale fold change >0.4.

Enzyme assays and total protein measurements

Cellulase activity assay against the MUL substrate is based on detecting the fluorescent hydrolysis product methyl umbelliferone (MU) released from the substrate 4-methyl umbelliferyl-β-D-

586 lactoside (MUL) (Sigma-Aldrich) as described in [64]. β-glucosidase activity was inhibited with
587 glucose to measure the combined activity of endoglucanase 1 (EGI) and cellobiohydrolase 1
588 (CBHI) enzymes. EGI activity was measured by adding cellobiose to inhibit CBHI and glucose to
589 inhibit β-glucosidase and CBHI activity was deduced by subtracting EGI activity from the
590 combined activity of CBHI and EGI enzymes. Endo-β-1,4-xylanase activity was assayed using 1.0
591 % birch glucuronoxylan as a substrate [65] and pure xylose (Merck) as a standard. The released
592 reducing sugars were detected with DNS. The amount of soluble proteins was measured with the
593 Bio-Rad protein assay (Bio-Rad Laboratories, Hercules, CA, USA) with Bovine serum albumin
594 (Sigma-Aldrich Chemie GmbH, Steinhem, Germany) as a standard.

595

596 **Abbreviations**: CAZy, carbohydrate active enzyme; CBHI, cellobiohydrolase 1; DNS, 2-hydroxy-
597 3,5-dinitrobenzoic acid; EGI, endo-β-1,4-glucanase 1; JGI, Joint Genome Institution; KOG,
598 Eukaryotic orthologous groups; MU, methyl umbelliferone; MUL, 4-methyl umbelliferyl-β-D-
599 lactoside

600

601 Competing interests

602 The authors declare that they have no competing interests.

603

604 Authors´ contributions

605 MH constructed the deletion strain, carried out enzymatic activity measurements and microarray
606 detection of the expression signals, participated in the analysis and interpretation of the microarray
607 data, and drafted the manuscript, DS carried out the bioreactor cultivations including sample

608 collection and monitoring of the cultivations, NA participated in the initial analysis of the effects of
609 the *pac1* gene, MS participated in the design and coordination of the study, TMP participated in the
610 design and coordination of the study, carried out microarray data analysis and helped to draft the
611 manuscript. All authors read and approved the final manuscript.

612

613 Acknowledgements

614 Aili Grundström, Eila Leino and Kirsi Multanen are acknowledged for extremely skilful technical
615 assistance. The work was co-funded by the VTT Technical Research Centre of Finland and by the
616 Academy of Finland (Decision no 133455).

617

618 References

619 1. Peñalva MA, Tilburn J, Bignell E, Arst Jr HN: **Ambient pH gene regulation in fungi: making**
620 **connections**. *Trends Microbiol* 2008, **16**:291–300.

621 2. Herranz S, Rodríguez JM, Bussink H-J, Sánchez-Ferrero JC, Arst HN, Peñalva MA, Vincent O:
622 **Arrestin-related proteins mediate pH signaling in fungi**. *Proc Natl Acad Sci U S A* 2005,
623 **102**:12141–12146.

624 3. Calcagno-Pizarelli AM, Negrete-Urtasun S, Denison SH, Rudnicka JD, Bussink H-J, Múnera-
625 Huertas T, Stanton L, Hervás-Aguilar A, Espeso EA, Tilburn J, Arst HN, Peñalva MA:
626 **Establishment of the ambient pH Signaling complex in *Aspergillus nidulans*: PalI assists**
627 **plasma membrane localization of PalH**. *Eukaryot Cell* 2007, **6**:2365–2375.

628 4. Hervás-Aguilar A, Galindo A, Peñalva MA: **Receptor-independent ambient pH signaling by**
629 **ubiquitin attachment to fungal arrestin-like PalF**. *J Biol Chem* 2010, **285**:18095–18102.

630 5. Galindo A, Calcagno-Pizarelli AM, Arst HN, Peñalva MÁ: **An ordered pathway for the**
631 **assembly of fungal ESCRT-containing ambient pH signalling complexes at the plasma**
632 **membrane**. *J Cell Sci* 2012, **125**:1784–1795.

633 6. Diez E, Alvaro J, Espeso EA, Rainbow L, Suarez T, Tilburn J, Arst HN, Penalva MA:
634 **Activation of the *Aspergillus* PacC zinc finger transcription factor requires two proteolytic**
635 **steps**. *EMBO J* 2002, **21**:1350–1359.

636 7. Hervás-Aguilar A, Rodríguez JM, Tilburn J, Arst HN, Peñalva MA: **Evidence for the direct**
637 **involvement of the proteasome in the proteolytic processing of the *Aspergillus nidulans* zinc**
638 **finger transcription factor PacC.** *J Biol Chem* 2007, **282**:34735–34747.

639 8. Espeso EA, Tilburn J, Sánchez-Pulido L, Brown CV, Valencia A, Arst HN, Peñalva MA:
640 **Specific DNA recognition by the *Aspergillus nidulans* three zinc finger transcription factor**
641 **PacC.** *J Mol Biol* 1997, **274**:466–480.

642 9. Tilburn J, Sarkar S, Widdick DA, Espeso EA, Orejas M, Mungroo J, Peñalva MA, Arst Jr HN:
643 **The *Aspergillus* PacC zinc finger transcription factor mediates regulation of both acid- and**
644 **alkaline-expressed genes by ambient pH.** *EMBO J* 1995, **14**:779–790.

645 10. Caddick MX, Brownlee AG, Arst Jr. HN: **Regulation of gene expression by pH of the growth**
646 **medium in *Aspergillus nidulans***. *Mol Gen Genet* 1986, **203**:346–353.

647 11. Arst Jr. HN, Bignell E, Tilburn J: **Two new genes involved in signalling ambient pH in**
648 ***Aspergillus nidulans***. *Mol Gen Genet* 1994, **245**:787–790.

649 12. Vankuyk PA, Diderich JA, MacCabe AP, Hererro O, Ruijter GJG, Visser J: ***Aspergillus niger***
650 **mstA encodes a high-affinity sugar/H+ symporter which is regulated in response to**
651 **extracellular pH.** *Biochem J* 2004, **379**:375– 383.

652 13. Maccabe AP, Orejas M, Pérez-González JA, Ramón D: **Opposite patterns of expression of**
653 **two *Aspergillus nidulans* xylanase genes with respect to ambient pH**. *J Bacteriol* 1998,
654 **180**:1331–1333.

655 14. Gielkens M, González-Candelas L, Sánchez-Torres P, van de Vondervoort P, de Graaff L,
656 Visser J, Ramón D: **The *abfB* gene encoding the major a-L-arabinofuranosidase of *Aspergillus***
657 ***nidulans*: nucleotide sequence, regulation and construction of a disrupted strain**. *Microbiology*
658 1999, **145**:735–741.

659 15. Ke R, Haynes K, Stark J: **Modelling the activation of alkaline pH response transcription**
660 **factor PacC in *Aspergillus nidulans*: Involvement of a negative feedback loop**. *J Theor Biol*
661 2013, **326**:11–20.

662 16. Penttilä M, Limón C, Nevalainen H: **Molecular biology of *Trichoderma* and biotechnological**
663 **applications.** In *Mycol Handb fungal Biotechnol*. 2. edition.; 2004:413–427.

664 17. Saloheimo M, Pakula TM: **The cargo and the transport system: secreted proteins and**
665 **protein secretion in *Trichoderma reesei* (*Hypocrea jecorina*)**. *Microbiology* 2012, **158**:46–57.

