Characterization and optimization of *Thermothelomyces thermophilus* as fungal production host

vorgelegt von
M. Sc.
Benedikt Siebecker

an der Fakultät III – Prozesswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
- Dr.- Ing. -

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender:
Prof. Dr. Peter Neubauer (TU Berlin)

Gutachter:
Prof. Dr. Vera Meyer (TU Berlin)
Dr. Stefan Haefner (BASF SE)

Tag der wissenschaftlichen Aussprache: 16. Dezember 2020

Berlin, 2021

Contents

1. Summary/Zusammenfassung

1.1 Summary

Thermothelomyces thermophilus is a thermophilic, filamentous fungus, and a very efficient natural producer of hydrolases, like lignocellulolytic enzymes. To optimize this fungus as a production host for hydrolases and heterologous proteins, the understanding of hydrolase expression and regulation is of great interest. With that the expression of specific hydrolases could either be enhanced for applications in, e.g. biorefineries or lowered to increase yield and purity of heterologously expressed proteins. Therefore, this dissertation aimed to 1) establish and apply molecular biological methods to delete regulators of cellulase expression, 2) establish and perform bioreactor cultivations with subsequent biochemical analyses of the above-mentioned strains, and 3) analyze the transcriptome with a focus on hydrolase expression and regulation.

Molecular biological methods were established and successfully used to delete known regulators of cellulase expression (Clr1, Clr2, Clr4, and Vib1) and a potentially new one (Vib2). The split marker approach was applied to delete the genes *clr2* and *vib2*, whereas *clr4* and *vib1* were deleted via the newly established CRISPR/Cas12a system. Deletion of *clr1* required the simultaneous use of the split marker approach and the CRISPR/Cas12a system.

Bioreactor cultivations in batch or chemostat operation were successfully established. This allows studies on *T. thermophilus* under controlled, industrially relevant conditions.

The transcriptomic analysis revealed a strong expression of genes encoding for predicted lignocellulolytic enzymes including cellulases, hemicellulases, pectinases, and esterases in response to cellulose. Almost all predicted LPMO genes were strongly expressed as well, supporting the application of *T. thermophilus* in biorefineries. In addition, the analysis uncovered highly expressed predicted protease genes and so far uncharacterized potential regulators of lignocellulolytic enzyme expression. These can be deleted or overexpressed to optimize *T. thermophilus* for industrial applications. Furthermore, the role of the *T. thermophilus* orthologs of known regulators of plant biomass degradation was examined in detail, revealing the essential function of the regulators Clr1 and Clr2 for cellulase expression. The expression of *clr2* and therefore the complete lignocellulolytic response to cellulose was observed to be Clr1 dependent. Moreover, the data indicate that the regulator Clr4, which was recently published to be an activator of cellulase expression, functions as a repressor, instead. Here, the deletion was leading to a higher gene expression of predicted lignocellulolytic enzymes and corresponding regulators (e.g. Clr2). Furthermore, the function of the potentially new regulator Vib2 was investigated and revealed a repressing function in the early lignocellulolytic response to cellulose. The deletion mutant of *vib1* showed growth defects on many carbon sources including cellulose on solid medium, whereas cellulose was successfully utilized during submerged cultivation. The transcriptome of that mutant was not analyzed in this study. Finally, the data suggest a strong involvement of genes predicted to be orthologs of other known regulators like Stk12 and McmA/1. Thus, more than 20 targets for future overexpression or deletion experiments were identified within this work.

1.2 Zusammenfassung

Der thermophile, filamentöse Pilz *Thermothelomyces thermophilus* ist ein sehr effizienter natürlicher Produzent von Hydrolasen, wie z.B. lignocellulolytischen Enzymen. Um die Produktion von Hydrolasen und heterologen Proteinen weiter zu optimieren, ist das Verstehen der Expression und Regulation von Hydrolasen von großem Interesse. Dadurch kann die Expression von spezifischen Hydrolasen für Anwendungen in z.B. Bioraffinerien gesteigert, oder mit dem Ziel einer erhöhten Reinheit und Ausbeute von heterolog exprimierten Proteinen, verringert werden. Die Dissertation hatte deshalb zum Ziel: 1) molekularbiologische Methoden zu etablieren und anzuwenden, um Regulatoren der Cellulaseexpression zu deletieren, 2) Bioreaktorkultivierungen zu etablieren, mit den erstellten Stämmen durchzuführen und anschließend biochemisch zu analysieren und 3) das Transkriptom im Hinblick auf Expression und Regulation von Hydrolasen zu analysieren.

Molekularbiologische Methoden wurden etabliert und erfolgreich angewendet, um bekannte Regulatoren der Cellulaseexpression (Clr1, Clr2, Clr4 und Vib1) und einen potenziell neuen (Vib2) zu deletieren. Die Gene *clr2* und *vib2* wurden mittels Split Marker Methode deletiert und *clr4* und *vib1* durch das neu etablierte CRISPR/Cas12a System. Die Deletion von *clr1* erforderte die gleichzeitige Nutzung der Split Marker Methode und des CRISPR/Cas12a Systems.

Protokolle für Bioreaktorkultivierungen mit *T. thermophilus* (Batch und Chemostat) wurden erfolgreich etabliert und ermöglichen zukünftige Studien unter kontrollierten, industriell relevanten Bedingungen.

Die Transkriptomanalyse zeigte eine starke Expression vorhergesagter lignocellulolytischer Enzyme, insbesondere von Cellulasen, Hemicellulasen, Pektinasen und Esterasen als Antwort auf Cellulose. Fast alle vorhergesagten LPMOs wurden ebenfalls stark exprimiert. Demnach ist *T. thermophilus* für die Anwendung in Bioraffinerien geeignet. Zusätzlich offenbarte die Transkriptomanalyse stark exprimierte vorhergesagte Proteasen und bisher unbekannte Regulatoren der Expression von lignocellulolytischen Enzymen, welche deletiert oder überexprimiert werden könnten, um *T. thermophilus* für industrielle Anwendungen zu optimieren. Des Weiteren konnte die Rolle von Orthologen zu bekannten Regulatoren des Abbaus von pflanzlicher Biomasse detailliert untersucht werden, wodurch die essenzielle Funktion von Clr1 und Clr2 für die Cellulaseexpression gezeigt werden konnte. Hierbei ist die Expression von *clr2* und damit die komplette lignocellulolytische Antwort auf Cellulose abhängig von Clr1. Die vorliegenden Daten für Clr4, bekannt als Aktivator der Cellulaseexpression, weisen stattdessen auf eine reprimierende Funktion hin, da eine Deletion zu einer höheren Expression putativer lignocellulolytischer Enzyme und den zugehörigen Regulatoren (z.B. Clr2) als Antwort auf Cellulose führte. Weiterhin konnte die Funktion des potenziell neuen Regulators Vib2 untersucht werden, welcher für eine Repression der frühen lignocellulolytischen Antwort auf Cellulose sorgt. Die Deletionsmutante von *vib1* zeigte Wachstumsdefekte auf vielen Kohlenstoffquellen (einschließlich Cellulose) auf festem Medium, wogegen Cellulose bei flüssiger Kultivierung erfolgreich genutzt werden konnte. Die Transkriptomanalyse dieser Mutante wurde innerhalb dieser Arbeit nicht durchgeführt. Letztlich suggerieren die Daten eine starke Beteiligung von Genen die als Orthologe von anderen bekannten Regulatoren, wie z.B. Stk12 und McmA/1, identifiziert wurden. Somit konnten über 20 Gene für zukünftige Überexpressions- und Deletionsexperimente innerhalb dieser Arbeit identifiziert werden.

2. List of Abbreviations

AA	auxiliary activities
ABTS	2,2′-azino-bis-(3-ethylbenzothiazoline-6-sulphonic acid)
amp	ampicillin
ampR	ampicillin resistance
An	*Aspergillus nidulans*
APS	ammonium persulfate
ATP	adenosine triphosphate
BG	beta glucosidase
bp	base pairs
BP	biological process
BSA	bovine serum albumin
CAZYs	carbohydrate-active enzymes
CBH	cellobiohydrolase
CBM	carbohydrate binding module
CC	cellular component
CCR	carbon catabolite repression
CE	carbohydrate esterase
CM	complete medium
CPEC	circular polymerase extension cloning
cps	counts per second
CTP	cystidine triphosphate
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DIG	digoxigenin
DME	dimethyl ether
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	desoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
DO	dissolved oxygen
DR	direct repeat
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
EC	enzyme commission
EDTA	ethylenediaminetetraacetic acid
EG	endoglucanase
ex	exponential state
FAA	fluoroacetamide
FDR	false discovery rate
GA	galacturonic acid
GH	glycoside hydrolase
glc	glucose
GO	gene ontology
GOI	gene of interest
GRAS	generally recognized as safe
GT	glycosyl transferase
GTP	guanosine triphosphate
HBT	hydroxybenzotriazole
HC	high cellulase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMF	hydroxymethylfurfural
IEA	international energy agency

Kan	kanamycin
kanR	kanamycin resistance
kb	kilobases
L2FC	log2 fold change
LB	lysogeny broth/luria bertani broth
LC	low cellulase
LPMO	lytic polysaccharide monooxygenase
MCB	master cell bank
MES	4-morpholineethanesulfonic acid
MF	molecular function
MM	minimal medium
MOPS	3-(N-morpholino)propanesulfonic acid
MTG	methanol-to-gasoline
NAD	nicotinamide adenine dinucleotide
NHEJ	non-homologous end joining
NMR	nitrogen metabolite repression
NTG	N-methyl-N-nitro-N-nitrosoguanidine
ORF	open reading frame
padj	adjusted p-value
PAGE	polyacrylamide gel electrophoresis
PCA	principal component analysis
PCR	polymerase chain reaction
PEG	polyethylene glycol
PGA	polygalacturonic acid
PKA	protein kinase A
PL	polysaccharide lyase
PPG	polypropylene glycol
qPCR	quantitative polymerase chain reaction
rcf	relative centrifugal force
RFU	Relative fluorescence units
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
SP	signal peptide
SS	steady state
TAE	tris acetate EDTA
Taq	*Thermus aquaticus*
TEMED	tetramethyl ethylenediamine
Tris	tris(hydroxymethyl)aminomethane
TTP	thymidine triphosphate
UPR	unfolded protein response
UTP	uridine triphosphate
WCB	working cell bank
xyl	xylose

3. Introduction

The following introduction contributes to understanding the background as well as the aims of the presented thesis. First, the definition of biorefineries, which role they play in today´s society, and the potential of filamentous fungi, to contribute to a circular economy is explained. Afterwards, the filamentous fungus *T. thermophilus*, which is investigated in detail within this work and its potential for industry, especially its hydrolase expression, is introduced. As the most important enzyme classes for industry, hydrolases (especially proteases and lignocellulolytic enzymes) of filamentous fungi are presented and discussed. Finally, the regulation of lignocellulolytic enzyme expression in filamentous fungi is discussed in detail to help understanding the potential for industrial applications in biorefineries or heterologous enzyme production, especially in the filamentous fungus *T. thermophilus*.

3.1 Biorefineries: toward a circular economy with the help of fungal biotechnology

3.1.1 Current obstacles and possible solutions toward a circular economy

Today's society still depends heavily on the utilization of fossil resources like crude oil to produce fuels or chemicals products. Fossil resources are formed by natural processes, but this process takes millions of years. Therefore, they are much faster depleted than new ones are generated. In contrast to this, renewable resources like solar energy, wind energy, and agricultural products cannot be depleted, or their regeneration only takes short periods of time. Despite the undeniable fact of climate change caused by the continuous and irreversible release of carbon dioxide from fossil resources, 80 % of the global energy usage still depend on the consumption of fossil fuels. In the energy sector, wind and solar energy are a good possibility to reduce the use of fossil resources to create clean energy (positive carbon balance), but they cannot replace fuels, since energy efficiency of liquid fuels is much higher and they are needed for transportation as well. To overcome this problem efficient alternative solutions have to be found using raw materials with a neutral or even negative net carbon (emitted-consumed) (Seh et al. 2017; Alalwan et al. 2019). The key for this will be found in biorefineries producing biofuels and precursors for chemical products based on renewable resources. In principle there are three types of biorefineries based on the utilized substrate: sugar and starch based biorefineries, lipid or triglyceride biorefineries, and lignocellulosic biorefineries (Cherubini et al. 2009; Konwar et al. 2018). Between the years 2008 and 2017 already more than 2 % increase in overall bio-based share in the product value of chemical products as well as more than 4 % bio-based share in organic chemical products could be observed in the European Union, showing the increase of the proportion of renewable resources used in industry (Porc et al. 2020). This development is also supported not only by the European Union but also by many international initiatives, programs, and agreements like the International Energy Agency (IEA) Bioenergy (annual budget 2010 >2 MUS$), especially Task 42: Biorefining and Task 43: Biomass feedstocks for energy markets, where experts from the whole world work together toward the common goal of a commercialization of bio-based industries (IEA Bioenergy; Van Ree 2017). In November 2017, 118 lipid based biorefineries that mainly produce biodiesel (64) or oleochemicals (54), 63 sugar and

starch based biorefineries that mainly produce bioethanol, and 30 lignocellulosic biorefineries that also mainly produce bioethanol, could be mapped across Europe (Bio based Industries Consortium). In those biorefineries the above-mentioned substrates are subsequently converted into marketable products via biochemical, thermal, and chemical processes. One major problem for biorefineries is the substrate used. In many cases the same raw materials used for production can serve as human nutrition or animal feed. According to this, biorefineries are furthermore classified as first, second, third, and fourth generation biorefineries. First generation biorefineries are working with raw materials like sugar beet and starch crops that are in competition with food and feed industry, therefore rising ethical, environmental, and political concerns. Second generation biorefineries are working with raw materials gained from agricultural or industrial biomass waste. Third generation biorefineries are working with aquatic feedstock like algae. Finally, fourth generation biofuels are working with bioengineered organisms like algae or crops that consume more CO_2 than they emit when they are used as a substrate. The second to fourth generation biorefineries are therefore not in competition with food and feed industry and are the most promising biorefineries for the future. The main reason why substrates for first generation biorefineries are preferred is that those are up to now much easier and cheaper to process in an industrial scale compared to plant waste materials or wood (Cherubini 2010; Moncada et al. 2014; Rulli et al. 2016; Bajracharya et al. 2017; Gambelli et al. 2017; Bhatia et al. 2017; Alalwan et al. 2019). The structure and composition of plant biomass will be discussed in 3.3.3. Nevertheless, this structure mediates a high hydrolytic stability as well as structural robustness and is therefore, difficult to be broken down (Sun and Cheng 2002; Payne et al. 2015; Ponnusamy et al. 2018). However, plant biomass, especially lignocellulose, is the most abundantly available raw material on earth as well as very likely the only viable alternative to fossil resources to produce biofuels and chemicals. Besides, together with fossil resources, it is the only carbon-rich material source available on earth, emphasizing it´s huge potential as a raw material for biorefineries (Cherubini 2010). In modern lignocellulosic biorefineries there are two different possibilities to convert biomass into products: thermochemical, or chemical together with biochemical processes (Cherubini et al. 2009; Konwar et al. 2018).

3.1.2 Thermochemical lignocellulosic biorefineries

In thermochemical biorefineries several methods for the treatment of lignocellulosic biomass are available: direct combustion, gasification, pyrolysis, and torrefaction, each used according to the desired product (Figure 3.1). Direct combustion of the biomass develops heat that can be used to generate electricity. Torrefaction converts biomass into a carbon enriched product (biocoal). Direct combustion and torrefaction are not a subject of present discussion, since in most biorefineries focus lies on the usage of lignocellulose for biofuel production. Gasification converts the lignocellulosic biomass into synthesis gas via exposure to high temperatures (800-900 °C) and limited oxygen supply. The resulting synthesis gas can be cleaned and used to produce dimethyl ether, methanol, and synthetic hydrocarbons (fuels) via different processes like Fisher Tropsch synthesis or the methanol-to-gasoline process. Furthermore, the clean synthesis gas can be used as fuel for gas engines to produce power at high efficiency. Synthesis gas can also be used to produce liquid fuels like heating oil, diesel gasoline, and many more using the so-called biomass to liquid technology, which is based on Fisher Tropsch

synthesis. A major problem of these processes is the poor quality (impurity) of the extracted synthesis gas which requires complicated cleaning steps and results in a low heating value of the gas. Furthermore, high investment and operation costs as well as the possibility of damaged equipment by tar and methane formation are a drawback. This makes it economically more attractive to use cheaper coal and natural gas for the production of synthesis gas (Bridgwater 1995; Huber et al. 2006; Maity 2015; Mikkola et al. 2015; Konwar et al. 2018). Some companies in Sweden like Chemrec AB and VärmlandsMethanol AB have already demonstrated successful production of methanol and dimethyl ether using the gasification process with biomass as substrate (Mikkola et al. 2015; Konwar et al. 2018).