666 18. Aro N, Pakula T, Penttilä M: **Transcriptional regulation of plant cell wall degradation by**
667 **filamentous fungi**. *FEMS Microbiol Rev* 2005, **29**:719–739.

668 19. Seiboth B, Herold S, Kubicek C: **Metabolic engineering of inducer formation for cellulase**
669 **and hemicellulase gene expression in *Trichoderma reesei***. In *Reprogramming Microb Metab*
670 *Pathways*. *Volume 64*. Edited by Wang X, Chen J, Quinn P. Springer Netherlands; 2012:367–390.
671 [*Subcellular Biochemistry*]

672 20. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov
673 O, Coutinho PM, Cullen D, Danchin EGJ, Grigoriev I V, Harris P, Jackson M, Kubicek CP, Han
674 CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N,
675 Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, et al.:
676 **Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn.
677 *Hypocrea jecorina*)**. *Nat Biotechnol* 2008, **26**:553–560.

678 21. Bailey M, Buchert J, Viikari L: **Effect of pH on production of xylanase by *Trichoderma*
679 *reesei* on xylan- and cellulose-based media**. *Appl Microbiol Biotechnol* 1993, **40**:224–229.

680 22. Li C, Yang Z, He Can Zhang R, Zhang D, Chen S, Ma L: **Effect of pH on cellulase production
681 and morphology of *Trichoderma reesei* and the application in cellulosic material hydrolysis**. *J
682 Biotechnol* 2013, **168**:470–477.

683 23. Adav SS, Ravindran A, Chao LT, Tan L, Singh S, Sze SK: **Proteomic analysis of pH and
684 strains dependent protein secretion of *Trichoderma reesei***. *J Proteome Res* 2011, **10**:4579–4596.

685 24. Bioconductor open source software for bioinformatics: **http://www.bioconductor.org/**. .

686 25. Smyth GK, Michaud J, Scott HS: **Use of within-array replicate spots for assessing
687 differential expression in microarray experiments**. *Bioinformatics* 2005, **21**:2067–2075.

688 26. v2.0 T reesei database: **[http://genome.jgi-psf.org/Trire2/Trire2.home.html]**. .

689 27. v1.0 T reesei database: **[http://genome.jgi-psf.org/trire1/trire1.home.html]**. .

690 28. Arvas M, Haiminen N, Smit B, Rautio J, Vitikainen M, Wiebe M, Martinez D, Chee C, Kunkel
691 J, Sanchez C, Nelson MA, Pakula T, Saloheimo M, Penttilä M, Kivioja T: **Detecting novel genes
692 with sparse arrays**. *Gene* 2010, **467**:41–51.

693 29. Kumar L, Futschik ME: **Mfuzz: a software package for soft clustering of microarray data.**
694 *Bioinformation* 2007, **2**:5–7.

695 30. Häkkinen M, Arvas M, Oja M, Aro N, Penttilä M, Saloheimo M, Pakula T: **Re-annotation of**
696 **the CAZy genes of *Trichoderma reesei* and transcription in the presence of lignocellulosic**
697 **substrates**. *Microb Cell Fact* 2012, **11**:134.

698 31. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The**
699 **Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics**.
700 *Nucleic Acids Res* 2009, **37**(suppl 1):D233–D238.

701 32. Database CAE: **[http://www.cazy.org]**. .

702 33. Häkkinen M, Valkonen MJ, Westerholm-Parvinen A, Aro N, Arvas M, Vitikainen M, Penttilä
703 M, Saloheimo M, Pakula TM: **Screening of candidate regulators for cellulase and hemicellulase**
704 **production in *Trichoderma reesei* and identification of a factor essential for cellulase**
705 **production**. *Biotechnol Biofuels* 2014, **7**:14.

706 34. Trushina N, Levin M, Mukherjee PK, Horwitz BA: **PacC and pH-dependent transcriptome**
707 **of the mycotrophic fungus *Trichoderma virens*.** *BMC Genomics* 2013, **14**:138.

708 35. Schmoll M, Kubicek CP: ***ooc1*, a unique gene expressed only during growth of *Hypocrea***
709 ***jecorina* (anamorph: *Trichoderma reesei*) on cellulose.** *Curr Genet* 2005, **48**:126–133.

710 36. Porciuncula de Oliveira J, Furukawa T, Shida Y, Mori K, Kuhara S, Morikawa Y, Ogasawara
711 W: **Identification of major facilitator transporters involved in cellulase production during**
712 **lactose culture of *Trichoderma reesei* PC-3-7**. *Biosci Biotechnol Biochem* 2013, **77**:1014–1022.

713 37. Schuster A, Kubicek CP, Schmoll M: **Dehydrogenase GRD1 represents a novel component**
714 **of the cellulase regulon in *Trichoderma reesei* (*Hypocrea jecorina*).** *Appl Environ Microbiol*
715 2011, **77**:4553–4563.

716 38. Zhang G, Zhu Y, Wei D, Wang W: **Enhanced production of heterologous proteins by the**
717 **filamentous fungus *Trichoderma reesei* via disruption of the alkaline serine protease SPW**
718 **combined with a pH control strategy**. *Plasmid* 2014, **71**:16–22.

719 39. Alkan N, Meng X, Friedlander G, Reuveni E, Sukno S, Sherman A, Thon M, Fluhr R, Prusky
720 D: **Global aspects of pacC regulation of pathogenicity genes in *Colletotrichum gloeosporioides***
721 **as revealed by transcriptome analysis.** *Mol plant-microbe Interact* 2013, **26**:1345–1358.

722 40. Penalva MA, Arst HN: **Regulation of gene expression by ambient pH in filamentous fungi**
723 **and yeasts**. *Microbiol Mol Biol Rev* 2002, **66**:426–446.

724 41. Hutchings H, Stahmann K-P, Roels S, Espeso EA, Timberlake WE, Arst Jr HN, Tilburn J: **The**
725 **multiply-regulated *gabA* gene encoding the GABA permease of *Aspergillus nidulans*: a score**
726 **of exons**. *Mol Microbiol* 1999, **32**:557–568.

727 42. Espeso EA, Arst HN: **On the mechanism by which alkaline pH prevents expression of an**
728 **acid-expressed gene**. *Mol Cell Biol* 2000, **20**:3355–3363.

729 43. Caracuel Z, Casanova C, Roncero MIG, Di Pietro A, Ramos J: **pH response transcription**
730 **factor PacC controls salt stress tolerance and expression of the P-Type Na+-ATPase Ena1 in**
731 ***Fusarium oxysporum***. *Eukaryot Cell* 2003, **2**:1246–1252.

732 44. Landraud P, Chuzeville S, Billon-Grande G, Poussereau N, Bruel C: **Adaptation to pH and**
733 **role of PacC in the rice blast fungus *Magnaporthe oryzae*.** *PLoS One* 2013, **8**:e69236.

734 45. Andersen MR, Lehmann L, Nielsen J: **Systemic analysis of the response of *Aspergillus niger***
735 **to ambient pH.** *Genome Biol* 2009, **10**:R47.

736 46. Negrete-Urtasun S, Reiter W, Diez E, Denison SH, Tilburn J, Espeso EA, Penalva MA, Arst Jr
737 HN: **Ambient pH signal transduction in *Aspergillus*: completion of gene characterization**. *Mol*
738 *Microbiol* 1999, **33**:994–1003.

739 47. Tenkanen M, Puls J, Poutanen K: **Two major xylanases of *Trichoderma reesei***. *Enzyme*
740 *Microb Technol* 1992, **14**:566–574.

741 48. Torronen A, Mach RL, Messner R, Gonzalez R, Kalkkinen N, Harkki A, Kubicek CP: **The two**
742 **major xylanases from *Trichoderma reesei*: Characterization of both enzymes and genes**. *Nat*
743 *Biotechnol* 1992, **10**:1461–1465.

744 49. Xu J, Takakuwa N, Nogawa M, Okada H, Okada H: **A third xylanase from *Trichoderma***
745 ***reesei* PC-3-7**. *Appl Microbiol Biotechnol* 1998, **49**:718–724.

746 50. Wang J, Zeng D, Mai G, Liu G, Yu S: **Homologous constitutive expression of Xyn III in**
747 ***Trichoderma reesei* QM9414 and its characterization.** *Folia Microbiol (Praha)* 2013, **59**:229–
748 233.

749 51. Xiong H, von Weymarn N, Leisola M, Turunen O: **Influence of pH on the production of**
750 **xylanases by *Trichoderma reesei* Rut C-30**. *Process Biochem* 2004, **39**:731–736.

751 52. Graaff LK, Broeck HC, Ooijen AJJ, Visser J: **Regulation of the xylanase-encoding *xlnA* gene**
752 **of *Aspergilius tubigensis***. *Mol Microbiol* 1994, **12**:479–490.

753 53. Stricker AR, Grosstessner-Hain K, Wurleitner E, Mach RL: **Xyr1 (xylanase regulator 1)**
754 **regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina***.
755 *Eukaryot Cell* 2006, **5**:2128–2137.

756 54. Aro N, Saloheimo A, Ilmén M, Penttilä M: **ACEII, a novel transcriptional activator involved**
757 **in regulation of cellulase and xylanase genes of *Trichoderma reesei***. *J Biol Chem* 2001,
758 **276**:24309–24314.

759 55. Aro N, Ilmen M, Saloheimo A, Penttilä M: **ACEI of *Trichoderma reesei* is a repressor of**
760 **cellulase and xylanase expression**. *Appl Environ Microbiol* 2003, **69**:56–65.

761 56. Saloheimo A, Aro N, Ilmén M, Penttilä M: **Isolation of the *ace1* gene encoding a Cys2-His2**
762 **transcription factor involved in regulation of activity of the cellulase promoter *cbh1* of**
763 ***Trichoderma reesei***. *J Biol Chem* 2000, **275**:5817–5825.

764 57. Ilmén M, Thrane C, Penttilä M: **The glucose repressor gene *cre1* of *Trichoderma*: Isolation**
765 **and expression of a full-length and a truncated mutant form**. *Mol Gen Genet* 1996, **251**:451–
766 460.

767 58. Brakhage AA: **Regulation of fungal secondary metabolism**. *Nat Rev Microbiol* 2013, **11**:21–
768 32.

769 59. Colot HV, Park G, Turner GE, Ringelberg C, Crew CM, Litvinkova L, Weiss RL, Borkovich
770 KA, Dunlap JC: **A high-throughput gene knockout procedure for *Neurospora* reveals functions**
771 **for multiple transcription factors**. *Proc Natl Acad Sci* 2006, **103**:10352–10357.

772 60. Christianson TW, Sikorski RS, Dante M, Shero JH, Hieter P: **Multifunctional yeast high-**
773 **copy-number shuttle vectors**. *Gene* 1992, **110**:119–122.

774 61. Penttilä M, Nevalainen H, Rättö M, Salminen E, Knowles J: **A versatile transformation**
775 **system for the cellulolytic filamentous fungus *Trichoderma reesei***. *Gene* 1987, **61**:155–164.

776 62. Sambrook J, Fritsch E, Maniatis T: *Molecular Cloning: A Laboratory Manual*. 2. edition. Cold
777 Spring Harbor, NY; 1989.

778 63. Bolstad BM, Irizarry RA, Åstrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias**. *Bioinformatics* 2003, **19**:185–193.

64. Bailey MJ, Tähtiharju J: **Efficient cellulase production by *Trichoderma reesei* in continuous cultivation on lactose medium with a computer-controlled feeding strategy**. *Appl Microbiol Biotechnol* 2003, **62**:156–162.

65. Bailey MJ, Biely P, Poutanen K: **Interlaboratory testing of methods for assay of xylanase activity**. *J Biotechnol* 1992, **23**:257–270.

Legends for figures

Figure 1. Venn-diagram from the pH-responsive genes of *T. reesei* and classification of the genes. A, the diagram shows the direction of change of gene expression in the parental strain when different pH conditions were compared and in the Δpac1 strain as compared to the parental strain; B, classification is based on Eukaryotic orthologous groups. Glycoside hydrolase, carbohydrate esterase and polysaccharide lyase genes were annotated according to [30].

Figure 2. Classification of the genes affected by the PACI transcription factor. Classification is based on Eukaryotic orthologous groups. Glycoside hydrolase, carbohydrate esterase and polysaccharide lyase genes were annotated according to [30]. A, genes that were either up-regulated or down-regulated in the *pac1* deletion strain as compared to the parental strain; B, genes that are induced or repressed by PACI according to statistically significant changes of expression both in the parental strain and in the Δpac1 strain.

800

801 **Figure 3. Expression profiles of the *palC* and *palF* gene homologues of *T. reesei* and of the**
802 ***pac1* gene.** The figure shows the log2 signal intensities of the genes at two different time points of
803 bioreactor cultivations at pH3, pH4.5 and pH6. The horizontal axis shows the different time points
804 and pH values.

805

806 **Figure 4. pH-responsive glycoside hydrolase, carbohydrate esterase and polysaccharide**
807 **lyase genes.** A, the number of pH-responsive genes from different CAZy families; B, the direction
808 of change of gene expression for cellulolytic and hemicellulolytic genes in the parental strain when
809 different pH conditions were compared and in the Δpac1 strain as compared to the parental strain**.**

810

811 **Figure 5. Heat map visualisation of expression data on the cellulase and hemicellulase**
812 **genes and of the characterized and putative regulatory genes.** The colour key indicates the
813 log2 scale fold change of the transcript signals in the comparison between two different pH
814 conditions or between two strains at a corresponding time point of cultivation (time point 1 or 2).
815 The genes are shown as rows and the samples as columns. The genes belonging to branches
816 including high pH up-regulated and low pH up-regulated genes are marked. QM, QM9414; Δ, *pac1*
817 deletion strain.

818

819 **Figure 6. Cellulase activity measured from the bioreactor cultivation of QM9414 and Δpac1**
820 **strains at pH6.** The fungal strains were cultivated in cellulose-containing medium for six days.
821 Panels A-C show the production of total MUL activity and of the specific CBHI and EGI activities.
822 CBHI, cellobiohydrolase 1; EGI, endo-β-1,4-glucanase 1. Error bars show the standard error of the
823 mean of three biological replicates.

824

825 **Figure 7. Xylanase activity and total protein production measured from the bioreactor**
826 **cultivation of QM9414 and ∆pac1 strains at pH6.** The fungal strains were cultivated in cellulose-
827 containing medium for six days. Error bars show the standard error of the mean of three biological
828 replicates. A, production of xylanase activity; B, production of total proteins; C, cumulative CO_2
829 production during the cultivation.

830

831 **Figure 8. Enzyme activities and total protein production measured from the bioreactor**
832 **cultivation of strain QM9414 in three different pH conditions.** The fungus was cultivated in
833 cellulose-containing medium for five days. Panels A-C show the production of total MUL activity,
834 xylanase activity and total proteins. Panel D shows the cumulative CO_2 production during the
835 cultivation. Error bars show the standard error of the mean of three biological replicates.