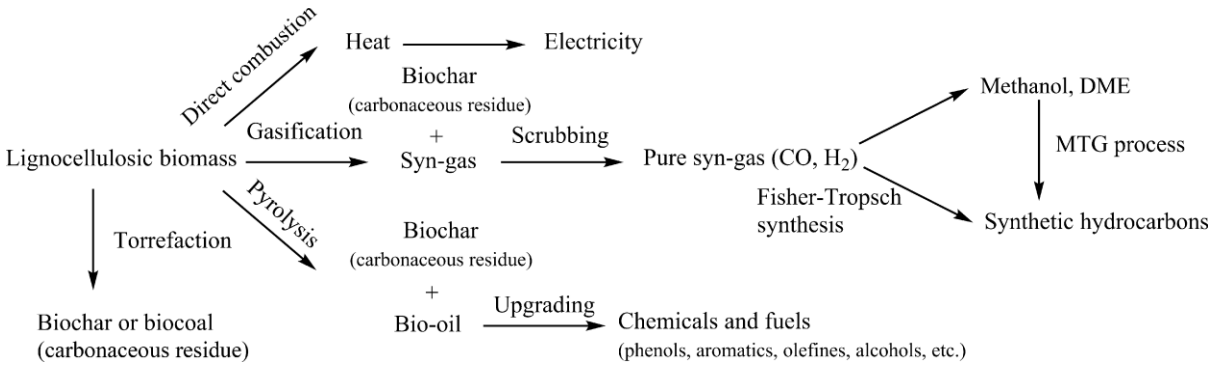

Figure 3.1: Processing of lignocellulosic biomass via thermochemical processes. DME= dimethyl ether, MTG= methanol-to-gasoline. Figure taken from Konwar et al. 2018 with friendly permission from Elsevier.

In the process of pyrolysis lignocellulosic biomass is heated to about 500 °C with a high heating rate in the absence of oxygen, producing bio-oil together with biochar (solid carbonaceous residue) (Zhang et al. 2007; Laird et al. 2009; Konwar et al. 2018). Bio-oil is a complex mixture of more than 300 compounds and can be used for the production of chemicals and hydrocarbons (Laird et al. 2009). Its exact physiochemical properties like high moisture content, oxygenated compounds, high viscosity, and acidic character (pH ~2-3) depend on the type of biomass used as well as on pyrolysis conditions. However, it´s complex composition makes it impossible to extract pure compounds via distillation or solvent extraction. Conversion of the bio-oil into hydrocarbons is possible via steam reforming, hydrodeoxygenation and zeolite catalysts. Advantages of pyrolysis and bio-oil are the requirement of less complicated process equipment, the usage of lower temperatures compared to the gasification process, and therefore lower investment costs. Furthermore, bio-oil has a higher heating value than the used biomass and can easily be stored and transported compared to synthesis gas. Disadvantages of bio-oil are mainly due to its complex nature, and the associated separation problems as well as difficulties in long term storage due to its low oxidative stability. Possibilities to increase the stability and concentration of specific compounds are chemical treatments like dehydration and hydrogenation. Up to now, bio-oils are only used in heavy duty diesel engines for firing boilers and running turbines (Czernik and Bridgwater 2004; Tang et al. 2009; Ma et al. 2015; Konwar et al. 2018). Biochar, the second product that is produced in the pyrolysis process besides bio-oil, can be used as an energy carrier (biocoal), soil amendment, and as precursor for adsorbents, catalysts, catalyst supports and capacitators (Cho et al. 2016; Konwar et al. 2018). Examples for companies that invested in thermochemical lignocellulosic biorefineries are Shell in the Netherlands and the United Kingdom (biomass to ethanol plant using

gasification) and Ensyn in the United States (biomass to fuels and chemicals using fast pyrolysis) (Wertz and Bédué 2013). Further examples can be found in Wertz and Bédué 2013.

3.1.3 Chemical and biochemical lignocellulosic biorefineries

The second possibility to convert lignocellulosic biomass into biofuels and chemicals is via chemical and biochemical processes (Figure 3.2). Due to the already described highly recalcitrant structure of lignocellulose a pretreatment prior to hydrolysis is necessary. These treatment steps include, for example, acid or alkali treatment and mechanical or organosolvent processes. The best way of pretreatment depends on the composition and properties of the used substrate. Different substrates have different contents of cellulose, lignin, and hemicellulose making the optimization of this step highly variable and laborious. The pretreatment process is very cost intensive and harsh pretreatment has negative consequences like the generation of artifacts as well as inhibitors of the following processes (Zheng et al. 2014; Konwar et al. 2018). Therefore, several attempts were made to develop more gentle pretreatment processes via biocatalysts from glycosyl hydrolase family 7, laccases, expansins, swollenins, or loosenins (Cosgrove 2000; Cosgrove et al. 2002; Brotman et al. 2008; Kern et al. 2013; Payne et al. 2015; Pollegioni et al. 2015; Lange 2017).

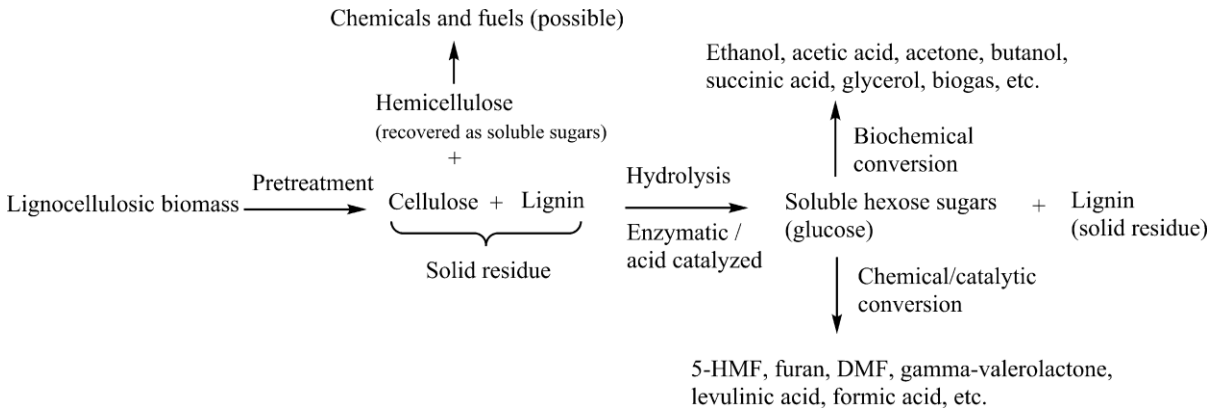

Figure 3.2: Processing of lignocellulosic biomass via biochemical and chemical processes. HMF= hydroxymethylfurfural, DMF= dimethylformamide. Figure taken from Konwar et al. 2018 with friendly permission from Elsevier.

After pretreatment, the biomass gets selectively broken down into its single components, first into cellulose, hemicellulose, and lignin and afterwards into monomeric sugars like glucose and xylose via enzymatic hydrolysis (biochemical) or acid hydrolysis (chemical). After detoxification and purification, those are converted to chemicals and biofuels via different biochemical and chemical processes. Through chemical conversion it is possible to produce, for example, hydroxymethylfurfural, furan, dimethylformamide, levulinic acid, gamma-valerolactone, and other chemical compounds. Biochemical conversion enables the production of ethanol, acetone, glycerol, acetic acid, butanol, biogas, succinic acid, and others from sugars (Alonso et al. 2010; Hasunuma et al. 2013; Climent et al. 2014; Besson et al. 2014; Li et al. 2016; Konwar et al. 2018). Due to the high costs of pretreatment and the enzymatic hydrolysis steps, commercialization of those types of biochemical lignocellulosic biorefineries is limited. Some examples for companies that invested in biochemical and chemical lignocellulosic biorefineries

are DuPont in the United States (lignocellulose to ethanol plant), Beta Renewables in Italy, and Indian River BioEnergy Center in the United states (both cellulosic ethanol plants) as well as GFBiochemicals in Italy (biomass to levulinic acid plant) (Wertz and Bédué 2013; Konwar et al. 2018). Further examples can be found in Wertz and Bédué 2013.

3.1.4 Potential of filamentous fungi for lignocellulosic biorefineries

Filamentous fungi are the optimal solution to the above-mentioned hurdles and disadvantages of existing lignocellulosic biorefineries due to their natural live style. Filamentous fungi are heterotrophic organisms which use lignocellulosic plant biomass as a carbon source for nutrition in nature. To break down the biomass they secrete a huge variety of enzymes, called “CAZYs (Carbohydrate-Active enZYmes)”. With the help of those enzymes, complex carbon sources like lignin or cellulose are subsequently converted into monosaccharides and transported into the cells for the generation of energy (Lange 2017; Meyer et al. 2020). Another advantage of filamentous fungi is their ability to selectively transform complex substrates into the desired product, facilitating cost intensive cleaning steps. Most of the modern lignocellulosic biorefineries using lignocellulose as substrate, like the above-mentioned biorefinery Beta Renewables, are already involving fungal enzymes for enzymatic breakdown and the production of the desired product like biofuels via fermentation (Lange 2017; Konwar et al. 2018; Meyer et al. 2020). However, integrated biorefineries are not yet suitable for commercialization due to the lack and effectiveness of possibilities that enable effective degradation of the provided substrate and the production of the desired product at the same time. Yet, the costs for the cultivation of filamentous fungi that are producing the enzymes needed for hydrolysis are relatively low. Therefore, this problem could be solved by finding or engineering fungal strains that efficiently break down the substrate biomass and simultaneously produce the desired product (e.g. ethanol). Another possibility could be the development of a co-cultivation system for two different types of fungi, one for the degradation of the substrate and one to generate the desired product. Finally, the on-site production of enzymes via fungi could also help to overcome the above-mentioned problem: the enzymes that are required for the breakdown of the substrate could be produced in a separate tank and the entire content of this tank, after inactivation or separation of the fungus, could be used to start hydrolysis in the main tank. Some problems that are so far hindering the above-mentioned solutions are intellectual property and technology protection issues, the restrictions on the use of genetically modified organisms in industry as well as suboptimized enzyme mixtures used for the enzymatic breakdown. Furthermore, the separation or inactivation of the enzyme producing fungus from the culture broth, containing the enzymes used for hydrolysis, without affecting the efficiency of those enzymes is difficult (Bech and Herbst 2015; Lange 2017).

In summary, the search for highly efficient natural enzyme producer strains, the optimization of those producer strains to grow on the selected substrate, and subsequently the ability of these strains to highly perform in an integrated lignocellulosic biorefinery, is of great interest. One of those strains is *T. thermophilus*, a highly efficient natural secretor of especially cellulases and hemicellulases. In industry, hypersecreting *T. thermophilus* strains were developed that produce 100 to 120 g/L of cellulases (Visser et al. 2011; Huuskonen 2020). The fact that this fungus is a thermophilic fungus is of special interest for lignocellulosic degradation processes, since the secreted enzymes are thermostable

(up to 85-90 °C) and allow for high process temperatures, which in turn reduce viscosity and hence increase the solubility of lignocellulosic biomass (Blumer-Schuette et al. 2014; Berezina et al. 2017; Meyer et al. 2020). Other advantages of thermostable enzymes are a higher stability, which allows elongated hydrolysis times, as well as a higher specific activity, which decreases the amount of enzyme needed. Those advantages are improving the overall performance of the enzymatic hydrolysis and therefore reduce the effective costs (Viikari et al. 2007). Furthermore, strains of *T. thermophilus* were already developed that are able to produce commodity chemicals like fumarate, succinate and malate directly from crystalline cellulose (~200 g/L malate), cellobiose (up to 11.3 g/L ethanol), and corncob (~ 100 g/L malate) (Li et al. 2019; Li et al. 2020a). Consequently, the potential of *T. thermophilus* for the development of integrated lignocellulosic biorefineries is huge. In the following chapter *T. thermophilus* will be described in more detail.

3.2 *T. thermophilus* as a fungal production host

T. thermophilus belongs to the ascomycota and is a heterothallic, anamorph, haploid, thermophilic fungus generating hyaline asexual conidia with an ornamented surface that turn rapidly into cinnamon-brown and can be isolated mainly from soil and compost (Van Oorschot 1977). The macroscopic and microscopic morphology of *T. thermophilus* is shown in Figure 3.3.

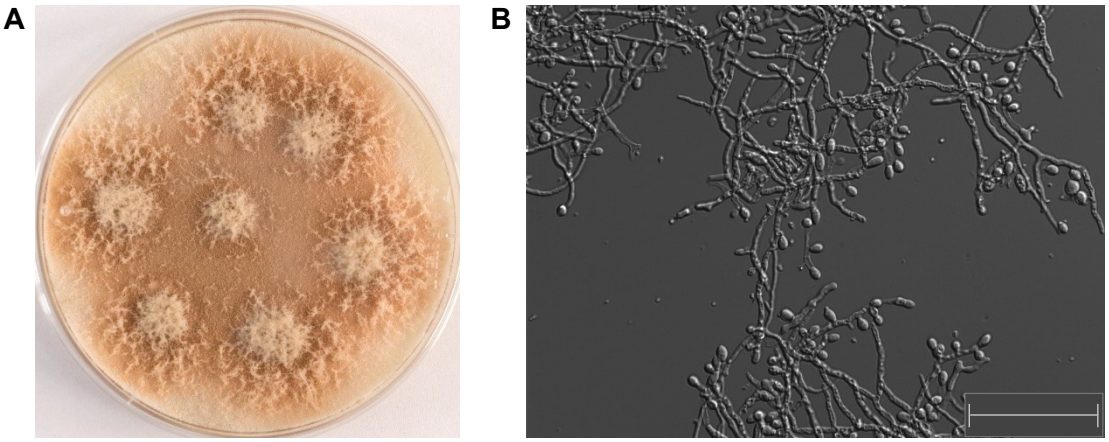

Figure 3.3: Morphology of *T. thermophilus*. (A) Macroscopic view of *T. thermophilus* grown on solid medium for 3 days at 37 °C. (B) Microscopic view of *T. thermophilus* hyphae and spores grown in bioreactor medium for 3 days. Diameter of the petri dish: 9 cm. Scale bar = 50 µm.

The fungus was first described as *Sporotrichum thermophilum*/*thermophile* in 1963 (Apinis 1963) and has been reclassified three times since then: in 1974 to *Chrysosporium thermophilum*/*thermophile* (Von Klopotek 1974)*, in* 1977 to *Myceliophthora thermophila* (Van Oorschot 1977) and 2015 to the current genus *Thermothelomyces thermophila*/*thermophilus* (Marin-Felix et al. 2015). *T. thermophilus* belongs to the order Sordariales and is therefore closely related to fungi of the genus *Neurospora* (Marin-Felix et al. 2015). Nevertheless, in today´s literature the obsolete classification *Myceliophthora thermophila* is still commonly used. The fungus grows from pH 4.5 to pH 7.0 and 25 °C to 55 °C, whereas the optimum growth temperature ranges from 45 to 50 °C (Van Oorschot 1977; Emalfarb et al. 1998; Maheshwari et al. 2000; Visser et al. 2011).

The potential of this strain to be used as a production strain was discovered in the early 1990s, when fungal isolates were screened for the production of neutral cellulases for applications in industrial textile treatment. This resulted in an isolate, called C1, with highly efficient natural secretion of especially cellulases as well as hemicellulases, which was found in alkaline soil in a forest in eastern Russia. C1 was first classified as *Chrysosporium lucknowense*, but after further molecular studies it was concluded that C1 is a *Myceliophthora thermophila* isolate. In the years after, the isolate was developed into Dyadic International Inc.´s proprietary fungal enzyme production platform. In the following the basic C1 strain development by Dyadic is described. The strain development started with random mutagenesis via UV light or N-methyl-N-nitro-N-nitrosoguanidine (NTG) resulting in a high cellulase producing strain (called HC). Due to several random mutagenesis steps the HC strain also acquired an additional low viscosity phenotype. The viscosity of liquid cultures of the HC strain are about 50 fold lower compared to the parental strain due to a distinct, fragmented mycelial morphology, called propagules (Verdoes et al. 2007; Visser et al. 2011). This phenotype has several big advantages, e.g. better transfer of oxygen and nutrients and low energy input for mixing. Mixing is especially important, because a heterogenous broth can severely impact protein production in industrial scale fermentations (Visser et al. 2011). Proteases are causing degradations of proteins of interest during fermentations, therefore reducing their activity enhances total protein yield. Consequently, the HC strain was further optimized by introducing additional mutations. A protease deficient strain was developed via UV mutagenesis, reducing the protease activity at pH 5,5 to 11 % compared to the HC strain. An additional targeted deletion of the alkaline protease *alp1* that was identified as a remaining protease with significant activity, reduced total protease activity at pH 5,5 even further to 6 % compared to the HC strain. The resulting high cellulase, low protease strain was shown to produce up to 100 g/L of extracellular protein. On the other hand, high expression of proteins that are not of interest has an influence on yield as well as the purity of proteins of interest. Therefore, reducing the high natural expression of cellulases in C1 is beneficial for the production of heterologous proteins. With that aim, the HC strain was exposed to UV-mutagenesis and screened for low cellulase activity. After obtaining a low cellulase strain (LC), the same steps as for the HC strain were conducted to create a protease deficient strain. The resulting low cellulase, low protease strain showed a cellulase activity of only 1 to 2 % compared to the HC strain (Visser et al. 2011). The above described basic C1 strain lineage is illustrated in Figure 3.4.

These strains serve as a basis for the C1 technology, which was successfully licensed, optimized for further targeted modifications, and transferred to third parties like Codexis, BASF, Abengoa Bioenergy, and others (Dyadic International website. C1 expression system). Those companies were further optimizing the C1 technology according to their needs especially with the availability of the C1 genome sequence in 2005 (Huuskonen 2020). Examples for those optimizations will be given in the respective chapters. The first commercially available product that was produced with the C1 technology was CeluStar CL which was granted GRAS status in 2009 (Dyadic International website. C1 cellulase enzyme) The highest level of enzymes produced was above 120 g/L and the highest production of an individual recombinant enzyme 80 g/L (Huuskonen 2020). In 2015, the C1 technology was sold to DuPont with co-exclusive rights for use in human and animal pharmaceutical applications (Dyadic International website. C1 expression system). Since 2016, Dyadic is collaborating with the VTT research center to optimize the C1 technology for protein production (Dyadic International website. C1 expression

system). During this collaboration, strains with deletions of up to 13 proteases were created, which improved the stability of target proteins like the ZAPI antigen (up to 1.8 g/L), monoclonal antibodies (up to 24.5 g/L) and Fab fragments (up to 14.5 g/L). Another project with VTT focusses on the optimization of glycosylation of therapeutic proteins to increase proportions of the human glycoforms GO, G2, FGO, and FG2 (Huuskonen 2020).

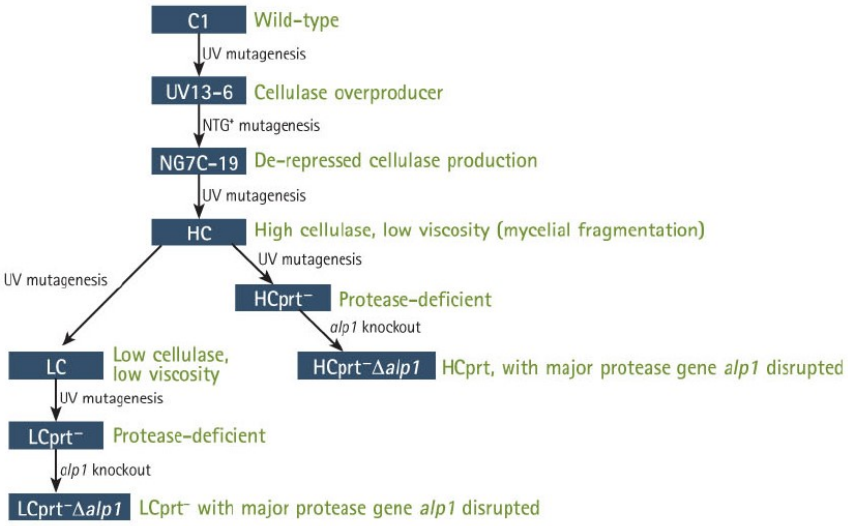

Figure 3.4 Basic C1 strain lineage. HC= high cellulase, LC= low cellulase, prt= protease. Figure taken from Visser et al. 2011 with friendly permission of Mary Ann Liebert, Inc.

While in industry work is continued in the already improved and patented C1 strain, current academic research is mainly focusing on work with the corresponding wildtype strain: *T. thermophilus* ATCC 42464, whose genome was successfully sequenced and made publicly available in 2011 (Berka et al. 2011). Initial research focused on the secreted thermostable enzymes like cellulases, laccases, xylanases, and phytases and their heterologous expression in other host organisms for applications in industry. The advantages of thermostable enzymes have already been discussed in 3.1.4. An overview of enzymes of commercial interest can be found in Singh 2014. After the genomic sequence of *T. thermophilus* was available, tools for targeted genetic and metabolic engineering were urgently needed, therefore transformation systems with different markers like *hygB* and *amdS* based on either *Agrobacterium tumefaciens* (Xu et al. 2015) or protoplasts (Yang et al. 2015; Wang et al. 2015; Kwon et al. 2019) were developed and improved. Recently also systems for genome editing via CRISPR/Cas9 (Liu et al. 2017) or CRISPR/Cas12a (Kwon et al. 2019) were developed, enabling precise as well as high throughput genetic and metabolic engineering of *T. thermophilus*. In the last 10 years mostly studies, focusing on the investigation of (ligno-) cellulosic degradation on cellular and genomic levels, biochemical characterization of enzymes, metabolic engineering for the production, and the improvement of the production of enzymes and chemicals have been performed (Berka et al. 2011; Topakas et al. 2012; Karnaouri et al. 2014b; Karnaouri et al. 2014a; Kolbusz et al. 2014; Volkov et al. 2015; Wang et al. 2018; Li et al. 2019; Liu et al. 2019a; Li et al. 2020a; Li et al. 2020b). Publications that are of special interest for this work will be discussed in more detail in the following chapters.

3.3 Important hydrolases in filamentous fungi

3.3.1 Hydrolases

Hydrolases are enzymes, that are capable of breaking chemical bonds by using water and belong to class 3 according to the EC (enzyme commission) classification. According to the chemical bonds they cleave, those enzymes are further divided into 13 different subclasses. Examples for those are the subclasses 3.1. esterases (cleavage of ester bonds), 3.2 glycosylases (cleavage of glycoside bonds), and 3.4 peptidases (cleavage of peptide bonds). Because several variations of those chemical bonds are existing, e.g. thiolester bonds or phosphoric monoester bonds, the enzymes are further subdivided into sub-subclasses (ExPASy Bioinformatics Resource Portal). In general, hydrolases are very important for degradative processes like the energy metabolism. Here, they contribute to the breakdown of energy storage molecules into smaller molecules that can be transported into the cell and subsequently used as an energy source for nutrition. This chapter will focus mainly on hydrolases of filamentous fungi important for industry: proteases, which belong to subclass 3.4 (peptidases) and to subclass 3.2 (glycosylases).

3.3.2 Proteases

As mentioned above, proteases are enzymes that are cleaving peptide bonds within proteins via hydrolysis. Due to their huge diversity in structure and mode of action they cannot be classified with the general system of enzymatic nomenclature (Martínez-Medina et al. 2019). Nevertheless, the most prominent classification is based on their catalytic residue. According to this, seven general groups exist: serine proteases (serine alcohol), threonine proteases (threonine secondary alcohol), cysteine proteases (cysteine thiol), glutamic proteases (glutamate carboxylic acid), metalloproteases (metal, mostly zinc), aspartic proteases (aspartate carboxylic acid), and asparagine proteases (asparagine, elimination reaction without water and therefore not classified as a hydrolase) (Oda 2012; Rawlings and Bateman 2019). Four of those groups (serine-, cysteine-, aspartic-, and metalloproteases) have been proposed already in 1960 (Hartley 1960) while threonine proteases were first described in 1995 (Löwe et al. 1995), glutamic proteases in 2004 (Fujinaga et al. 2004), and asparagine proteases in 2011 (Rawlings et al. 2011). All of those proteases are either capable of hydrolyzing proteins to peptides (limited proteolysis) or to their constituent amino acids (unlimited proteolysis) (Martínez-Medina et al. 2019). Proteases have various functions within a cell, especially in the regulation of pathways like, e.g. cell differentiation, cell division, blood-clotting cascade, stem cell mobilization, cellular senescence, and apoptosis (López-Otín and Bond 2008; Paschkowsky et al. 2019). Due to these functions, research on protease mechanisms is very important in medical research of human diseases like cancer, cardiovascular diseases, neurodegenerative diseases, infection diseases, and inflammatory diseases. Protease inhibitors are already used in the treatment of some of the above-mentioned diseases like, e.g. HIV, hepatitis C, hypertension, cancer and blood coagulation (Drag and Salvesen 2010; Opar 2010; Chakraborti et al. 2017; Paschkowsky et al. 2019).

Besides the applications and research in medicine, research on proteases is also of great interest for industry. With 60 % of the industrial enzymes market, they are the largest group of commercially available enzymes worldwide due to their broad range of applications in food and beverage industry (e.g. coagulation of milk for cheese production), animal feed industry (to increase amino acid and peptide uptake from feed), cleaning products (to remove protein derived stains), pharmacy (see above), and cosmetics (anti-aging) (Li et al. 2013; Martínez-Medina et al. 2019). The field of application of those proteases depends on their substrate specificity, catalytic mechanism, optimum pH, and temperature as well as their stability. Proteases can be gained from plant (e.g. papain, bromelain), animal (e.g. trypsin, pepsin), or microbial resources. Because the amount and availability of plant and animal material is influenced by many factors and extraction processes are long and expensive, microbial enzymes are preferred. Advantages of microbial compared to plant or animal production are a fast production, uniform product quality with high stability, and simple nutritional requirements. In addition, microbes and enzymes can be optimized easily and fast in comparison to plants or animals by genetic engineering, which enables increasing production yield and enzyme activities. Microorganisms commonly used for the production of proteases are bacteria (mainly *Bacillus* species) and filamentous fungi (mainly *Aspergillu*s species). Due to their strong natural secretion of mainly hydrolases originating from their lifestyle combined with the above-mentioned advantages, filamentous fungi are perfectly suited as production hosts for proteases (Martínez-Medina et al. 2019). Examples and overviews of some commercially available proteases, mainly obtained via production with filamentous fungi and bacteria can be found in Guadix et al. 2000; Li et al. 2013 and Martínez-Medina et al. 2019. Research on the regulation of protease expression has two aims: on the one hand to enhance protease expression for industrial applications and on the other hand to reduce or even eliminate protease expression. The advantages of reducing or eliminating protease activities have been already discussed in chapter 3.2. Some examples are the enhancement of cellulase production (six-fold increase in cellulase activity compared to the parental strain) in *Trichoderma reesei* via deletion of multiple protease genes (Qian et al. 2019), the enhancement of thaumatin production (up to 125 % higher yield compared to the parental strain) in *Aspergillus awamori* via silencing the aspergillopepsin B *(pepB*) gene (Moralejo et al. 2002) or the enhancement of the production of heterologous bovine chymosin and human lysozyme (up to 35 % higher compared to the parental strain) in *Aspergillus oryzae* via a ten-protease gene disruptant (Yoon et al. 2011). Besides academic research, various existing patents are stressing the importance and effectiveness of the inactivation or deletion of proteases for the industrial production of homologous and heterologous proteins. Some examples of those patents are the inactivation of *apsB* and *cpsA* in filamentous fungi (Wang 2009), the deletion of aspergillopepsin A (*pepA*) in filamentous fungi (Berka et al. 2003), and the inhibition of peptidase C (*pepC*) in *Aspergilli* (Lehmbeck and Udagawa 2008) to improve protein production.

In *T. thermophilus* research on proteases is performed as well. This was already described in detail in chapter 3.2 for the industrial isolate C1 (Visser et al. 2011). Furthermore, a patent covering amongst others the use of multiple protease deficient *T. thermophilus* strains was released (Landowski et al. 2013). Publications on protease deletion strains of *T. thermophilus* are limited to one article. There, five different proteases (Mt*alp1*, Mt*pep4*, Mt*gp1*, Mt*cpa1*, and Mt*ga1*) were deleted and the resulting strains analyzed for their cellulase production. A significant increase of cellulase productivity was observed for

all single deletion mutants. This effect was even stronger when multiple proteases were deleted (Li et al. 2020b). The investigation of proteases of *T. thermophilus* for industrial applications could also be of interest, especially because of their thermostability. This characteristic is beneficial for applications in laundry detergents or leather processing, where higher temperatures are used (Li et al. 2013). Additionally, conversion rates are faster and catalytic efficiencies are higher at higher temperatures. Up to now most thermostable proteases were obtained by protein engineering or immobilization methods, proteases from thermophilic organisms are the natural choice for exploring the inherent heat stability (Sinha and Khare 2013).

3.3.3 Lignocellulolytic enzymes

Lignocellulose is the major component and basic structure of plant cell walls, its total amount accounts for 30 to 50 % of the plants total dry weight (Zhang et al. 2019). The structure provides stability and strength to the plant cell wall for protection against wind and weather (abiotic factors) and also acts as a barrier against microbial attacks (biotic factors) (Lange 2017; Sharma et al. 2019). Lignocellulose consists of four major components: cellulose (38-50 %), hemicellulose (23-32%) lignin (12-25%) and pectin (usually very low percentages) (Lange 2017; Ponnusamy et al. 2018). The structure of the primary plant cell wall as well as its composition is shown in Figure 3.5. The exact composition of the single polymers will be discussed in the following chapters. Cellulose forms a very recalcitrant structure via stacking of polysaccharide chain layers by hydrogen bonds. These cellulose microfibrils are coated by hemicellulose via covalent bonds. Pectin is crosslinked with hemicellulose and cellulose while lignin (not shown in Figure 3.5 A) encapsulates cellulose, hemicellulose, and pectin like a matrix via covalent and non-covalent linkages. Considering the amount and complexity of the different non-covalent and covalent bonds and the structure of cellulose itself, lignocellulose forms a highly recalcitrant structure (Lange 2017; Sharma et al. 2019; Zhang et al. 2019).

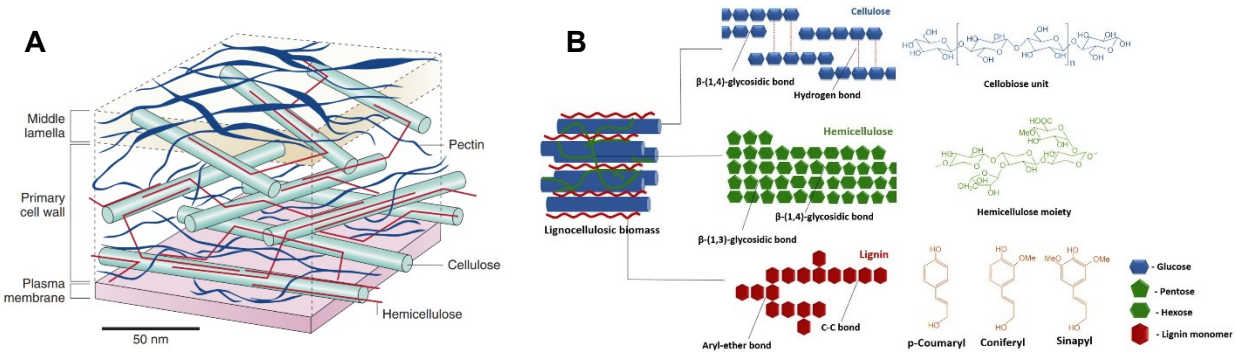

Figure 3.5: Structure and composition of the primary plant cell wall. (A) Schematic structure of the organization of the plant cell wall. Taken from Smith 2001 with friendly permission of Springer Nature. (B) Schematic structure and composition of lignocellulose. Taken from Baruah et al. 2018. Distributed under the terms of the Creative Commons Attribution License (CC BY 4.0).

To break down this structure, the interaction of several enzymes is required which belong to the Carbohydrate Active Enzymes (CAZYs). Besides lignocellulolytic enzymes also enzymes with substrates like, e.g. starch or chitin are amongst the CAZYs. These enzymes are usually required for

cell wall remodeling or starch metabolism and are therefore not required to degrade lignocellulose. However, their regulation is often coupled to the one of lignocellulolytic enzymes, especially for enzymes acting on starch or glucose (carbon catabolite repression). Detailed regulation processes will be discussed in the next chapter.

CAZYs consist of five families of enzymes that degrade, modify or create glycosidic bonds: glycoside hydrolases, glycosyl transferases, polysaccharide lyases, carbohydrate esterases, and enzymes with auxiliary activities (CAZY database. Overview). Glycoside hydrolases (GH) are the most important family of enzymes that are required for the breakdown of lignocellulose. They hydrolyze glycosidic bonds, which are the most prominent bonds in cellulose as well as hemicellulose. They are grouped in 167 families (up to now) according to their amino acid sequence and are continuously updated (CAZY database. Glycoside hydrolases). The counterparts of glycoside hydrolases are glycosyl transferases (GTs). GTs are forming new glycosidic bonds via transfer of sugar moieties from an activated donor to specific acceptor molecules. Therefore, they are not directly involved in polysaccharide degradation, but in their synthesis (CAZY database. Glycosyl transferases). The second family of enzymes that are involved in the degradation of lignocellulosic biomass are carbohydrate esterases (CEs). They assist GHs by catalyzing the de-acylation of substituted saccharides, e.g. in hemicellulose. Up to now they are grouped in 17 families according to their amino acid sequence (CAZY database. Carbohydrate esterases). The third family of enzymes that are involved in the breakdown of lignocellulose are polysaccharide lyases (PLs). They help depolymerizing polysaccharides by cleaving uronic acid containing polysaccharide chains, which are found in pectin. So far they are grouped in 40 families according to their amino acid sequence (CAZY database. Polysaccharide lyases). The fourth and final family of enzymes involved in plant biomass degradation are enzymes with auxiliary activities (AA). This family includes redox enzymes that are acting in conjunction with CAZYs to boost lignocellulolytic activities. Until now they are grouped in 9 families of ligninolytic enzymes and 6 families of lytic polysaccharide monooxygenases (LPMOs) (CAZY database. Auxiliary activities). GH, GT, CE, PL and AA families are frequently polyspecific, meaning the classification into one specific family does not provide information about one specific substrate or product for this family, since several substrates ad products are possible (CAZY database. Overview). Associated with the CAZY families are the carbohydrate binding modules (CBMs). Those are amino acid sequences within a CAZY with a discreet fold allowing carbohydrate-binding. They can affect efficacy and speed of the enzyme activity (Lange 2017). Thus far they are grouped in 86 families according to their amino acid sequence (CAZY database. Carbohydrate binding module).