836

837 # Additional files

838 **Additional file 1 as XLSX**

839 **Additional file 1. Transcriptional profiling data of the pH-responsive genes.** Gene identifiers
840 are as in the *T. reesei* database version 2.0 and 1.2 (v1_2 before the gene ID) or as in [28]. Fold
841 changes (log2 scale), signal intensities (log2 scale) and significance test (R package limma, p <
842 0.01, log2 fold change > 0.4) are shown for the genes from two different time points (1 and 2) of
843 cultivations. 1 indicates induction and -1 repression. The intensity of the red and blue colour
844 indicates the strength of positive and negative fold changes, respectively. Colour scales of yellow,
845 red and green indicate different intensities of signals, red representing the strongest and green the
846 weakest signals. II, QM9414 samples from the cultivation of QM9414 and ∆pac1. Annotations are

847 shown according to the Eukaryotic orthologous group classification and as the manual annotations

848 from the JGI *T. reesei* 2.0 database. Functional Interpro domain identifiers are as in the InterPro

849 database. Annotations of glycoside hydrolase, carbohydrate esterase and polysaccharide lyase genes

850 are as in [30].

851

852 **Additional file 2 as XLSX**

853 **Additional file 2. Transcriptional profiling data of the genes affected by the *pac1* deletion.**

854 Gene identifiers are as in the *T. reesei* database version 2.0 or 1.2 (v1_2 before the gene ID). Fold

855 changes (log2 scale), signal intensities (log2 scale) and significance test (R package limma, $p <$

856 0.01, log2 fold change > 0.4) are shown for the genes from two different time points (1 and 2) of

857 cultivations. 1 indicates induction and -1 repression. The intensity of the red and blue colour

858 indicates the strength of positive and negative fold changes, respectively. Colour scales of yellow,

859 red and green indicate different intensities of signals, red representing the strongest and green the

860 weakest signals. II, QM9414 samples from the cultivation of QM9414 and Δpac1. Annotations are

861 shown according to the Eukaryotic orthologous group classification and as the manual annotations

862 from the JGI T. reesei 2.0 database. Functional Interpro domain identifiers are as in the InterPro

863 database. Annotations of glycoside hydrolase, carbohydrate esterase and polysaccharide lyase genes

864 are as in [30].

865

866 **Additional file 3 as XLSX**

867 **Additional file 3. Transcriptional profiling data of the CAZy genes.** Gene identifiers are as in

868 the *T. reesei* database version 2.0. Fold changes (log2 scale), signal intensities (log2 scale) and

869 significance test (R package limma, $p < 0.01$, log2 fold change > 0.4) are shown for the genes from

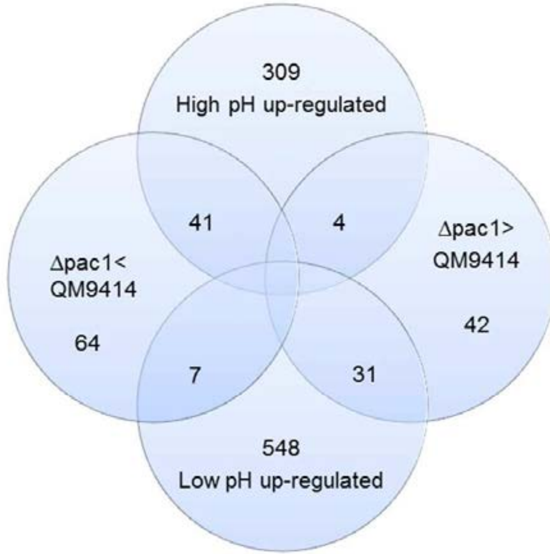
870 two different time points (1 and 2) of cultivations. 1 indicates induction and -1 repression. The
871 intensity of the red and blue colour indicates the strength of positive and negative fold changes,
872 respectively. Colour scales of yellow, red and green indicate different intensities of signals, red
873 representing the strongest and green the weakest signals. II, QM9414 samples from the cultivation
874 of QM9414 and Δpac1. Annotations of the genes are as in [30].

875

876 **Additional file 4 as XLSX**

877 **Additional file 4. Genes from the heat map visualisation.** Gene identifiers are as in the *T. reesei*
878 database version 2.0. Annotations are shown according to the Eukaryotic orthologous group
879 classification and as the manual annotations from the JGI *T. reesei* 2.0 database. Functional Interpro
880 domain identifiers are as in the InterPro database. Annotations of glycoside hydrolase, carbohydrate
881 esterase and polysaccharide lyase genes are as in [30]. References for the characterized and
882 candidate regulatory genes of the cellulase and hemicellulase genes are shown.

883



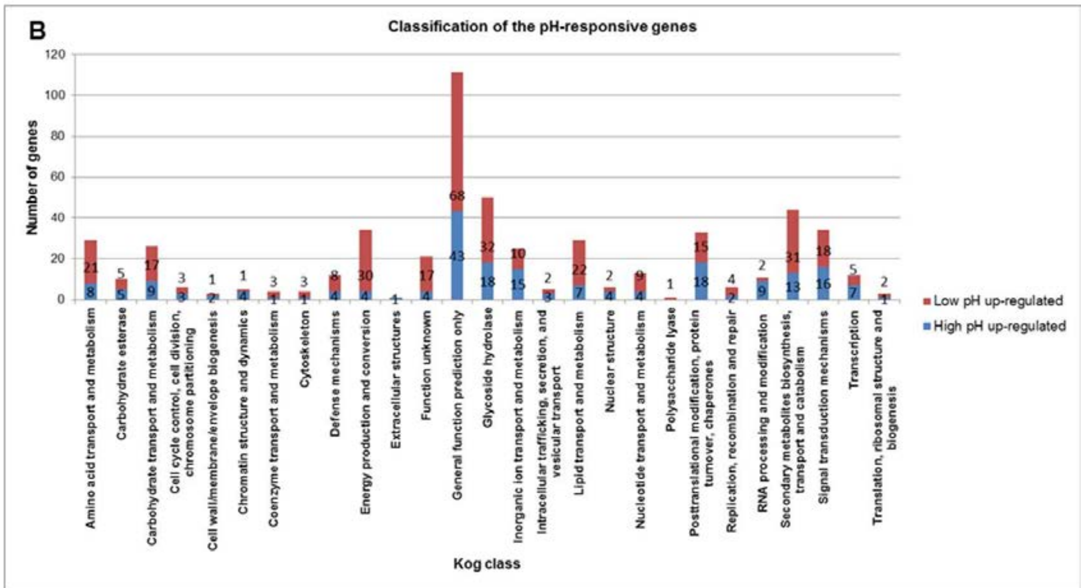
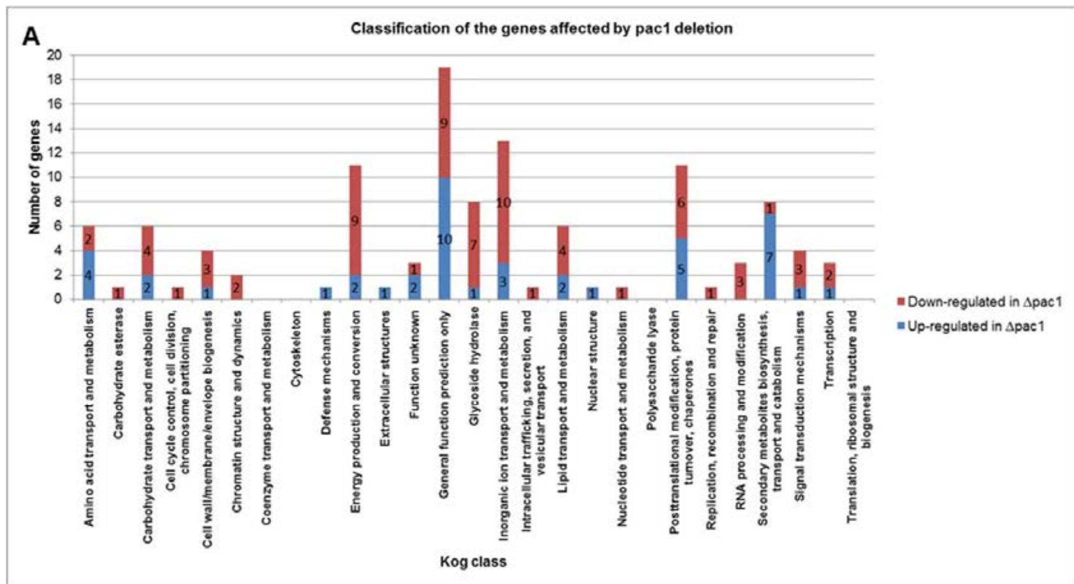