The exact composition of the different polymers as they can be found in the plant cell wall (Table 3.1) and the enzymes needed for their degradation will be discussed in the next chapters. Due to the aims of this thesis, detailed descriptions will be mainly done for substrates and enzymes that are relevant for this work.

Table 3.1: Components of the plant cell wall. Adapted from Benocci et al. 2017.

polymer type	polymer	monomers
cellulose		D-glucose
hemicellulose	xylan	D-xylose
	glucuronoxylan	D-glucuronic acid, D-xylose
	arabinoglucuronoxylan	L-arabinose, D-xylose
	arabinoxylan	L-arabinose, D-xylose
	galacto(gluco)mannan	D-galactose, D-glucose, D-mannose
	mannan/galactomannan	D-galactose, D-mannose
	xyloglucan	D-xylose, D-glucose, D-fructose, D-galactose
	β(1,3)/(1,4)-glucan	D-glucose
pectin	homogalacturonan	D-galacturonic acid
	xylogalacturonan	D-xylose, D-galacturonic acid
	rhamnogalacturonan I	L-rhamnose, D-galacturonic acid, D-galactose, L-arabinose, ferulic acid, D-glucuronic acid
	rhamnogalacturonan II	L-rhamnose, D-galacturonic acid, D-galactose, L-arabinose, L-fucose, D-glucose, D-manno-octulosonic acid (KDO), D-lyxo-heptulosaric acid (DHA), D-xylose, D-apiose, L-acetic acid
starch	amylose	D-glucose
	amylopectin	D-glucose
lignin		monolignols: ρ-coumaryl alcohol, coniferyl alcohol, sinapyl alcohol

3.3.3.1 Cellulases

Cellulose is a linear polysaccharide that is composed of D-glucose molecules that are linked via β-1.4 glycosidic bonds. It exists as ordered microfibrils in the cell wall and contains crystalline and amorphous regions, the ratio affects its degradation efficiency (Igarashi et al. 2007; Zhang et al. 2019). Degradation is achieved via combination of different enzymes which are grouped together as cellulases: endoglucanases, cellobiohydrolases, β-glucosidases, and lytic polysaccharide monooxygenases. Endoglucanases (EGs) randomly hydrolyze the amorphous part of cellulose (Shrotri et al. 2017). They are widespread among GH families with examples that were described for families 5-9, 12, 44, 45, 48, 51, and 74 (Karnaouri et al. 2014a). Except acting on cellulose, endoglucanases can also act on xylan and mixed β-glucan (Chen and Wang 2017). Cellobiohydrolases (CBHs) are cleaving at the resulting reducing (CBH I, GH 7) and non-reducing (CBH II, GH 6) end groups to form the disaccharide cellobiose which is then hydrolyzed by β-glucosidases (BGs, GH1 and GH3) to glucose. BGs can also act on short (soluble) cello-oligosaccharides (Karnaouri et al. 2014a; Lange 2017). Finally, LPMOs are playing an important role in plant biomass degradation which was already demonstrated for cellulose and soluble cello-oligosaccharides (Vaaje-Kolstad et al. 2010; Forsberg et al. 2011; Isaksen et al. 2014). When combined with other cellulases they have an enhancing cellulolytic effect by cleaving cellulose chains via oxidation of carbons C1 and/or C4 and C6. LPMOs that are known to enhance cellulose degradation belong to the following families: AA9, formerly known as GH61 family, AA10, formerly known as CBM33, and AA16 (CAZY database. AA9/GH61; Forsberg et al. 2011; Phillips et al. 2011; Horn et al. 2012; Hemsworth et al. 2015; Filiatrault-Chastel et al. 2019). Finally, enzymes with a variety of different CBM modules are known (see CAZY database. Carbohydrate binding module for detailed lists), but the most prominent amongst them in filamentous fungi is CBM1, which is known to bind cellulose (CAZY database. Carbohydrate binding module; Lange 2017).

3.3.3.2 Hemicellulases

Hemicellulose is a highly branched polysaccharide that consist of many different monosaccharides like glucose, xylose, arabinose, fructose, mannose, and galactose that are linked to each other via various α and β glycosidic bonds. The different polymers as well as monomers hemicellulose consists of are listed in Table 3.1. The major component of hemicellulose is xylan which consists of a backbone of β-1.4 linked xyloses that can be substituted with arabinofuranosyl, 4-O-methylglucopyranosyl, feruloyl, and acetyl groups, depending on its origin (Shibuya and Iwasaki 1985; Karnaouri et al. 2014a). Other major components of hemicellulose are xyloglucans, mannans, and glucomannans (Ward 2015). Xyloglucan consists of a β-1.4 linked glucose backbone with α-1.6 linked xylose sidechains which are often capped (β-1.2 link) with a galactose residue that is sometimes followed by a fucose residue (α-1.2 link) (Fincher 2016). Mannans are linear polysaccharides that are composed of β-1.4 linked mannose molecules and glucomannans are polysaccharides that consist of a backbone of mannose and glucose molecules that are β-1.4 linked with α-1.6 linked galactose sidechains. The exact composition of hemicellulose and the presence and amounts of side chains always varies depending on the organism (Karnaouri et al. 2014a; Fincher 2016). Given the complexity of hemicellulose, degradation of it requires a bigger set of enzymes compared to cellulose. Xylanases and xylosidases are required to hydrolyze the xylan backbone. Xylanases appear mostly in GH families 10 and 11, and xylosidases in families 3 and 43, respectively. To remove arabinose residues and break down arabinan, arabinases, and arabinofuranosidases with examples in GH families 43, 51, 62, and 93 are required. Mixed link glucanases act on different hemicellulolytic chains containing glucose residues like, e.g. xyloglucan with examples in GH family 16. Degradations of chains with mannose residues are done by mannanases and mannosidases with examples in GH families 5, 26. 76, 38, 47 and 92. Degradations of chains with galactose residues are performed by galactanases and galactosidases with examples in GH families 5, 53, 27, and 2. Note that some of the above-mentioned enzymes can be also involved in pectin degradation, depending on the saccharide residues of the respective chains. Esterases like feruloyl esterases, acetyl esterases, and glycuronoyl esterases also play an important role in the degradation of lignocellulosic material. This includes the cleavage of bonds between hemicellulose backbones and many types of their side chains. Examples can be found in CE 1, 3, 4, 5, 8, 9, 12, 15, and 16 (Karnaouri et al. 2014a). Finally, LPMOs from families AA9 and AA14 are known to be involved in hemicellulose degradation (Agger et al. 2014; Frommhagen et al. 2015; Couturier et al. 2018). Given the huge variety of different enzymes and mechanisms involved in degradation of hemicellulose, this was only a brief summary. Detailed information can be found in the CAZY database or literature like Karnaouri et al. 2014a.

3.3.3.3 Pectinases

Pectin is like hemicellulose a mixture of polymeric compounds. It consists of homogalacturonan, xylogalacturonan, and rhamnogalacturonan I and II as described in Table 3.1. The major components are homogalacturonan and rhamnogalacturonan I (Varga and Samson 2008; Caffall and Mohnen 2009; Wu et al. 2020). Homogalacturonan consists of a linear α-1.4 linked galacturonic acid backbone to which methyl or acetyl residues can be attached. Xylogalacturonan has the same backbone as

homogalacturonan, but with xylose residues attached to the main chain. Rhamnogalacturonan I has a backbone of α-1.4 linked galacturonic acid residues that is interrupted by α-1.2 linked rhamnose residues. Long chains of arabinose (α-1.5 linked), galactose (β-1.4 linked), or a combination of both can be attached to the rhamnose residues. Those side chains can contain feruloyl residues on terminal arabinose or galactose residues. Rhamnogalacturonan II has approximately 30 residues and consists of a homogalacturonan backbone with 4 side chains that are consisting of uncommon sugars like 2-O-methyl-L-fucose and 3-deoxy-D-manno-2-octulosonic acid and many other sugars as listed in Table 3.1. The exact composition of pectin and the presence and amounts of side chains always varies depending on the organism (Varga and Samson 2008). Due to its diverse structure, degradation of pectin involves more enzymes compared to the degradation of cellulose. Involved are polygalacturonases (e.g. GH28), rhamnosidases (e.g. GH78), pectin lyases (e.g. PL 1, 3, 4, 20), and pectin esterases (e.g. CE 8 and 12) (Karnaouri et al. 2014a; Zhang et al. 2019). Depending on the respective saccharide residues that are incorporated, also other enzymes can be involved as mentioned above.

3.3.3.4 Ligninases

Lignin is a highly heterogenous alkyl-aromatic polymer and therefore very different in its structure compared to cellulose, hemicellulose, and pectin. It is composed of a mixture of three hydroxyphenylpropanoid units: p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. The exact composition of lignin varies depending on the organism (Vanholme et al. 2010; Sharma et al. 2019; Zhang et al. 2019). Degradation is achieved via laccases, lignin peroxidases, manganese peroxidases, and versatile peroxidases. Laccases belong to the family AA 1 and the mentioned peroxidases to family AA 2. These enzymes do not react independently, mediators are required (CAZY database. Auxiliary activities; Lange 2017; Zhang et al. 2019).

3.3.4 Lignocellulolytic enzymes in industry

In industry lignocellulolytic enzymes have a broad variety of applications. In food industries they are applied in fruit and vegetable juice making, vegetable oil processing, winemaking, brewing, and baking (Toushik et al. 2017). In fruit and vegetable juice making, for example, cellulases are used for improving extraction methods, clarification, and stabilization while xylanases and pectinases are increasing the yield of juices. In the animal feed industry lignocellulolytic enzymes are used to improve the nutritional value and absorbance of feed. In the textile industry laccases are used for dye decolorization, denim finishing, and cotton bleaching and cellulases are used in laundry detergents to remove plant derived stains. The paper and pulp industry also uses laccases for delignification of lignocellulose as well as biopulping, and biobleaching (Nguyen et al. 2018).

Nevertheless, the most promising application of lignocellulosic enzymes lies within lignocellulosic biorefineries for a sustainable production of bioethanol and commodity chemicals. Detailed processes and the role and potential of filamentous fungi has already been discussed in this context (see chapter 3.1.4), while the other above-mentioned branches of industry are also able to benefit from the potential of filamentous fungi for the production of lignocellulolytic enzymes. Lignocellulolytic enzyme cocktails that are applied in lignocellulosic biorefineries can contain cellulases, LPMOs, and laccases.

Hemicellulases are usually not applied because many bonds that are usually cleaved by those enzymes are already hydrolyzed during pretreatment (Lange 2017). Even though it is known that accessory hemicellulases like xylanases and mannanases beside cellulases and LPMOs have beneficial effects on cellulose degradation by increasing cellulose accessibility (Hu et al. 2015), remaining hemicellulolytic residues after pretreatment are usually degraded via hemicellulolytic side activities of cellulases in the cellulase cocktail (Hu et al. 2013; Chylenski et al. 2017; Chylenski et al. 2019). Degradation of lignin can be performed via laccases only, since the peroxidases have a preference toward coupling of phenoxy radicals *in vitro*, which leads to polymerization rather than depolymerization of lignin samples (Conesa et al. 2002; Sharma et al. 2019). Together with a suitable synthetic mediator like 2,2′-azino-bis-(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) and hydroxybenzotriazole (HBT) laccases can catalyze oxidative degradation of lignin also *in vitro* (Jeon and Chang 2013; Sharma et al. 2019). By doing this, it is possible to delignify plant biomass, leaving behind cellulose and hemicellulose, but economically the pretreatment method, leaving behind lignin as a solid residue, is still favored (Konwar et al. 2018; Sharma et al. 2019). Following pretreatment, degradation of cellulose is the most important process in lignocellulosic biorefineries due to its abundance in biomass and its simple composition. The degradation is achieved, as already described, mainly via endoglucanases, cellobiohydrolases, and β-glucosidases (Chylenski et al. 2019). It has been shown that the addition of LPMOs (especially from family AA9) to cellulase enzyme cocktails containing the above-mentioned cellulases leads to a huge increase in overall efficiency of the enzyme cocktail (relative increase up to 93 %). A detailed overview of some commercial cellulase cocktails and the increase of efficiencies after adding LPMOs can be found in Chylenski et al. 2019. That´s why in most modern cellulase cocktails one or more LMPOs are included (Johansen 2016; Chylenski et al. 2019). Especially via application of LPMOs the costs of cellulolytic enzyme cocktails could be greatly reduced during the past decades, but they are still accounting for 25 % of the costs of cellulosic bioethanol and are therefore the primary bottleneck for successful commercialization of lignocellulose derived fuels and chemicals (Humbird et al. 2011; Johansen 2016; Chylenski et al. 2019). Due to this harnessing the potential of cellulolytic enzymes including LPMOs especially in filamentous fungi is of major interest for industry.

3.3.5 Lignocellulolytic enzymes in *T. thermophilus*

In *T. thermophilus* lignocellulolytic enzymes have also been an important topic of research due to their various possible applications in industry as already mentioned in 3.2. Furthermore, heterologous expression of lignocellulolytic enzymes from *T. thermophilus* like laccases was achieved in, e.g. *Aspergillus oryzae* and *Saccharomyces cerevisiae* (Berka et al. 1997; Bulter et al. 2003). Together with the availability of the genomic sequence, a study coupling cultivation on different carbon sources (glucose, alfalfa, barley) with transcriptomics allowed to gain insight into the regulation and expression of lignocellulolytic enzymes (Berka et al. 2011). This publication is the base for almost all publications about *T. thermophilus* after 2011. The analysis of the genomic sequence with regards to lignocellulolytic enzymes was extended by Karnaouri et al. 2014a. They combined genomic data from Berka et al. 2011 with published enzymatic activities of several isolated and characterized enzymes to categorize lignocellulolytic enzymes (e.g. into different cellulolytic and hemicellulolytic enzymes). According to this

work, *T. thermophilus* possesses 48 cellulases (25 AA9 LPMOs included), 51 hemicellulases, 15 pectinases, and 10 enzymes with auxiliary activities including ligninases. For comparison, *Aspergillus niger* possesses 31 predicted cellulases and 36 predicted hemicellulases (De Souza et al. 2011) and *Neurospora crassa* 23 predicted cellulases and 19 predicted hemicellulases (Sun and Glass 2011). Kolbusz et al. 2014 expanded the analysis of Berka et al. 2011 with the addition of further substrates like oat, triticale, canola, and flax allowing to analyze differences between substrates from monocot and dicot plants in a more detailed manner. More recent publications are investigating lignocellulolytic gene expression on defined carbon sources rather than complex substrates mostly in combination with investigations of new or putative regulators of lignocellulolytic enzymes. Those carbon sources include for example cellobiose (Li et al. 2020a), arabinose, xylose (Dos Santos Gomes et al. 2019), and cellulose (Wang et al. 2018; Xu et al. 2018; Liu et al. 2019a). Regulation of lignocellulolytic enzyme expression will be discussed in detail in the following chapter.

The potential of *T. thermophilus* for the production of lignocellulolytic enzymes especially for lignocellulolytic biorefineries due to its huge amount of lignocellulolytic enzymes and their thermostability (Singh 2014) as well as the advantages of thermostable enzymes has already been discussed. An additional outstanding characteristic of *T. thermophilus* compared to other filamentous fungi is the high amount of LPMOs in its genome. With 25 AA9 family members this is the highest number among filamentous fungi (Berka et al. 2011; Karnaouri et al. 2014a). For comparison, *N. crassa* possesses 14 AA9 enzymes and *A. niger* only 7 (CAZY database. *Neurospora crassa* OR74A CAZYs; CAZY database. *Aspergillus niger* CBS 513.88 CAZYs). The effects of LPMOs have been discussed previously. Nevertheless, one drawback of LPMOs is their requirement of oxygen, which is expensive in large industrial reactors. To achieve this, maintaining high aeration and high stirrer speed is needed (Garcia-Ochoa et al. 2010; Chylenski et al. 2019). Here the C1 low viscosity phenotype is of additional benefit, since oxygen transfer is facilitated due to its low viscosity and therefore, the stirrer speed can be reduced (Visser et al. 2011). Furthermore, one LPMO from C1 (*Mt*LPMO9A) is the only LPMO known yet that can oxidatively cleave both, xylan and cellulose and might therefore be of interest for lignocellulosic biomass biorefineries (Frommhagen et al. 2015).