Figure 1



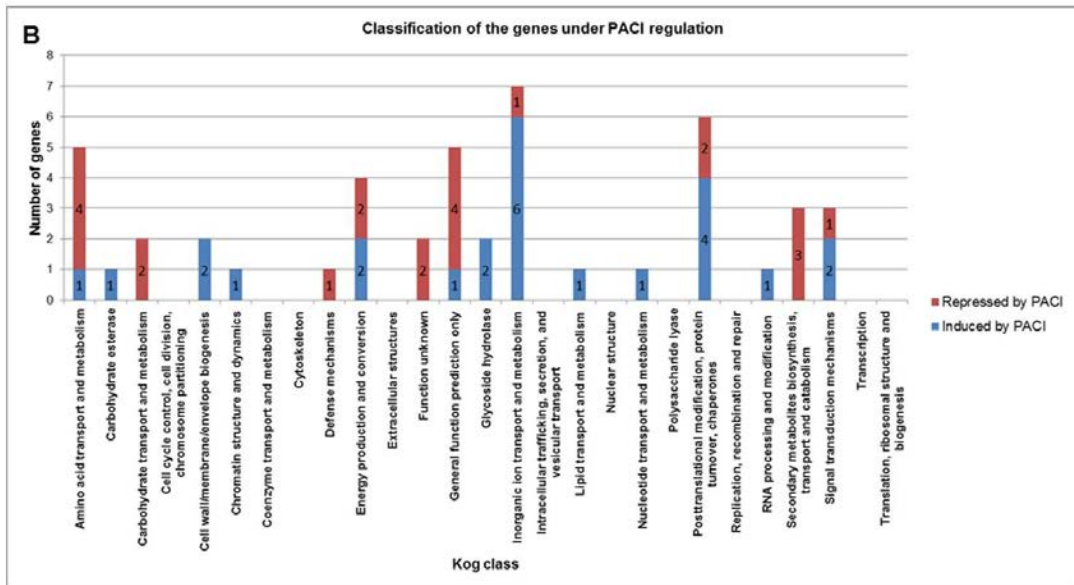
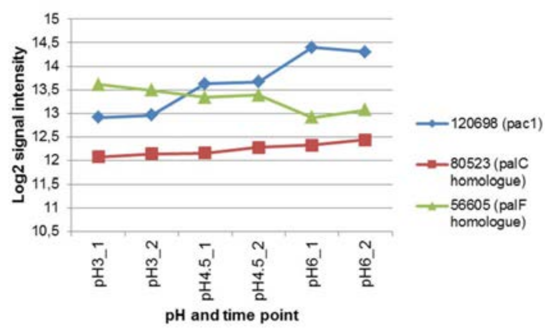
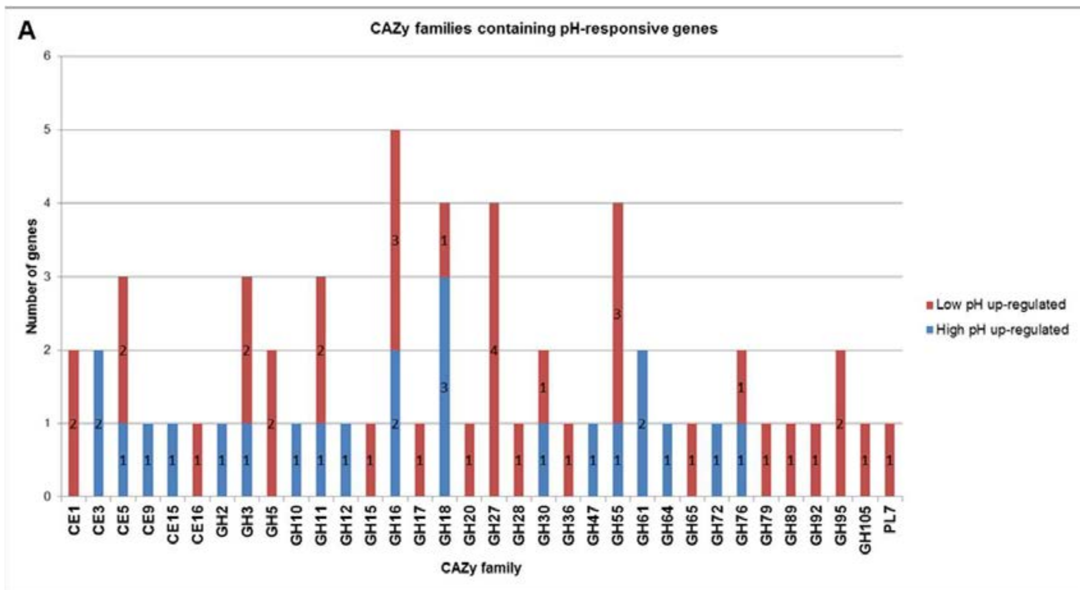


Figure 2


15
14,5
14
13,5
13
12,5
12
11,5
11
10,5
Log2 signal intensity
pH3_1
pH3_2
pH4.5_1
pH4.5_2
pH6_1
pH6_2
pH and time point
120698 (pac1)
80523 (palC homologue)
56605 (palF homologue)