3.4 Regulators of the expression of lignocellulolytic enzymes

3.4.1 General mechanisms

Regulation of plant biomass degradation requires complex interactions between different pathways and mechanisms, including carbon catabolite repression (CCR), nutrient (inducer) sensing pathways that can be coupled with CCR, direct transcriptional regulation, and post-translational regulation via feedback from the secretory pathway (Huberman et al. 2016; Benocci et al. 2017). Generally, the sensing of inducers, which varies amongst different filamentous fungi, starts a signaling cascade that finally leads to the activation of transcription factors required for the expression of lignocellulolytic enzymes (De Vries and Visser 2001; Benocci et al. 2017). To successfully utilize plant biomass as a carbon source, fungi need to recognize the components (sugars) and adapt their metabolism to be able to convert these into bioproducts and energy. The components that are recognized are defined as inducers, which are small molecules able to enter the fungal cell (Benocci et al. 2017). It is assumed that inducers are produced

via enzymatic hydrolysis of polysaccharides by small amounts of either constitutively produced enzymes, or carbon starvation induced enzymes known as scouting enzymes. This has been observed in various filamentous fungi like *A. niger*, *T. reesei*, and *N. crassa* in response to different lignocellulosic compounds (Kubicek et al. 1993; De Vries et al. 2002; Foreman et al. 2003; Yuan et al. 2008; Tian et al. 2009; Benocci et al. 2017). As already mentioned, inducers vary amongst filamentous fungi. Common inducers for hemicellulases and cellulases are, e.g. xylose and cellobiose (Marui et al. 2002; De Souza et al. 2011; Amore et al. 2013). A summary of known inducers and genes that are regulated via those inducers in *N. crassa*, *T. reesei*, and *A. niger* can be found in Amore et al. 2013. Inducer activated transcription factors that are involved in the regulation of lignocellulolytic enzyme expression mainly belong to the zinc finger family, which is characterized by zinc finger formations in the binding domains. Activators belong mostly to the Zn_2Cys_6 class and repressors primarily to the Cys_2His_2 class (Benocci et al. 2017). Transcription factor based regulation can be further fine-tuned via continuous feedback mechanisms like carbon catabolite repression and feedback from the secretory pathway. This ensures the highest possible energy efficiency and is dependent on the composition of the nutrient to grown on (Huberman et al. 2016; Benocci et al. 2017).

In the following, the most important known regulators and other genes involved in lignocellulose degradation will be discussed in detail. An overview of some regulators involved in plant biomass degradation taken from the latest review about this topic, can be found in Figure 3.6. The strong interaction between pathways for degradation of different polysaccharides leads to blurry boundaries, meaning that regulators mainly known for starch degradation can also be involved in (hemi)-cellulose degradation. Therefore, the classification into the following subsections is based on the main known function of the regulators discussed there. Table 3.2 summarizes regulators, their involvement in plant biomass degradation and the respective references. A detailed description follows in the next chapter.

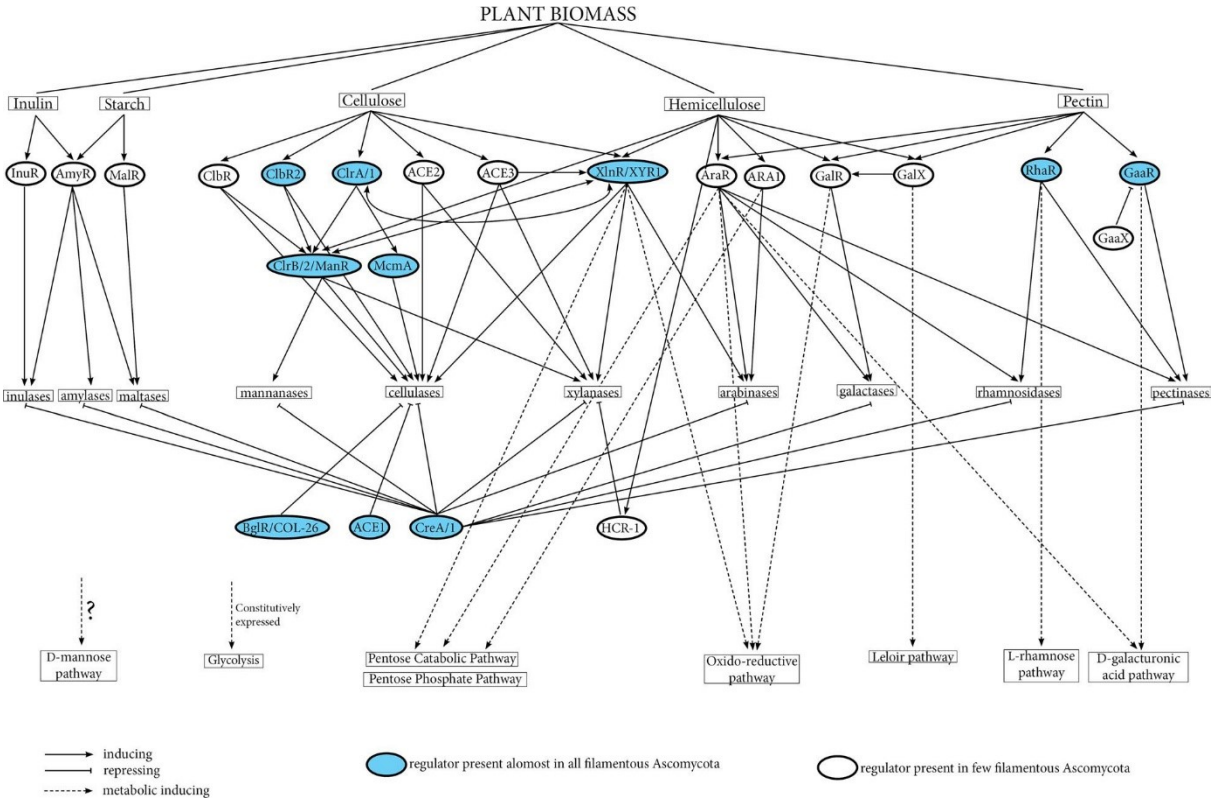

Figure 3.6: Regulators and their role in plant biomass degradation as well as other metabolic processes. Figure taken from Benocci et al. 2017. Distributed under the terms of the Creative Commons Attribution License (CC BY 4.0).

Table 3.2. Regulators and their involvement in plant biomass degradation together with the respective references. The references that are written bold, are the known publications that investigate the function of the respective regulator in *T. thermophilus*. The dash indicates the different names of the same orthologs in different filamentous fungi.

regulator	regulation of/ function	references
Ap3	cellulose degradation	**Liu et al. 2019b**
Clr1/A		Coradetti et al. 2012; Coradetti et al. 2013; Craig et al. 2015; **Haefner et al. 2017a**
Clr2/B/ManR		Ogawa et al. 2013; Coradetti et al. 2013; Yao et al. 2015; Raulo et al. 2016; **Haefner et al. 2017b**
ClrC		Lei et al. 2016
Crz1		Chen et al. 2016
Hcr1		**Liu et al. 2019b**
Hp1/HepA		Zhang et al. 2016
Ire1, Hac1		Huberman et al. 2016; Benz et al. 2014; Fan et al. 2015; Montenegro-Montero et al. 2015
LaeA/1		Karimi-Aghcheh et al. 2013; Fekete et al. 2014
McmA/1		Yamakawa et al. 2013; Tani et al. 2014a; Fujii et al. 2015
Mhr1		**Wang et al. 2018**
Rca1		**Liu et al. 2019b**
Rce1		Cao et al. 2017
Res1		**Liu et al. 2019b**
Sah2		Reilly et al. 2015
Stk12		Lin et al. 2019
Wc1/Blr1, Wc2/Blr2		Gyalai-Korpos et al. 2010
Ace1, Ace2, Ace3	(hemi)-cellulose degradation	Saloheimo et al. 2000; Aro et al. 2003, Aro et al. 2001; Häkkinen et al. 2014

AreA/Nit2, AreB, NmrA/1	(hemi)-cellulose degradation	Arst and Cove 1973; Fu and Marzluf 1990; Berger et al. 2008; Gonçalves et al. 2011; Amore et al. 2013; Conlon et al. 2001; Lamb et al. 2004; Berger et al. 2006; Macios et al. 2012
ClbR, ClbR2, ClbR3		Kunitake et al. 2013; Kunitake et al. 2015; Tani et al. 2014b; Benocci et al. 2017
Clr4		**Liu et al. 2019a**
Hap2,3,4,5		Narendja et al. 1999; Zeilinger et al. 2003
Mhr1		**Wang et al. 2018**
PacC/1		Rossi et al. 2013, De Graaff et al. 1994; Kunitake et al. 2016; He et al. 2014
VeA/Vel1, VelB		Seiboth et al. 2012; Karimi-Aghcheh et al. 2014; Bayram et al. 2008; Benocci et al. 2017
Xyr1/XlnR/Xlr1		Van Peij et al. 1998b; Van Peij et al. 1998a; Rauscher et al. 2006; Tamayo et al. 2008; Sun et al. 2012; Hasper et al. 2000; Stricker et al. 2006; Li et al. 2015; **Wang et al. 2015; Dos Santos Gomes et al. 2019**
Hcr1	hemicellulose degradation	Li et al. 2014
Xpp1		Derntl et al. 2015; Derntl et al. 2017
AraR/1	hemicellulose and pectin degradation	Battaglia et al. 2011; Klaubauf et al. 2016
GalR, GalX		Kowalczyk et al. 2015, Christensen et al. 2011
GaaR	pectin degradation	Alazi et al. 2016
GaaX		Niu et al. 2017
Pdr1		Thieme et al. 2017
Pdr2		Wu et al. 2020
RhaR		Gruben et al. 2014
MalR	starch degradation	Hasegawa et al. 2010; Suzuki et al. 2015
AmyR	starch degradation, glucose sensing	Li et al. 2015; Xiong et al. 2017; Benocci et al. 2017; **Xu et al. 2018**
BglR/Col26	carbon catabolite repression, glucose sensing	Xiong et al. 2014; Huberman et al. 2016; **Lai et al. 2017**
Vib1		Hynes and Kelly 1977; Kelly and Hynes 1977; Lockington and Kelly 2002; Boase and Kelly 2004; Nitta et al. 2012; Xiong et al. 2014
CreA/1	carbon catabolite repression	Orejas et al. 1999; Orejas et al. 2001; Tamayo et al. 2008; Sun and Glass 2011; **Yang et al. 2015; Liu et al. 2017**
CreB, CreC, CreD		Lockington and Kelly 2002; Boase and Kelly 2004; Denton and Kelly 2011; Hunter et al. 2013

3.4.2 Carbon catabolite repression and glucose sensing

Carbon catabolite repression is a mechanism that downregulates genes needed to metabolize less preferred carbon sources, when a more preferred carbon source is available and therefore prevents the production of unnecessarily secreted enzymes. When, for example, glucose is available as a carbon source, most genes that are required for growth on lignocellulosic biomass will be repressed (Ronne 1995; Aro et al. 2005; Brown et al. 2014; Benocci et al. 2017). The main regulator of carbon catabolite repression in filamentous fungi is the conserved transcription factor CreA/1. Studies in *A. nidulans* and *N. crassa* were able to show that a deletion or mutation of *creA/1*, leads to transcription of xylanase genes in the presence of glucose or xylose and an increased production of cellulases in strains grown on cellulose, indicating a repression of those genes via CreA/1 (Orejas et al. 1999; Orejas et al. 2001; Tamayo et al. 2008; Sun and Glass 2011). It was also observed that CreA/1, even though it represses expression of transcription factors, is required for full expression of (hemi)-cellulases under some inducing conditions (Portnoy et al. 2011; Antoniêto et al. 2014). In general, the deletion of *creA/1* causes a derepression of the transcription of (hemi)-cellulolytic genes under repressing as well as inducing

conditions. This includes (hemi)-cellulases as well as transcription factors involved in lignocellulolytic degradation (Huberman et al. 2016; Benocci et al. 2017). In *T. reesei*, for example, the Cre1 mechanism is different: it acts as a switch to turn the transport of inducer or repressor molecules on or off and therefore controls the expression of transcription factors (Ries et al. 2013; Antoniêto et al. 2014). CreA/1 also plays a role in chromatin remodeling by affecting chromatin structure (acetylation, packaging, nucleosome position etc.) under repressing conditions as shown, e.g. for *T. reesei* and *A. nidulans* (García et al. 2004; Ries et al. 2014; Mello-de-Sousa et al. 2015; Mello-de-Sousa et al. 2016). CreA/1 is also involved in many other processes like pH control, penicillin production, and nitrogen as well as amino acid transport or metabolism (Cepeda-García et al. 2014; Bi et al. 2015; Ries et al. 2016). Deletion of genes encoding for CreA/1 can also lead to severe phenotypes including lethality, revealing a central role of CreA/1 that is still not completely understood (Shroff et al. 1997; Nakari-Setälä et al. 2009; Sun and Glass 2011; Fujii et al. 2013; Bi et al. 2015). CreA/1 itself is regulated via SnfA/1, which phosphorylates CreA/1 and therefore controls its import and export into the nucleus. Phosphorylation can also be achieved in a SnfA/1 independent way, stressing that genetic regulation of CreA/1 varies amongst filamentous fungi (Cziferszky et al. 2002; Cziferszky et al. 2003; Brown et al. 2013). Further enzymes that affect CreA/1 stability and proteasomal degradation and therefore its function, are the deubiquitinating CreB/2, WD40 motif protein CreC, and ubiquitin ligase interacting CreD (Lockington and Kelly 2002; Boase and Kelly 2004; Denton and Kelly 2011; Hunter et al. 2013).

Furthermore, the glucose-sensing regulators BglR/Col26 are involved in carbon catabolite repression. Deletion of *bglR*/*col26*, for example, increases cellulase production when the strains are grown on cellobiose, presumably due to lacking glucose signaling for CCR. (Xiong et al. 2014; Huberman et al. 2016).

Another protein that plays a role in CCR is the cAMP dependent protein kinase A (PKA), which often acts opposing to SnfA/1 (Thompson-Jaeger et al. 1991; Santangelo 2006; Barrett et al. 2012). It was observed that *pka* deletion mutants growing on cellulose are expressing lignocellulolytic enzymes earlier at higher levels in the cellulose response, indicating that the cells are glucose-blind and therefore allow production of cellulases under conditions in which CCR is active in wild type cells (De Assis et al. 2015). Finally, the transcription factor Vib1 is involved in CCR via regulating *creA/1*, *creB*, *creD* and *bglR*/*col26* homologs in *N. crassa* (Hynes and Kelly 1977; Kelly and Hynes 1977; Lockington and Kelly 2002; Boase and Kelly 2004; Nitta et al. 2012; Xiong et al. 2014).

There are still many mechanisms that are not fully understood and require further research, especially in *T. thermophilus*, where only a few research papers focus on understanding carbon catabolite repression. Nevertheless, Cre1 also seems to play a role in carbon catabolite repression in *T. thermophilus*. In strains where *cre1* was deleted or silenced, a higher production of cellulases like endoglucanases and ß-glucosidases compared to the parental strains could be observed (Yang et al. 2015; Liu et al. 2017). BglR has also been investigated via overexpression mutants. Here overexpression increased the beta-glucosidase activity under inducing conditions compared to the parental strain (Lai et al. 2017).

3.4.3 Regulators of cellulose degradation

Although the composition of cellulose is rather simple compared to other components of the plant biomass, many regulators are involved in the control of cellulase expression. The main regulators that are also conserved amongst many filamentous fungi, are Clr1/A and Clr2/B/ManR (Benocci et al. 2017). Clr2/B/ManR is regulating genes essential for cellulose degradation in *A. nidulans*, *N. crassa*, *A. niger*, *A. oryzae* and *Penicillium oxalicum* (Ogawa et al. 2013; Coradetti et al. 2013; Yao et al. 2015; Raulo et al. 2016). A constitutive expression of *clr2* in *N. crassa* causes an almost full cellulase expression even in the presence of repressive carbon sources. However, for other orthologs like *clrB* in *A. nidulans*, the constitutive overexpression does not result in constitutive expression of cellulolytic genes (Coradetti et al. 2013).

In *N. crassa* only, *clr2* expression is regulated via Clr1. The regulator Clr1 directly regulates genes that are necessary for the hydrolysis of cellulose and the import of soluble degradation products in many other filamentous fungi (Coradetti et al. 2012; Coradetti et al. 2013; Craig et al. 2015). In *N. crassa* and *A. nidulans* Clr1/A also has a role in cellulose sensing, meaning the presence of cellulose or its hydrolysis products (e.g. cellobiose) activates Clr1/A (Coradetti et al. 2013). In *Aspergillus spp*. ClrA is not required for the induction of cellulases. *T. reesei* does not have a homolog of Clr1, this stresses that different strategies for cellulose degradation and sensing exist in filamentous fungi (Tani et al. 2014b; Raulo et al. 2016). Constitutive overexpression of *clr1* does not lead to the expression of target genes, indicating the requirement of an activation or derepression similar to *xyr1* in *T. reesei* (Coradetti et al. 2012; Craig et al. 2015). In contrast, deletion of *clr1* as well as *clr2* in *N. crassa* causes an inability to grow on cellulose as well as a complete loss of cellulase activity when grown on cellulose (Coradetti et al. 2012). Clr1/2 also regulates the expression of other transcription factors involved in plant biomass degradation like *xlr1*, *vib1,* and *col26* in *N. crassa*. In *Aspergilli* the opposite can be observed: ClrA and ClrB are regulated via XlnR (Craig et al. 2015; Raulo et al. 2016).

Other only recently characterized regulators in the Clr family of transcriptional regulators are ClrC, which is known to modulate cellulase expression in *P. oxalicum* and Clr4, which positively regulates genes encoding (hemi)-cellulases and other regulators like Clr1, Clr2 and Xyr1 in *N. crassa* and *T. thermophilus* (Lei et al. 2016; Liu et al. 2019a).

In *A. aculeatus* an additional regulator controls cellulase expression. ClbR takes part in induction of cellulase expression via XlnR dependent and independent pathways, therefore overexpression results in an increase in xylanolytic and cellulolytic activities (Kunitake et al. 2013; Kunitake et al. 2015). The paralog of ClbR, which shares 42 % amino acid identity is named ClbR2. Orthologs of ClbR are found in only one order of fungi, whereas ClbR2 orthologs can be found in several orders. ClbR2 has not been extensively characterized, but it seems to compete with ClbR for the same DNA binding regions (Tani et al. 2014b; Benocci et al. 2017). In some orders of fungi also another paralog with a so far unknown function can be found: ClbR3. It is suggested that ClbR2 is the common ancestor and the other two paralogs originated via duplication events (Benocci et al. 2017).

Originally assigned as an inducer, the regulator Ace1 represses cellulase and xylanase production in *T. reesei* (Saloheimo et al. 2000; Aro et al. 2003). Although homologs of Ace1 have been found in many other filamentous fungi only few have been characterized. The deletion of *ace1* caused an increase in

cellulase and hemicellulase expression in *T. reese*i cultures grown on cellulose (Aro et al. 2003). Interestingly, *ace1* expression is induced via lactose and Cre1 mediated CCR (Portnoy et al. 2011). In contrast to Ace1 the regulators Ace2 and Ace3 are known to positively regulate the expression of cellulases and hemicellulases in *T. reesei* (Aro et al. 2001; Häkkinen et al. 2014). With cellulose as the sole carbon source the deletion of *ace2* caused a reduction in the expression of all main cellulase genes and a xylanase encoding gene in *T. reesei* (Aro et al. 2001). For Ace3 the regulation of cellulase and xylanase production was proven in *T. orientalis*, when deletion of *ace3* significantly reduced transcriptional levels of cellulase and hemicellulase genes as well as extracellular protein concentration (Liu et al. 2018). However, in *T. reesei* a truncation of *ace3* drastically increases cellulase titers (Chen et al. 2020). Apart from *Trichoderma spp.* those regulators have not been extensively investigated in other filamentous fungi.

Another regulator of cellulose degradation that is only characterized in a few filamentous fungi is McmA/1, which positively controls cellulase expression in *A. nidulans*, presumably via interaction with ClrB. On the contrary, McmA/1 has almost no effect on cellulase production in the fungus *Talaromyces cellulolyticus* (Yamakawa et al. 2013; Tani et al. 2014a; Fujii et al. 2015).

Also, a calcium dependent regulation of cellulase expression via the regulator Crz1 could be observed, e.g. in *T. reesei* (Chen et al. 2016).

Even antagonization for binding sites plays a role in cellulase regulation. This was observed in *T. reesei*, where a novel regulator, Rce1, was observed to regulate cellulase gene expression by blocking binding sites of Xyr1 in promotor regions of cellulase genes (Cao et al. 2017).

A recent publication discovered another new regulator: Stk12, whose deletion results in 7-fold higher total cellulase production in *N. crassa* compared to the wildtype. Due to the observations that the deletion of *stk12* enables a constantly high expression of cellulases over several days of cultivation, it is suggested that it acts as a transcriptional brake to control cellulase genes in *N. crassa* (Lin et al. 2019).

In *T. thermophilus* only a few of those regulators have been investigated in detail via deletion or overexpression mutants so far. As already mentioned, Clr4 positively regulates (hemi)-cellulase expression (Liu et al. 2019a) and a potential new regulator MHR1 represses cellulase activities (Wang et al. 2018). Furthermore, in a recent study, some new potential regulators that negatively affect cellulase production could be identified: Res1, Hcr1, Rca1, and Ap3 (Liu et al. 2019b). Regarding the regulators Clr1 and Clr2 the company BASF SE possesses patents for producing proteins in the industrial *T. thermophilus* strain C1. Here an elimination or decrease of Clr1 and Clr2 activity drastically increased the effectiveness of the production of recombinant polypeptides (Haefner et al. 2017a; Haefner et al. 2017b).

3.4.4 Regulators of hemicellulose degradation

The main and only exclusive regulator for hemicellulose degradation is Xyr1/ XlnR/ Xlr1 (Van Peij et al. 1998b; Van Peij et al. 1998a; Rauscher et al. 2006; Tamayo et al. 2008; Sun et al. 2012). Orthologs of this regulator exist in almost every filamentous fungus, where it controls xylose catabolism and the expression of xylanolytic enzymes (Van Peij et al. 1998b; Stricker et al. 2006; Calero-Nieto et al. 2007; Brunner et al. 2007; Sun et al. 2012; Todd et al. 2014). This regulator is also involved in the regulation

of other plant biomass degrading enzymes like cellulases but differs depending on the organism. In filamentous fungi like, e.g. *T. reesei*, *P. oxalicum*, and *A. niger* it activates transcription of cellulases (Hasper et al. 2000; Stricker et al. 2006; Li et al. 2015), whereas in *N. crassa* and *A. nidulans* it is not directly involved in the degradation of cellulose (Tamayo et al. 2008; Sun et al. 2012). Nevertheless, in *N. crassa* Xlr1 can also induce Vib1, which under starvation, represses glucose sensing and therefore CCR as mentioned above. This leads to the induction of cellulase production via Clr2 activation, meaning that Xlr1 is indirectly involved in cellulose degradation (Xiong et al. 2014). In *T. reesei* the overexpression of *xyr1* increases transcription of downstream targets, but the complete hemicellulolytic response is not activated, meaning that an activation or derepression is required, which can be achieved via a single amino acid substitution (Mach-Aigner et al. 2008; Derntl et al. 2013). Interestingly, in *T. reesei, xyr1* expression is induced by lactose and cellulose but not via xylan, which is the so far only reported case where Xyr1/ XlnR/ Xlr1 expression is not induced via xylan (Mach-Aigner et al. 2008; Portnoy et al. 2011; Bischof et al. 2013).

Other regulators of hemicellulose degradation are AraR/1, GalR, as well as GalX and will be discussed in context of pectin degradation since pectin and hemicellulose share some similar sugar residues in their side chains.

The regulator Hcr1 was discovered and characterized in 2014. It represses hemicellulase expression (mainly xylanases) during growth on arabinose and xylan and appears to be conserved amongst several cellulolytic filamentous fungi (Li et al. 2014).

In *T. thermophilus* the role of Xyr1 has also been investigated. Overexpression of *xyr1* results in enhanced xylanase activity and xylanase production (Wang et al. 2015), whereas the deletion of *xyr1* results in no growth on xylan and xylose, but similar growth on cellulose compared to the wildtype. The expression of cellulolytic genes is not significantly affected by the deletion, suggesting a similar role for Xyr1 as in *N. crassa* where it is an exclusive regulator for hemicellulose degradation (Dos Santos Gomes et al. 2019). Recently also a new regulator, Mhr1 was characterized, which represses cellulase as well as xylanase activities, explaining that silencing of *mhr1* caused an increased expression of the main cellulase genes and *xyr1* (Wang et al. 2018).

3.4.5 Regulators of pectin degradation

Due to the heterogeneity of pectin composition, regulation of pectin degradation is much more complex and involves more inducer activated regulators compared to, e.g. starch or cellulose. The best studied regulators that are involved in pectin degradation are: AraR/1, which responds to arabinose, RhaR, which responds to rhamnose, GalR and GalX, which respond to galactose, GaaR which responds to galacturonic acid and GaaX, a repressor, which also responds to galacturonic acid and presumably regulates expression of *gaaR* (Battaglia et al. 2011; Christensen et al. 2011; Gruben et al. 2014; Kowalczyk et al. 2015; Alazi et al. 2016; Niu et al. 2017). From those regulators only RhaR, GaaR, and GaaX are pectin-specific, since lots of similar side chains in pectin are also found in hemicellulose and lead to the induction of the previously mentioned regulators, because they share the same inducers. Therefore, identifying pectin specific transcriptional responses is challenging.

So far pectin degradation was mainly investigated in *Aspergilli* but orthologs of regulators are also existing in other fungi, some of them are summarized in Huberman et al. 2016. Other, fairly new regulators like a homolog of RhaR for example has been identified and characterized in *N. crassa* and named PDR-1 (Thieme et al. 2017).

In *T. reesei* and *Magnaporthe oryzae*, a different regulator compared to AraR which is induced via arabinose was identified and named Ara1 (Battaglia et al. 2011; Klaubauf et al. 2016). This regulator was also recently identified and characterized in *N. crassa* together with another new regulator named Pdr2. The deletion of *ara1* and *pdr2* for example is causing a deficiency in the utilization of arabinose and galactose (*ara1*) as well as galacturonic acid and pectin (*pdr2*). Furthermore, the *pdr2* deleted mutant shows a severe growth defect in medium containing pectin or galacturonic acid as the sole carbon source as well as reduced pectate lyase and polygalacturonase activity (Wu et al. 2020).

Deletion mutants of many other mentioned regulators, like an *A. niger* Δ*gaaR* strain, which cannot release and utilize galacturonic acid from pectin, do not have severe growth defects on pectin. This is due to the fact that other regulators enable growth on other sugars contained in pectin (Alazi et al. 2016; Benocci et al. 2017). The involvement of inducer activated regulators of pectin degradation in the degradation of cellulose has so far not been described.

In *T. thermophilus* regulation of pectin degradation has not been investigated, yet.

3.4.6 Regulators of starch degradation

Starch degradation is mainly regulated via the transcription factors AmyR and MalR and studies were mainly done in *Aspergilli*, in which the regulation of amylase genes and maltose utilizing cluster genes via those transcription factors has already been proven (Petersen et al. 1999; Gomi et al. 2000; Suzuki et al. 2015). Nevertheless, a similar function of AmyR orthologs was also found in other filamentous fungi like *N. crassa* and *P. oxalicum* (Li et al. 2015; Xiong et al. 2017). The expression of *amyR* is induced via starch, maltose, and isomaltose, whereas the expression of *malR* is induced via maltose only (Benocci et al. 2017). It is suggested that MalR is essential for the metabolism of maltose and the resulting subsequent production of isomaltose is the main inducer of AmyR in *A. oryzae* (Suzuki et al. 2015). AmyR is phylogenetically related to BglR/Col26 and therefore shares its glucose sensing function for CCR in various filamentous fungi (Benocci et al. 2017). This was observed for *P. oxalicum* and *N. crassa*, where AmyR directly represses cellulase gene transcription on cellulose. The deletion of *amyR* results in increased cellulase activity, suggesting an important role in the regulation of cellulolytic genes (Nitta et al. 2012; Li et al. 2015). Deletion of *amyR* in *A. oryzae* results in reduced growth on starch and in *A. nidulans* in no growth on starch and maltose, the phenotype on cellulose was not investigated (Petersen et al. 1999; Gomi et al. 2000; Murakoshi et al. 2012; Kowalczyk et al. 2014). A deletion of *malR* in *A. oryzae* leads to a growth defect on maltose and poor growth on starch (Hasegawa et al. 2010; Suzuki et al. 2015).

In *T. thermophilus*, the overexpression of *amyR* significantly increases amylase activity on starch and the deletion of *amyR* increases lignocellulase activities on cellulose (Xu et al. 2018).

3.4.7 Other factors involved in plant biomass degradation

In the following, a brief overview on other factors involved in plant biomass degradation is given. Factors that are playing a role are, e.g. chromatin remodeling, light, nitrogen source, pH value, and the secretion pathway.

The access to chromatin is generally important for the regulation of gene expression and therefore also for the regulation of plant biomass degradation. Here the most important players are the Hap complex, which consists of Hap2, Hap3, Hap4, and Hap5; as well as LaeA/1 which is important for the methylation and Hp1/HepA which controls the heterochromatin status (Narendja et al. 1999; Zeilinger et al. 2003; Karimi-Aghcheh et al. 2013; Fekete et al. 2014; Karimi-Aghcheh et al. 2014; Zhang et al. 2016). The Hap complex binds to the CCAAT box, which is present in promotor regions in about 30 % of all eukaryotic genes, and presumably opens chromatin structure (Narendja et al. 1999; Zeilinger et al. 2003; Benocci et al. 2017). In hemicellulase and cellulase genes CCAAT sequences are also present and are essential in *T. reesei* for cellulase gene expression. These sequences can be mutated to enhance transcription of xylanases (Zeilinger et al. 1998; Würleitner et al. 2003). Deletion of *hepA* in *P. oxalicum* caused a derepression of gene expression, an increase in the expression of *creA* and *clrB* as well as a downregulation of prominent cellulolytic enzyme genes (Zhang et al. 2016).

Surprisingly light also influences plant biomass degradation in filamentous fungi. Examples have been shown for *T. reesei*, *N. crassa*, and several *Aspergilli* (Purschwitz et al. 2008; Purschwitz et al. 2009; Castellanos et al. 2010; Idnurm and Heitman 2010; Gyalai-Korpos et al. 2010; Schmoll et al. 2012; Tisch et al. 2014; Craig et al. 2015). The two known complexes responsible for light dependent regulations are the White collar complex (WCC) together with VIVID (VVD) and the VELVET complex (Schafmeier and Diernfellner 2011; Bayram and Braus 2012; Karimi-Aghcheh et al. 2014; Hurley et al. 2015; Proietto et al. 2015). The White collar complex consists of the blue light/ UV-A photoreceptors Wc1/Blr1 and Wc2/Blr2 whereas the VELVET complex consists of VeA/Vel1, VelB, and LaeA (Bayram et al. 2008; Bayram and Braus 2012; Schmoll et al. 2012; Tisch and Schmoll 2013). In *T. reesei* for example Lae1 and Vel1 are essential for the expression of (hemi)-cellulase encoding genes (Seiboth et al. 2012; Karimi-Aghcheh et al. 2014).

Nitrogen sources affect plant biomass degradation via Nitrogen Metabolite Repression (NMR) in a similar manner compared to CCR (Marzluf 1997; Tudzynski 2014). Therefore, during nitrogen starvation, NMR negatively affects the expression of hydrolytic enzymes including (hemi)-cellulases which was observed for example in *A. nidulans* and *N. crassa* (Arst and Cove 1973; Fu and Marzluf 1990; Berger et al. 2008; Gonçalves et al. 2011; Amore et al. 2013). The key regulators here are the activator AreA/Nit2 and the two repressors AreB and NmrA/1. AreA directly activates transcription of the target genes and the two repressors are modulating AreA activity (Arst and Cove 1973; Tomsett et al. 1981; Fu and Marzluf 1987; Kudla et al. 1990; Caddick 1994; Andrianopoulos et al. 1998; Conlon et al. 2001; Lamb et al. 2004; Berger et al. 2006; Macios et al. 2012).

Lignocellulolytic enzymes are furthermore regulated by PacC/1, the key transcription factor for pH regulation (Rossi et al. 2013). This transcription factor directly regulates (hemi)-cellulase production, but it also indirectly effects plant biomass degradation by modulating activities of transcription factors like

XlnR and ClrB (De Graaff et al. 1994; Kunitake et al. 2016). In *T. reesei* for example PAC1 induces the expression of *xyr1* and *ace2* and therefore increases cellulase production (He et al. 2014).
Interestingly also a homolog of the Sterol Regulatory Element Binding Protein (SREBP) pathway, Sah2 is involved in cellulase production in *T. reesei* and *N. crassa*. Here, deletion of *sah2* resulted in cellulase hyper-production phenotypes (Reilly et al. 2015).
Furthermore, the unfolded protein response (UPR) plays an important role in the regulation of lignocellulolytic enzymes. Usually, the accumulation of unfolded proteins activates Ire1, which then releases Hac1 from transcriptional inhibition (Sidrauski and Walter 1997). Hac1 then activates transcription of genes that enhance folding capacity and allow an enhanced protein trafficking (Mori et al. 1996; Kaufman 1999). It is suggested that UPR also plays an important role in balancing protein secretion with processing capacity during the lignocellulolytic response (Huberman et al. 2016). In *N. crassa* it was observed, that *ire1* and *hac1* are upregulated when the cells are grown on cellulose (Benz et al. 2014). Furthermore, the deletion of *ire1* or *hac1* lowers cellulase secretion, but cellulase transcription itself remains unaffected (Fan et al. 2015; Montenegro-Montero et al. 2015). Moreover, it was observed that in a *T. reesei* cellulase producer strain which was selected based on increased cellulase secretion, UPR is activated earlier in the lignocellulolytic response compared to the respective wild type (Wang et al. 2014).
Finally, the transcription factor Xpp1 is involved in the regulation of hemicellulase expression. Xpp1 is usually responsible for a switch between primary and secondary metabolism and represses xylanase-encoding genes in *T. reesei* as a secondary effect (Derntl et al. 2015; Derntl et al. 2017).
None of the described factors have been studied in *T. thermophilus*, yet. This is also the case in many other filamentous fungi, stressing the potential of investigating these regulation mechanisms to be able to control lignocellulolytic enzyme expression for already discussed industrial applications.