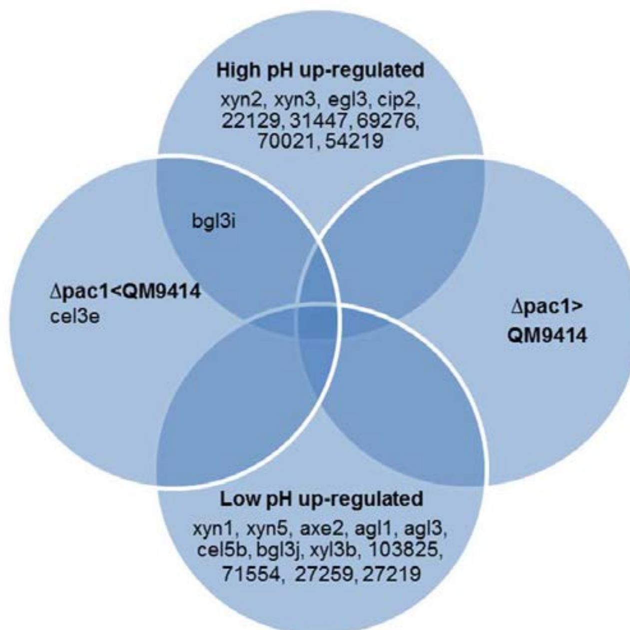


Figure 4

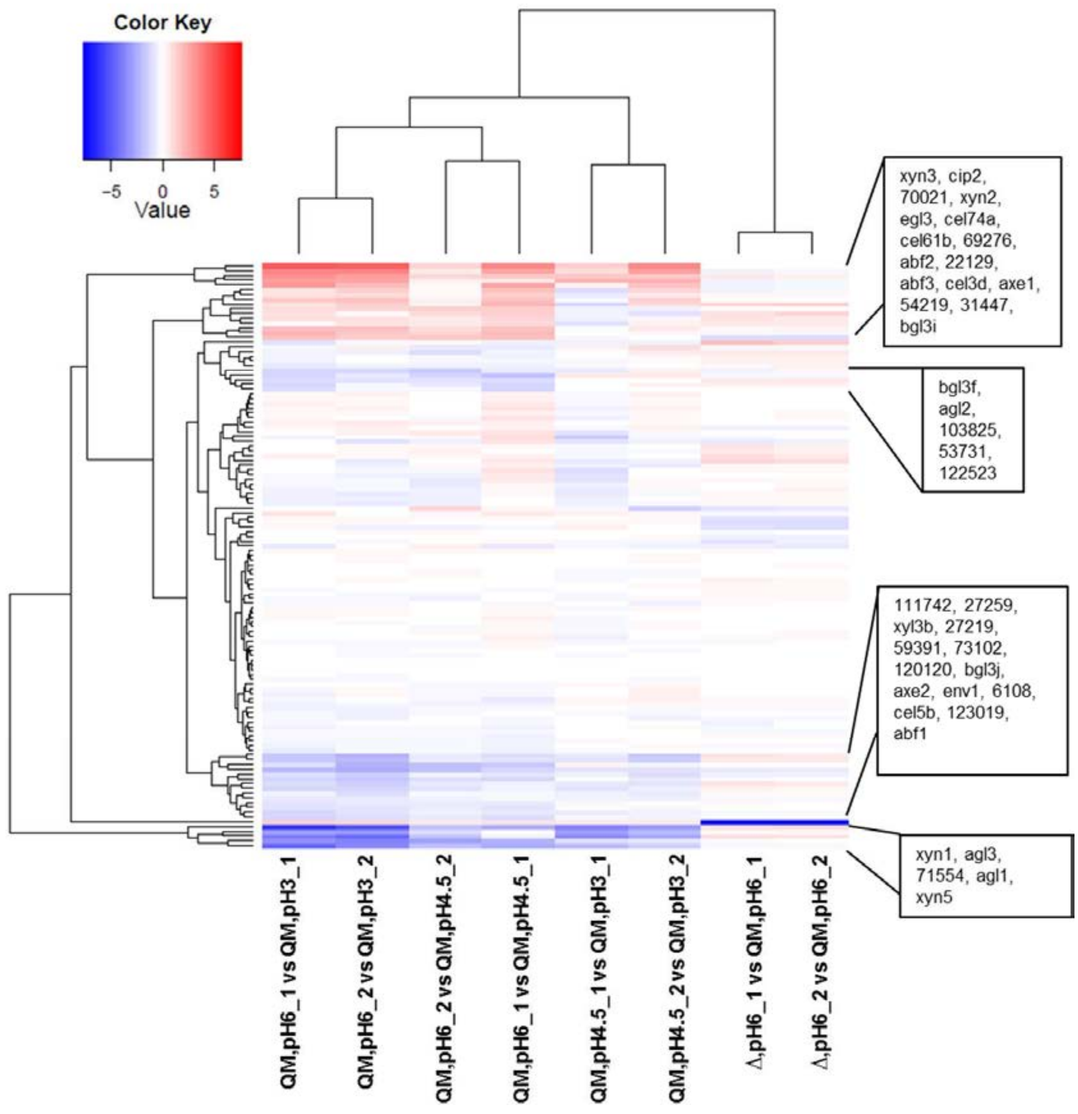
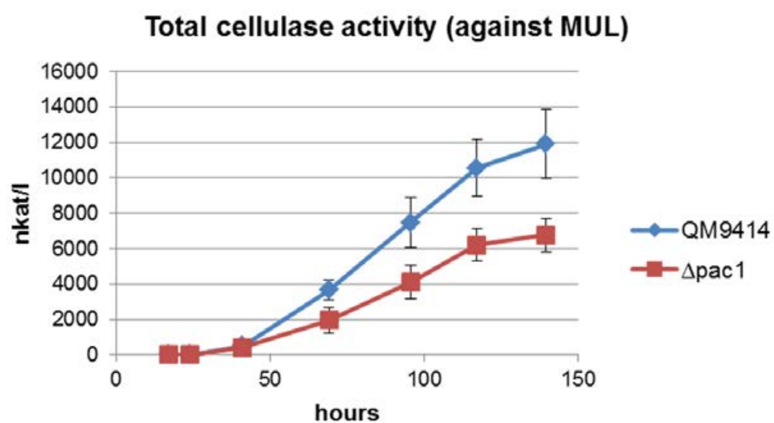
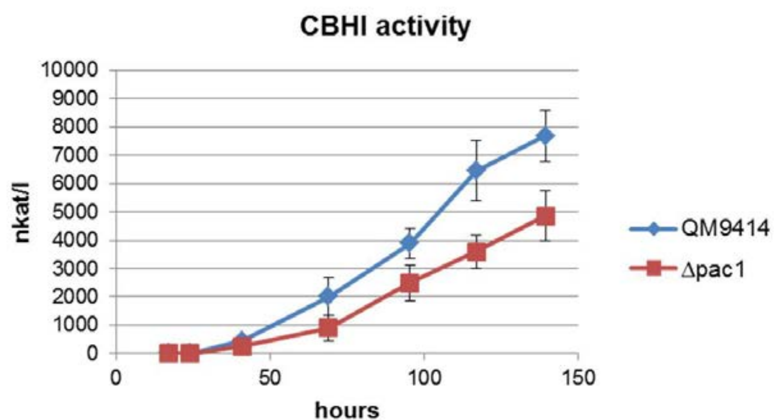


Figure 5

A



B



C

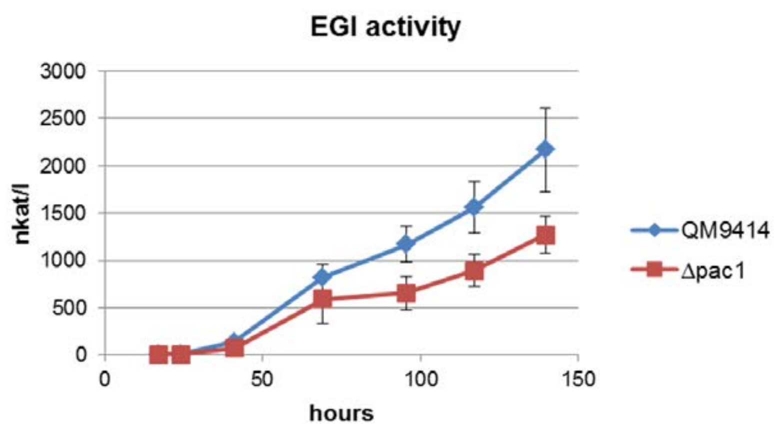
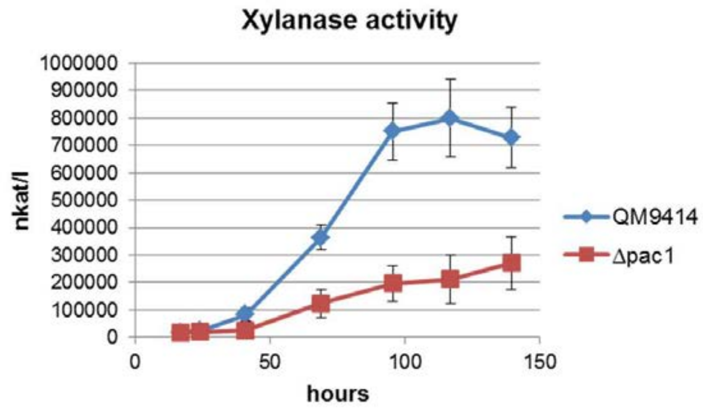
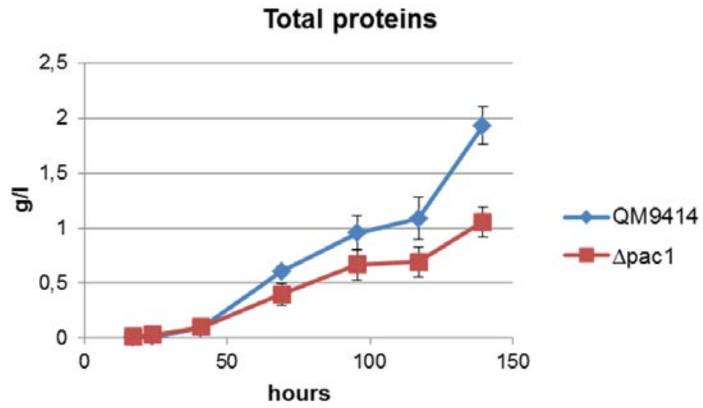