3.5 Aim of this study

T. thermophilus is a fungus with huge biotechnological potential as already discussed in the previous chapters. Industrial strains like the C1 strain are already established in industry to produce various proteins and enzymes. Nevertheless, transcriptional regulation of the strongly secreted lignocellulolytic enzymes under industrial relevant conditions, are so far poorly investigated. Understanding of these regulation processes has various benefits. It will help lignocellulosic biorefineries by providing optimized enzyme cocktails according to the used substrate which in turn helps to overcome the major bottlenecks and problems in lignocellulolytic biorefineries. Besides the ability to identify regulation processes that enable a higher production of certain lignocellulolytic enzymes, also the opposite becomes possible when regulation processes are known. This is especially important when producing heterologous proteins, as production of lignocellulolytic enzymes or proteases are lowering yields and purity of the protein of interest. The benefit of reducing this secretory burden has already been described in the BASF patents by Haefner et al. 2017a and Haefner et al. 2017b, where an elimination or decrease of the activity of the cellulase regulators Clr1 and Clr2 caused a significant advantage regarding yield and purity of heterologously expressed proteins in industrial *T. thermophilus* strains compared to their parental strain. Therefore, to enable the manipulation of the secretome of *T. thermophilus* to improve

this fungus as a production host for heterologous proteins and lignocellulolytic enzyme cocktails, the following aims were targeted during this work:

1) Establishment of molecular biological methods for *T. thermophilus* and application of those methods to delete known and potentially new regulators of cellulase expression

2) Establishment of stable bioreactor cultivation protocols with subsequent chemostat bioreactor cultivations of all generated strains and their parental strain under inducing and non-inducing conditions including biochemical characterizations

3) Transcriptomic analysis of hydrolase (proteases, CAZYs) and transcription factor expression via RNA sequencing and differential expression analysis of samples gained from the bioreactor cultivations to investigate regulation dependencies and unravel regulatory mechanisms regarding lignocellulolytic enzyme expression

4. Materials and Methods

4.1 Materials

4.1.1 Bacterial Strains

Table 4.1: Bacterial strains used in this study.

strain name	relevant genotype	reference
E. coli TOP10	F-*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*Z ΔM15 Δ*lac*X74 *nup*G *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1	Invitrogen (Karlsruhe, Germany)

4.1.2 Fungal Strains

Table 4.2: *T. thermophilus* strains used in this study. An= *Aspergillus nidulans*, DR= direct repeat.

strain name	genotype	description	reference
ATCC 42464	wildtype	wildtype strain	DSMZ (Braunschweig, Germany)
MJK19.4	Δ*ku80* :: DR-P An *gpdA*–An *amdS*-T An *amdS*-DR	*ku80* deleted via *amdS* split marker transformation of deletion cassettes from MT125 and MT126 in ATCC 42464	this study, created by Min Jin Kwon
MJK20.2	Δ*ku80*	marker recycling of MJK19.4 via FAA plating, mixed clone with ATCC 42464	this study, created by Min Jin Kwon
MJK20.3	Δ*ku80*	MJK20.2 subcultivated and re-analyzed	this study
BS3.6	Δ*ku80*, *vib2* :: DR-P An *gpdA*–An *amdS*-T An *amdS*-DR	*vib2* deleted via *amdS* split marker transformation of deletion cassettes from MT134 and MT136 in MJK20.2	this study
BS3.8	Δ*ku80*, *vib2* :: DR-P An *gpdA*–An *amdS*-T An *amdS*-DR	*vib2* deleted via *amdS* split marker transformation of deletion cassettes from MT134 and MT136 in MJK20.2	this study
BS3.12	Δ*ku80*, *vib2* :: DR-P An *gpdA*–An *amdS*-T An *amdS*-DR	*vib2* deleted via *amdS* split marker transformation of deletion cassettes from MT134 and MT136 in MJK20.2	this study
BS3.15	Δ*ku80*, *vib2* :: DR-P An *gpdA*–An *amdS*-T An *amdS*-DR	*vib2* deleted via *amdS* split marker transformation of deletion cassettes from MT134 and MT136 in MJK20.2	this study
BS4.1	Δ*ku80*, Δ*vib2*	marker recycling of BS3.8 via FAA plating	this study
BS4.2	Δ*ku80*, Δ*vib2*	marker recycling of BS3.15 via FAA plating	this study
BS5.1	Δ*ku80*, *clr2* :: DR-P An *gpdA*–An *amdS*-T An *amdS*-DR	*clr2* deleted via *amdS* split marker transformation of deletion cassettes from MT121 and MT497 in MJK20.2	this study
BS5.6	Δ*ku80*, *clr2* :: DR-P An *gpdA*–An *amdS*-T An *amdS*-DR	*clr2* deleted via *amdS* split marker transformation of deletion cassettes from MT121 and MT497 in MJK20.2	this study
BS5.13	Δ*ku80*, *clr2* :: DR-P An *gpdA*–An *amdS*-T An *amdS*-DR	*clr2* deleted via *amdS* split marker transformation of deletion cassettes from MT121 and MT497 in MJK20.2	this study
BS5.14	Δ*ku80*, *clr2* :: DR-P An *gpdA*–An *amdS*-T An *amdS*-DR	*clr2* deleted via *amdS* split marker transformation of deletion cassettes from MT121 and MT497 in MJK20.2	this study
BS6.1	Δ*ku80*, Δ*clr2*	marker recycling of BS5.6 via FAA plating	this study
BS6.2	Δ*ku80*, Δ*clr2*	marker recycling of BS5.13 via FAA plating	this study
BS6.3	Δ*ku80*, Δ*clr2*	marker recycling of BS5.1 via FAA plating	this study
BS6.4	Δ*ku80*, Δ*clr2*	marker recycling of BS5.14 via FAA plating	this study

BS7.4	Δ*ku80*, Δ*clr1*	*clr1* deleted via *amdS* split marker transformation of deletion cassettes from MT122 and MT227 in MJK20.3 with CRISPR/Cas12a. Counterselection was not necessary because the strain removed *amdS* marker itself during subcultivation.	this study
BS7.8	Δ*ku80*, Δ*clr1*	*clr1* deleted via *amdS* split marker transformation of deletion cassettes from MT122 and MT227 in MJK20.3 with CRISPR/Cas12a. Counterselection was not necessary because the strain removed *amdS* marker itself during subcultivation.	this study
BS7.10	Δ*ku80*, Δ*clr1*	*clr1* deleted via *amdS* split marker transformation of deletion cassettes from MT122 and MT227 in MJK20.3 with CRISPR/Cas12a. Counterselection was not necessary because the strain removed *amdS* marker itself during subcultivation.	this study
BS7.13	Δ*ku80*, Δ*clr1*	*clr1* deleted via *amdS* split marker transformation of deletion cassettes from MT122 and MT227 in MJK20.3 with CRISPR/Cas12a. Counterselection was not necessary because the strain removed *amdS* marker itself during subcultivation.	this study
JK1.1	Δ*ku80*, *clr4* :: DR-P An *gpdA*–An *amdS*-T An *amdS*-DR	*clr4* deleted via transformation of pBS1.13 and CRISPR/Cas12a in MJK20.3	this study, created by Jonas Karsten
JK1.8	Δ*ku80*, *clr4* :: DR-P An *gpdA*–An *amdS*-T An *amdS*-DR	*clr4* deleted via transformation of pBS1.13 and CRISPR/Cas12a in MJK20.3	this study, created by Jonas Karsten
JK2.1	Δ*ku80*, Δ*clr4*	marker recycling of JK1.1 via FAA plating	this study, created by Jonas Karsten
JK2.8	Δ*ku80*, Δ*clr4*	marker recycling of JK1.8 via FAA plating	this study, created by Jonas Karsten
BS9.1	Δ*ku80*, Δ*vib1*	*vib1* deleted via transformation of pBS4.2 and CRISPR/Cas12a in MJK20.3, marker recycling immediately after transformation	this study
BS9.2	Δ*ku80*, Δ*vib1*	*vib1* deleted via transformation of pBS4.2 and CRISPR/Cas12a in MJK20.3, marker recycling immediately after transformation	this study
BS9.3	Δ*ku80*, Δ*vib1*	*vib1* deleted via transformation of pBS4.2 and CRISPR/Cas12a in MJK20.3, marker recycling immediately after transformation	this study
BS9.4	Δ*ku80*, Δ*vib1*	*vib1* deleted via transformation of pBS4.2 and CRISPR/Cas12a in MJK20.3, marker recycling immediately after transformation	this study
BS9.5	Δ*ku80*, Δ*vib1*	*vib1* deleted via transformation of pBS4.2 and CRISPR/Cas12a in MJK20.3, marker recycling immediately after transformation	this study

4.1.3 Plasmids

Table 4.3: Plasmids used in this study.

name	description	reference
MT121	puc Ori, 5´*clr2*, P An *gpdA*, An *amdS* part 1, kanR	BASF SE (Ludwigshafen, Germany)
MT122	puc Ori, 5´*clr1*, P An *gpdA*, An *amdS* part 1, kanR	BASF SE (Ludwigshafen, Germany)
MT125	puc Ori, 5´*ku80*, P An *gpdA*, An *amdS* part 1, kanR	BASF SE (Ludwigshafen, Germany)
MT126	puc Ori, An *amdS* part 2, T An *amdS*, 5´*ku80*, 3´*ku80*, kanR	BASF SE (Ludwigshafen, Germany)
MT134	puc Ori, 5´*vib2*, P An *gpdA*, An *amdS* part 1, kanR	BASF SE (Ludwigshafen, Germany)
MT136	puc Ori, An *amdS* part 2, T An *amdS*, 5´*vib2*, 3´*vib2*, kanR	BASF SE (Ludwigshafen, Germany)
MT227	puc Ori, An *amdS* part 2, T An *amdS*, 5´*clr1*, 3´*clr1*, kanR	BASF SE (Ludwigshafen, Germany)
MT28	puc Ori, An *amdS*, ampR; positive control for transformation	BASF SE (Ludwigshafen, Germany)
MT497	puc Ori, An *amdS* part 2, T An *amdS*, 5´*clr2*, 3´*clr2*, kanR	BASF SE (Ludwigshafen, Germany)

MT1402	puc Ori, *lacZ,* 5´*vib1*, 3´*vib1*, ampR	BASF SE (Ludwigshafen, Germany)
pBS1.13	puc Ori, 5´*clr4*, P An *gpdA*, An *amdS,* T An *amdS*, 5´*clr4*, 3´*clr4*, kanR	this study
pMJK19.7	ColE1 Ori, 5´*ku70*, P An *gpdA*, An *amdS*, 3´*ku70*, ampR	this study, created by Min Jin Kwon
pMJK18.1	ColE1 Ori, 5´*ku80*, P An *gpdA*, An *amdS*, 3´*k80*, ampR	this study, created by Min Jin Kwon
pBS4.2	puc Ori, 5´*vib1*, P An *gpdA*, An *amdS,* T An *amdS*, 5´*vib1*, 3´*vib1*, ampR	this study

4.1.4 Oligonucleotides

All oligonucleotides used in this study were ordered at Eurofins Genomics (Ebersberg, Germany). Oligonucleotides arrived lyophilized and were subsequently diluted to 100 µM with H_2O MQ. Storage took place at -20 °C. Primer numbers are referring to the numbering system of the in-house database.

Table 4.4: Oligonucleotides for the amplification of split marker deletion cassettes used in this study.

number	sequence (5´→ 3´)	description
1619	AAATCCCGTAGCGGCCG	*amdS* part 1 rev primer, amplification of split marker deletion cassette 1
1620	AAATCCCGTAGCGGCCGCA	amplification of split marker deletion cassette 2 from MT227 together with primer 1624
1624	AAATCTCGAGAGGCCTGACGTCG	amplification of split marker deletion cassette 2 from MT227 together with primer 1620
1625	AAATCTCGAGCGACAGCGA	amplification of split marker deletion cassette 1 from MT125 together with primer 1619
1635	TTCTGACAACCATGCTCCG	amplification of split marker deletion cassette 2 from MT126 together with primer 1636
1636	AACTGACGCTCGACTGG	amplification of split marker deletion cassette 2 from MT126 together with primer 1635
1637	AATCTCGAGAGGCCTGACG	amplification of split marker deletion cassette 1 from MT122 together with primer 1619
1657	CATGCACGCCCGTAATGAAG	amplification of split marker deletion cassette 1 from MT121 together with primer 1619
1658	GACTCGGTTCTGACAACCATG	*amdS* part 2 fw primer, amplification of split marker deletion cassette 2
1659	GTACTCTCTGCTTCACCATCG	amplification of split marker deletion cassette 2 from MT497 together with primer 1658
1660	TGAGCTGGGTGCAGGATG	amplification of split marker deletion cassette 1 from MT134 together with primer 1619
1661	GAAATGTGCCACCATCACTGC	amplification of split marker deletion cassette 2 from MT136 together with primer 1658

Table 4.5: Oligonucleotides used for the diagnostic PCR of potential regulator deletion clones used in this study.

number	sequence (5´→3´)	description
1681	CCACCAGGGCTACGAAACATC	An *amdS*, fw primer
1689	CGGAAGAGTAGATGCACAACGG	*ku80* 3´Locus, rev primer
1732	CTTTGCGGGCTTGAACTACG	*ku80* ORF, fw primer
1684	ACCGTCTGTTGAGCACCTG	*ku80* 5´Locus, fw primer
1817	CAAGAGATGCCAAATGCAGG	*clr2* 3´Locus, rev primer
1734	CATGATGACGGGCTGGGTTC	*clr2* ORF, fw primer
1687	CCACAGCGGAGCATCAGGC	*clr2* 5´Locus, fw primer
1690	TGTCCGTCATGGGCAAAG	*vib2* 3´Locus, rev primer
1736	GTGCGGAGCAAAGCAAAGC	*vib2* ORF, fw primer
1685	TCGTCGGCTCAGGCAACCTT	*vib2* 5´Locus, fw primer
1733	GTTCTTTCTCCCAGCCTCAGC	*clr1* ORF, fw primer
1691	GATGGAAACACTTCTCCGATC	*clr1* 3´Locus, rev primer
1686	CCATGTTGCGGACTTGCTCC	*clr1* 5´Locus, fw primer
2163	GAAGGACTCAAAGCCGAAGC	*clr4* ORF, fw primer
2162	CGTACAGTGTCAAGGCAAATGG	*clr4* 3´Locus, rev primer

2161	CAGTAGCCGTGGTAGGACG	*clr4* 5´Locus, rev primer
2525	GATACGGGTACACTTGGAACG	*vib1* ORF, fw primer
2381	CAGTCGACTGGTCTCGATCG	*vib1* 3´Locus, rev primer
2382	GAGCAACCGCGATCTTCAC	*vib1* 5´Locus, fw primer

Table 4.6: Oligonucleotides used for the amplification of probes for Southern analyses of potential regulator deletion clones used in this study.

number	sequence (5´→3´)	description
1741	CTAAGACCCACACCACCACC	amplification of 5´ flank probe for Southern analysis of potential *clr2* deletion clones
1742	GTAGTGTAACAGGCGAGAGCG	
1743	CTCGGGTTCACCTTTCAAATGC	amplification of 3´ flank probe for Southern analysis of potential *clr2* deletion clones
1744	GATTTGTGGTTGGCGGCTG	
1745	GTGTCGTGGTCAGAGAGCTAAG	amplification of 5´ flank probe for Southern analysis of potential *vib2* deletion clones
1746	GCTCCTTAGGTGGGCAAACC	
1747	GTTTCGGATTTGGCCCTTGG	amplification of 3´ flank probe for Southern analysis of potential *vib2* deletion clones
1748	GTTGCGTTGGCTTGCTTGTC	
1749	GAACGCGGAAGGAGATGATGC	amplification of 5´ flank probe for Southern analysis of potential *ku80* deletion clones
1750	GACGGAAGCTCATGCAGTGAC	
1751	GTGAGCATTGCGCTTGTGG	amplification of 3´ flank probe for Southern analysis of potential *ku80* deletion clones
1752	GTTCTGGAAGGAGCTGCTGG	
1737	GTTGGATGGAATGCTCGGACC	amplification of 5´ flank probe for Southern analysis of potential *clr1* deletion clones
1738	GAGCTGCACAACGCGATCAG	
1739	CGCAGCGAGTCACAATTTGG	amplification of 3´ flank probe for Southern analysis of potential *clr1* deletion clones
1740	GTGGCTTGATCTCATGGACGC	
2081	GAGCTCGAATTGGACGCTGATGGTATGTCAACGGTTGATCC	amplification of 5´ flank probe for Southern analysis of potential *clr4* deletion clones
2128	CGTAGGTAAGCAAGATCATCG	
2383	CAGGATTCCTCGGAGGATTC	amplification of 5´ flank probe for Southern analysis of potential *vib1* deletion clones
2384	GTGGATGCGGATTGTGTC	

Table 4.7: Oligonucleotides used for the amplification of the DNA template for the guide RNA used in CRISPR/Cas12a based transformations in this study.

number	sequence (5´→3´)	description
1697	ATGTAATACGACTCACTATAGGTAATTTCTACTGTTGTAGAT	general fw primer for amplification of the sgRNA template carrying T7 promotor sequence and direct repeat
2048	CTTCTTTGATGATTTCAGCGATCTACAACAGTAGAAATTA	amplification of a sgRNA template for the deletion of *clr1* via CRISPR/Cas12a together with primer 1697
2049	CTCAAATTCTCCTCCAGCTTATCTACAACAGTAGAAATTA	amplification of a sgRNA template for the deletion of *clr1* via CRISPR/Cas12a together with primer 1697
2152	GGTCAGTGGTGAATGCGCGCTGTATCTACAACAGTAGAAATTA	amplification of the sgRNA template for the deletion of *clr4* via CRISPR/Cas12a together with primer 1697
2390	ATGCTTGTCTCCAACCTTCTTATCTACAACAGTAGAAATTA	amplification of the sgRNA template for the deletion of *vib1* via CRISPR/Cas12a together with primer 1697