Figure 6

A



B



C

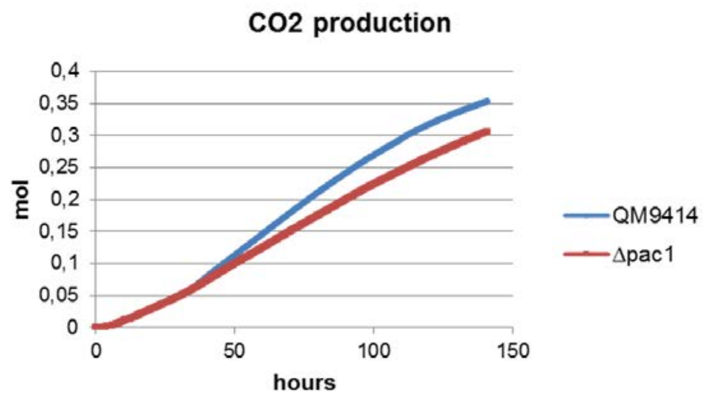
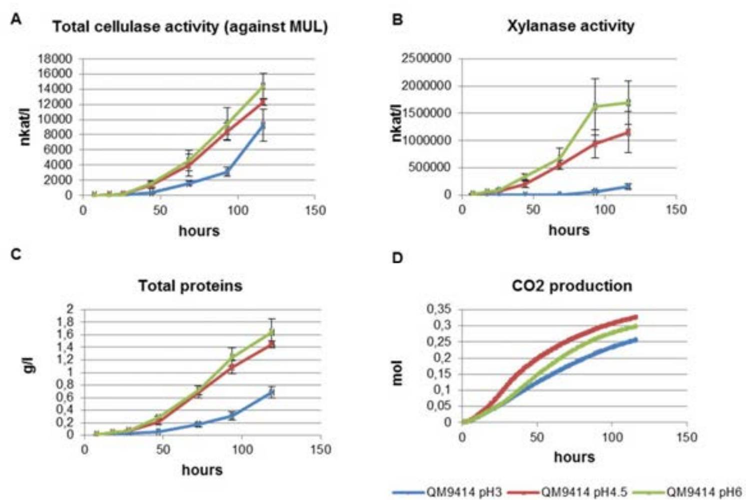


Figure 7


A
Total cellulase activity (against MUL)
nkat/l
hours
B
Xylanase activity
nkat/l
hours
C
Total proteins
g/l
hours
D
CO2 production
mol
hours
QM9414 pH3
QM9414 pH4.5
QM9414 pH6

Additional files provided with this submission:

Additional file 1: Additional file 1.xlsx, 413K
http://www.microbialcellfactories.com/imedia/1232886930143792/supp1.xlsx
Additional file 2: Additional file 2.xlsx, 103K
http://www.microbialcellfactories.com/imedia/1426523706143793/supp2.xlsx
Additional file 3: Additional file 3.xlsx, 128K
http://www.microbialcellfactories.com/imedia/1033236425143793/supp3.xlsx
Additional file 4: Additional file 4.xlsx, 25K
http://www.microbialcellfactories.com/imedia/3848546121437934/supp4.xlsx

Series title and number

VTT Science 65

Title	**Transcriptional analysis of *Trichoderma reesei* under conditions inducing cellulase and hemicellulase production, and identification of factors influencing protein production**
Author(s)	Mari Häkkinen
Abstract	Utilisation of non-edible, renewable lignocellulosic biomass for the production of second generation biofuels and chemicals is hindered especially by the high price of enzymes needed for biomass degradation. Filamentous fungi are natural producers of enzymes active against plant cell wall polymers. Especially the ascomycota fungus *Trichoderma reesei* is widely utilised in the industry for the production of cellulases and hemicellulases. However, the efficiency of enzyme production needs to be further improved in order to ensure economical production of biobased products. Several environmental factors affect protein production by filamentous fungi. Cellulase and hemicellulase genes of *T. reesei* are activated by inducer molecules derived from different substrates. The need for cooperation of different hydrolytic enzymes for the total degradation of plant cell wall material has led to coordinated expression of these genes. However, the extent and timing of induction can vary between different genes and especially the hemicellulase genes are differentially induced by various substrates. The direct regulation of cellulase and hemicellulase genes by transcriptional regulators has been widely studied and several activators and repressors of these genes have been characterized in detail. However, little is still known concerning the exact regulatory pathways and mechanisms utilised by the fungus for the accurate timing and composition of the hydrolytic enzymes produced. In this study, a genome-wide transcriptional analysis of *T. reesei* gene expression at different ambient pH conditions was conducted in order to identify genes affected by extracellular pH. The role of a *T. reesei* orthologue for the characterized pH regulator, PacC, in the expression of cellulase and hemicellulase genes was also studied. An extensive induction experiment together with transcriptional profiling was then utilised to study the effects of several different substrates on the expression of genes encoding carbohydrate active enzymes (CAZy). In addition, transcriptomics data was utilised for the identification of novel candidate regulators affecting cellulase and xylanase production by *T. reesei*. Transcriptional profiling identified pH as an important determinant of *T. reesei* gene expression. Ambient pH was also found to affect the expression of several cellulase and hemicellulase genes and more information on the role of a PacC orthologue in the expression of cellulase and hemicellulase genes was gained. A profiling study utilising different substrates as inducers together with a thorough annotation of the *T. reesei* CAZy genes revealed the expression patterns of novel candidate genes possibly involved in the degradation of different types of cellulosic and hemicellulosic substrates. In addition, a phylogenetic analysis indicated that functional diversification of the carbohydrate active enzymes of *T. reesei* is a rather common phenomenon and is reflected in the differential regulation of the corresponding genes. A transcription factor gene named *ace3* was identified from the profiling data and was shown to be essential for cellulase production and for the expression of cellulase genes. Over-expression of *ace3* led to improved production of cellulase and xylanase activities. Several other candidate regulators were also identified as interesting subjects for more detailed studies. Overall, the use of genome-wide methods increased understanding concerning the genome organisation of *T. reesei* and its possible evolutionary benefits, and enabled identification of co-regulated genomic regions possibly involved in enzyme production.
ISBN, ISSN	ISBN 978-951-38-8161-0 (Soft back ed.) ISBN 978-951-38-8162-7 (URL: http://www.vtt.fi/publications/index.jsp) ISSN-L 2242-119X ISSN 2242-119X (Print) ISSN 2242-1203 (Online)
Date	October 2014
Language	English, Finnish abstract
Pages	89 p. + app. 102 p.
Name of the project	
Commissioned by	
Keywords	Trichoderma reesei, cellulase, hemicellulase, gene regulation, transcription factors, transcriptional profiling, ambient pH
Publisher	VTT Technical Research Centre of Finland P.O. Box 1000, FI-02044 VTT, Finland, Tel. 020 722 111

Julkaisun sarja ja numero

VTT Science 65

Nimeke

Transkriptioanalyysi *Trichoderma reesei* -sienelle sellulaasien ja hemisellulaasien tuottoa aktivoivissa olosuhteissa ja proteiinin tuottoon vaikuttavien faktoreiden tunnistaminen

Tekijä(t)

Mari Häkkinen

Tiivistelmä

Biomassan hajotukseen tarvittavien entsyymien korkea hinta vaikeuttaa uusiutuvien lignoselluloosasta rakentuvien biomassamateriaalien käyttöä toisen sukupolven biopolttoaineiden ja kemikaalien tuotantoon. Rihmamaiset sienet tuottavat luonnostaan kasvien soluseinämateriaalia hajottavia entsyymejä. Varsinkin *Trichoderma reesei* -sientä käytetään laajasti teollisuudessa sellulaasien ja hemisellulaasien tuottoon. Sienen entsyymintuottokykyä tulee kuitenkin parantaa vielä entisestään, jotta biopohjaisten tuotteiden kustannustehokas tuotto voidaan varmistaa. Useilla eri ympäristötekijöillä tiedetään olevan vaikutusta rihmamaisen sienen proteiinintuottoon. *T. reesei* -sienen sellulaasi- ja hemisellulaasigeenit aktivoituvat erilaisista substraateista muodostuvien indusorien välityksellä. Kasvimateriaalin hajotukseen tarvitaan useiden eri entsyymien yhteistyötä, mikä on johtanut entsyymejä koodaavien geenien koordinoituun ekspressioon. Induktion voimakkuus ja ajoitus voivat kuitenkin vaihdella eri geenien välillä, ja erityisesti hemisellulaasigeenien induktiossa on havaittu vaihtelua myös eri substraattien välillä. Sellulaasi- ja hemisellulaasigeenien säätelyä spesifisten transkriptiofaktoreiden välityksellä on tutkittu laajasti ja useita aktivaattoreita ja repressoreja on karakterisoitu. Entsyymien tuoton ajoittamiseen ja optimaalisen entsyymiseoksen tuottamiseen tarvittavista säätelymekanismeista tiedetään silti vielä melko vähän.

Solun ulkopuolisen pH:n vaikutusta *T. reesei* -sienen geeniekspressioon tutkittiin genominlaajuisella transkriptioanalyysillä. Analyysin tavoitteena oli tunnistaa pH:n muutokseen reagoivia geenejä. Lisäksi tunnetun, pH-säätelystä vastaavan pacC-geenin ortologin roolia *T. reesei*n sellulaasi- ja hemisellulaasigeenien ekspression säätelyssä tutkittiin. Laajaa induktiokoetta yhdistettynä transkriptioanalyysiin hyödynnettiin tutkittaessa eri substraattien vaikutusta CAZy-geenien ekspressioon. Lisäksi transkriptiodataa hyödynnettiin uusien, sellulaasi- ja ksylanaasiaktiivisuuden tuottoon vaikuttavien säätelytekijöiden tunnistamisessa.

Transkriptioprofiloinnin tulosten perusteella pH:n todettiin olevan tärkeä *T. reesei* -sienen geeniekspressioon vaikuttava tekijä. Solunulkopuoleinen pH vaikutti myös useiden sellulaasi- ja hemisellulaasigeenien ekspressioon. Lisäksi saatiin uutta tietoa PacC-ortologin roolista sellulaasi- ja hemisellulaasigeenien ekspressiossa. Profilointi yhdistettynä useiden eri substraattien käyttöön indusoreina sekä CAZy-geenien huolelliseen uudelleenannotointiin paljasti uusien, mahdollisesti erilaisten selluloosaa ja hemiselluloosaa sisältävien materiaalien hajotuksessa mukana olevien geenien ekspressioprofiilit. Lisäksi fylogeneettinen analyysi antoi viitteitä siitä, että funktionaalinen eriytyminen on melko yleistä *T. reesei*n CAZy-entsyymeille ja se näkyy myös entsyymejä koodaavien geenien erilaisena säätelynä. Sellulaasigeenien ekspressiolle ja sellulaasien tuotolle välttämätön transkriptiofaktorigeeni, *ace3*, tunnistettiin profilointidatan avulla. Geenin ylituotto lisäsi sellulaasien ja ksylanaasien tuottoa. Lisäksi profilointidatan avulla tunnistettiin useita muita mahdollisia uusia regulaattoreita, jotka ovat mielenkiintoisia kohteita lisätutkimuksille. Genominlaajuisten metodien käyttö lisäsi ymmärrystä *T. reesei*n genomin organisaatiosta ja sen mahdollisesti tuomista evoluutionaarisista eduista sekä paljasti entsyymien tuottoon mahdollisesti osallistuvia yhteisesti säädeltyjä genomialueita.

ISBN, ISSN

ISBN 978-951-38-8161-0 (nid.)
ISBN 978-951-38-8162-7 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN-L 2242-119X
ISSN 2242-119X (Painettu)
ISSN 2242-1203 (Verkkojulkaisu)

Julkaisuaika

Lokakuu 2014

Kieli

Englanti, suomenkielinen tiivistelmä

Sivumäärä

89 s. + liitt. 102 s.

Projektin nimi

Rahoittajat

Avainsanat

Trichoderma reesei, sellulaasi, hemisellulaasi, geenisäätely, transkriptiofaktorit, transkriptioprofilointi, solunulkoinen pH

Julkaisija

VTT
PL 1000, 02044 VTT, puh. 020 722 111

VTT TECHNOLOGY 65

Transcriptional analysis of *Trichoderma reesei* under...

Transcriptional analysis of *Trichoderma reesei* under conditions inducing cellulase and hemicellulase production, and identification of factors influencing protein production

Enzymes degrading cellulase and hemicellulase polymers are widely used in the industry for different applications. Depletion of fossil fuels together with environmental concerns related to the usage of non-renewable resources has increased the incentive to find alternative sources for petroleum-based fuels and chemicals. Second generation biofuels and chemicals are derived from lignocellulosic biomass and other plant waste materials, the production of which does not compete with food production. Polymers of the cell wall need to be degraded into simple sugars by the coordinated action of several different enzymes. However, utilisation of renewable biomass materials is hindered by the high price of enzymes needed for biomass degradation.

The filamentous fungus *Trichoderma reesei* is widely utilised in the industry especially for the production of cellulose- and hemicellulose-degrading enzymes. This thesis focuses on studying the expression of genes encoding carbohydrate active enzymes (CAZy) and especially the cellulases and hemicellulases of *T. reesei*. The effects of ambient pH and of different biomass substrates on the gene expression were studied by a microarray method. New knowledge was gained on the different expression patterns of CAZy genes in the presence of various inducing substrates. Ambient pH was shown to be an important determinant of gene expression and to affect the expression of several cellulase and hemicellulase genes. The data enabled identification of candidate regulators for cellulase and hemicellulase genes. A regulator named ACEIII was identified as being essential especially for the production of cellulase activity.

ISBN 978-951-38-8161-0 (Soft back ed.)
ISBN 978-951-38-8162-7 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN-L 2242-119X
ISSN 2242-119X (Print)
ISSN 2242-1203 (Online)