Table 4.8: Oligonucleotides used for the amplification of overlapping fragments for CPEC and analysis of the resulting plasmid after CPEC in this study.

number	sequence (5´→3´)	description
2128	CGTAGGTAAGCAAGATCATCG	amplification of 5´flank fragment A for CPEC
2081	GAGCTCGAATTGGACGCTGATGGTATGTCAACGGTTGATCC	

2131	ACTGATTGATGACGGCTGAGG	amplification of 3´flank fragment for CPEC
2087	GGACTGGCTTTCTACGTGTTCGTGTTTTCCACGAAGCGAGC	
2130	CGTAGGTAAGCAAGATCATCG	amplification of 5´flank fragment B for CPEC
2085	CCTCAGCCGTCATCAATCAGTTGGTATGTCAACGGTTGATCC	
2129	TCAGCGTCCAATTCGAGCTC	amplification of *amdS* fragment for CPEC
2083	CGATGATCTTGCTTACCTACGCATGGGTTGAGTGGTATGG	
2132	GAACACGTAGAAAGCCAGTCC	amplification of backbone fragment for CPEC
2089	CGATGATCTTGCTTACCTACGGACGTCAGGCCTCTCGAGAT	
2136	GAGCTCGAATTGGACGCTGA	fusion of backbone and 5´ flank fragment A together with primer 2132
2139	GGACTGGCTTTCTACGTGTTC	fusion of 3´flank and 5´ flank fragment B together with primer 2130
2145	CTCCATATTCTCCGATGATGC	primers used for sequencing of the resulting plasmid after CPEC
2146	GGCACAAGTGTCTCTCACC	
2147	GAATCCCAATCTTAACGCTACC	
2148	CATTCGACGTAACAGCTCG	
2149	GCAAGCAGCAGATTACGC	

Table 4.9: Oligonucleotides used for the amplification of overlapping fragments for Gibson cloning and analysis of the resulting plasmid after Gibson cloning in this study.

number	sequence (5´→3´)	description
2291	GGCTCGAGTTTTTCAGCAAGATCGCTCTTGTCGCTGATGATTTCG	amplification of 5´flank fragment for Gibson cloning
2292	TGACGGAAGCGTGCAGATGG	
2295	CCATACCACTCAACCCATGCGCGATCTCCTCCTCTGG	amplification of 5´3´flank fragment for Gibson cloning
2296	GTCTTATGCTGCGGGTCG	
2293	CCATCTGCACGCTTCCGTCATCAGCGTCCAATTCGAGCTC	amplification of *amdS* fragment for Gibson cloning
2294	CATGGGTTGAGTGGTATGG	
2297	CGACCCGCAGCATAAGACCTCCTACAATATTCTCAGCTGC	amplification of backbone fragment for Gibson cloning
2298	ATCTTGCTGAAAAACTCGAGCC	
2145	CTCCATATTCTCCGATGATGC	primers used for sequencing of the resulting plasmid after Gibson cloning
2146	GGCACAAGTGTCTCTCACC	
2323	GGTATGGGTTGGTGACATATG	
2324	GGAGCAGGTTCCATTCATTG	

Table 4.10: Oligonucleotides used for qPCR in this study.

number	sequence	description
2587	CTCGTCGATCAAATCGATCC	primer for qPCR; target gene: MYCTH_2316610
2588	CGACGTGCTTCAGGAAC	
2589	GATTGCCCAGGTCGTCTC	primer for qPCR; target gene: MYCTH_2294321
2590	GCTCGAAGCACTGGAAG	
2591	CAAGACGCAGTTCGTCAAC	primer for qPCR; target gene: MYCTH_2303045
2592	CTCCTGCCCGTCTTCTTC	
2593	CAAGGCAGCGTACAAGG	primer for qPCR; target gene: MYCTH_2300079
2594	GTCGACCTCGTGATGATG	

4.1.5 Consumables

Table 4.11: Consumables used in this study.

item	manufacturer	article number
1.5 ml reaction tube	Sarstedt	72.690.001
10 µl pipette filter tips	Sarstedt	70.1130.210
10 µl pipette tips	Sarstedt	70.13

1000 µl pipette filter tips	Sarstedt	70.762.211
1000 µL pipette tips	Sarstedt	70.762
15 mL reaction tube	Sarstedt	20003468
2 ml reaction tube	Sarstedt	72.691
2.5 mL dispenser tips	Ritter	H682.1
20 µl pipette filter tips	Sarstedt	70.760.213
200 µl pipette filter tips	Sarstedt	70.760.211
20-200 µL pipette tips	Sarstedt	70.760.002
50 mL reaction tube	Sarstedt	10535253
96 well plate	Greiner Bio-One	655098
96 well plate	Perkin Elmer	6005181
biomass filter 3hw	Sartorius	FT-3-303-045
chromatography/whatman paper	VWR	588-3184
clingfilm	Ecopla	N9686
cotton sticks	Carl Roth	EH11.1
cover glass	Carl Roth	657
cryotubes	Greiner Bio-One	122.278
filters bioreactor (Midisart 2000)	Sartorius	90148113
latex and nitrile gloves	Ansell	IH-92670, IH-69318
luer-locks, tubing connections	Carl Roth	E774.1, E791.1, E772.1, E793.1, CT61.1, CT65.1, CT70.1, CT69.1
microscopy slide	Carl Roth	656
miracloth	VWR	475855-1R
nylon membrane	Carl Roth	K058.1
parafilm M	Carl Roth	CNP8.1
PCR stripes	Biozym	711057
PCR tubes (single)	Sarstedt	72.737.002
Petri dishes (150x20mm)	Sarstedt	82.1184.500
Petri dishes (92x16mm)	Sarstedt	82.1473.001
SDS gel loading pipette tips	Sarstedt	70.1190.100
serological pipettes (10 ml)	Sarstedt	86.1685.001
serological pipettes (2 ml)	Sarstedt	86.1252.001
serological pipettes (25 ml)	Sarstedt	86.1254.001
serological pipettes (5 ml)	Sarstedt	86.1253.001
spatula	Carl Roth	PC57.1
sterile filter (0.2 µm)	Sarstedt	83.1826.001
syringe cannula	Erasa/Braun	0749/X132.1
syringes	Braun	4617207
tinfoil	Carl Roth	0954.1
toothpicks	Franz Mensch GmbH	3882
tubings bioreactor	Carl Roth	9568.1

4.1.6 Kits

In all kits, where a special elution buffer was provided, H_2O MQ was used instead.

Table 4.12: Kits used in this study.

kit name	manufacturer
Bio-Rad Protein Assay Kit II	Biorad
Biozym Blue S´Green qPCR Kit	Biozym
DNA-free™ DNA Removal Kit	Invitrogen
Glucose Fluid GOD-PAP	Mti Diagnostics
innuPREP DOUBLEpure Kit	Analytik Jena
innuPREP Plasmid Mini Kit 2.0	Analytik Jena
innuPREP RNA Mini Kit 2.0	Analytik Jena

MEGAscript™ T7 Transcription Kit	Thermo Fisher Scientific
Pierce™ BCA Protein Assay Kit	Thermo Fisher Scientific
PureYield™ Plasmid Midiprep System	Promega
RevertAid H Minus First Strand cDNA Synthesis Kit	Thermo Fisher Scientific
D-xylose Assay Kit	Megazyme

4.1.7 Antibodies

Table 4.13: Antibodies used in this study.

antibody name	manufacturer	article number
anti-digoxigenin-AP, Fab fragments	Roche	11093274910

4.1.8 Equipment

Table 4.14: Equipment used in this study.

instrument/machine	type	manufacturer
autoclaves	Timo	international pbi
	VX150	Systec
	Vapour Line Lite	VWR
analytical balance	Entris124-1S	Sartorius
balances	Entris3202-1S	Sartorius
	IND560	Mettler Toledo
	EL4001	Mettler Toledo
bioreactor	Bioflo & Celligen 310	New Brunswick
cell counting chamber	Thoma	Marienfeld-Superior GmbH
centrifuge	Biofuge primo R	Heraeus instruments
	5427R	Eppendorf
	Rotina 38 R	Hettich-Zentrifugen
	Spectrafuge Mini	Labnet
	Biofuge Stratos	Heraeus instruments
	Megafuge 16R	Heraeus instruments
ChemiDoc imager	ChemiDoc MP	Biorad
crosslinker	UV Stratalinker® 1800	Stratagene
dispenser	Stepmate	Abimed
DNA gel electrophoresis chamber	with Biorad PowerPac HC power supply	Biorad
DO probe (bioreactor)	InPro 6860i	Mettler Toledo
freeze dry system	Freezone 2.5	Labconco
gel documentation	Gel logic 212pro	Carestream
hybridization oven	UVP Hybridizer	Analytik Jena
ice machine	SPR-80L	NordCap
incubator	HERATherm	Thermo Fisher Scientific
	B5042	Heraeus instruments
incubator (shaking)	Multitron II	Infors HT
laminar flow cabinet	BDK-SB 1200	BDK Luft-und Reinraumtechnik GmbH
magnetic stirrer	IKA C-MAG MS10	IKA Labortechnik
microscope	DM5000CS	Leica Scientific
microwave	MW764	Clatronic
multiplate reader	VictorX	PerkinElmer
	Glomax	Promega
offgas analyzer (bioreactor)	EX-2000	New Brunswick
pH probe (bioreactor)	405-DPAS-SC-K8S/225 combination pH	Mettler Toledo
pipettes, pipetman classic	G2, G10, G20, G200, G1000	Gilson

platform shaker	Duomax 1030	Heidolph Instruments
pumps	120U	Watson Marlow
	Masterflex L/S	Cole Parmer
qPCR system	CFX96	Biorad
recirculation cooler (bioreactor)	Frigomix 1000	Sartorius
RNA gel electrophoresis system	SEA2000 with GEPS 200/2000 power supply	Elchrom Scientific
SDS-PAGE system	Mini-PROTEAN Tetra handcast systems	Biorad
sequencer (bioreactor)	8-Channel Gas Sequencer	New Brunswick
spectrophotometer	BioSpectrometer	Eppendorf
temperature sensor (bioreactor)	RTD Cable Assembly	Eppendorf
thermoshaker	TSC Thermoshaker	Biometra
tilting shaker	WT 17	Biometra
vacuum manifold	Vac-Man® Laboratory Vacuum Manifold	Promega
vacuum pump	XF5423050	Millipore
vortex	VF2	IKA Labortechnik
water purification system (MQ)	PURELAB Classic	Elga LabWater

4.1.9 Chemicals, media and buffers

4.1.9.1 Chemicals

Table 4.15: Chemicals used in this study.

chemical	manufacturer	article number
2-mercaptoethanol	Sigma Aldrich	M7154
2-propanol	Carl Roth	6752.2
40% acrylamide:bisacrylamide (29:1)	Carl Roth	A121.1
6x DNA loading dye	Thermo Fisher Scientific	R0611
acetamide	Fluka	160-250G
agar-agar	Serva	11392.04
agarose	Biozym	840004
ammonia solution 25%	VWR	1336-21-6
ammonium molybdate tetrahydrate	Fluka	261823
ammonium persulfate (APS)	Serva	13375
ammonium sulfate	Merck	1.01211.500
ampicillin sodium salt	Applichem	A0839.0025
arabinogalactan (larch wood)	Sigma	10830-25G
aurintricarboxylic acid ammonium salt	Sigma-Aldrich	A36883
betaine monohydrate	Sigma	B2754
biotin	Serva	15060
blocking reagent (Southern)	Roche	1096176
boric acid	Merck	1001651000
bromophenol blue	Amersham Biosciences	17-1329-01
bromophenol blue	Amersham Biosciences	17-1329-01
BSA	Pharmacia	27-8915-01
calcium chloride dihydrate	Carl Roth	5239.3
casamino acids	BD Biosciences	223050
CDP-star	Roche	11759051001
cellulose microcrystalline	Merck	1.02331.9025
cesium chloride	Biomol	2452
chloroform	Carl Roth	3313.2
cobalt (II) chloride hexahydrate	Aldrich	25,559-9
coomassie brilliant blue G250	Merck	115444
copper (II) sulfate pentahydrate	Merck	2790
D-cellobiose	Carl Roth	5840.1

dextrane sulfate	Carl Roth	5956.1
D-galactose	Merck	104062
D-galacturonic acid monohydrate	Sigma	G-2125
D-glucose monohydrate	Carl Roth	6887.5
DIG labelled dNTPs	Roche	11585550910
dipotassium hydrogen phosphate	Merck	4876
disodium hydrogen phosphate	Merck	1065862500
D-lactose monohydrate	Carl Roth	6868.2
D-maltose monohydrate	Carl Roth	8951.4
D-mannose	Carl Roth	4220.1
DMSO	Merck	29521000
dNTPs	Thermo Fisher Scientific	R0181
DreamTaq™ Green buffer	Thermo Fisher Scientific	B71
D-sorbitol	Carl Roth GmbH+Co. KG	6213.3
D-sucrose	Carl Roth GmbH+Co. KG	4621.5
DTT	Gerbu	1008
D-xylose	Sigma-Aldrich	W360600
EDTA	Carl Roth	CN06.3
EDTA disodium salt dihydrate	Sigma	E6635
ethanol	Carl Roth	T171.2
fluoroacetamide	Aldrich	128341
glycerol	Carl Roth	75301
glycine	Carl Roth	3908.2
glyoxal (40 % in H_2O)	Sigma	G3140
HEPES	Carl Roth	9105.3
hydrochloric acid 37%(w/w)	Carl Roth	X942.1
hydrogen peroxide	Carl Roth	96813
iron (II) sulfate heptahydrate	Merck	2447456
kanamycin sulfate	Serva	26898
lactose monohydrate	Carl Roth	6868.2
L-arabinose	Carl Roth	5118.1
L-rhamnose monohydrate	Carl Roth	4655.1
magnesium chloride hexahydrate	Merck	1725711000
magnesium sulfate heptahydrate	Merck	A466286
maleic acid	Merck	8003800500
manganese (II) sulfate monohydrate	Fluka	63554
manganese (II) chloride tetrahydrate	Carl Roth	T881.1
MES	Carl Roth	4256.3
methanol	Carl Roth	AE71.2
midori green	Nippon Genetics Europe	MG04
MOPS	Carl Roth	6979.3
NAD	Roche	10127981001
ortho phosphoric acid	Merck	563
oxoid agar	Oxoid	CP0011
pectin from citrus	Serva	31650
PEG4000	Carl Roth	0156.1
PEG6000	Merck	8170071000
PEG8000	Carl Roth	0263.2
peptone	Merck	10859
phusion GC buffer	Thermo Fisher Scientific	F530S
polygalacturonic acid	Sigma	P-3889
potassium acetate	Merck	4820.1000
potassium chloride	Carl Roth	6781.1
potassium dihydrogen phosphate	Carl Roth	3904.1
potassium hydroxide	Merck	1050331000
PPG 2000	Merck	8.21037.1000
protein assay dye reagent	Biorad	500-0006
Q5 High GC enhancer	NEB	M0491S
Q5 reaction buffer	NEB	M0491S

QuickExtract™ Plant DNA Extraction Solution	Epicentre	QEP70750
rubidium chloride	Aldrich	R2252
SDS	Merck	13760
sodium chloride	Carl Roth	3957.2
sodium dihydrogen phosphate	Carl Roth	T878.2
sodium hydroxide	Fluka	71689
sodium molybdate dihydrate	Carl Roth	8601.1
sodium nitrate	Carl Roth	8601.1
starch	Fluka	85642
TEMED	Serva	35925
Tris	Carl Roth	5429.3
trisodium citrate	Sigma-Aldrich	I-2752
trizol	Thermo Fisher Scientific	15596018
tween 20	Merck	8221840500
urea	Amersham Biosciences	17-1319-01
uridine	Carl Roth	714.3
xylan (beechwood)	Carl Roth	4414.2
xylene cyanol FF	Fluka	95600
yeast extract	Ohly	10901006
zinc sulfate	Merck	8883

4.1.9.2 Enzymes

Table 4.16: Enzymes used in this study.

enzyme	manufacturer	article number
all restriction enzymes	Thermo Fisher Scientific	
Cas12a	in-house (Kwon et al. 2019)	
lysing enzymes from *Trichoderma harzianum*	Sigma-Aldrich	L1412-25G
Phusion DNA polymerase	Thermo Fisher Scientific	F530S
Q5 DNA polymerase	NEB	M0491S
RNAse A	Carl Roth	7156.1
T5 exonuclease	Epicentre	T5E4111K
Taq DNA ligase	NEB	M0208L
Taq DNA polymerase	in-house	

4.1.9.3 Size markers

Table 4.17: Size markers used in this study.

size marker	manufacturer	article number
Gene ruler 1kb plus DNA ladder	Thermo Fisher Scientific	SM1331
Gene ruler DNA ladder mix	Thermo Fisher Scientific	SM0331
PageRuler unstained protein ladder	Thermo Fisher Scientific	26614

4.1.9.4 Buffers and solutions

All buffers and solutions used in this study were prepared with H_2O MQ unless stated otherwise.

Table 4.18: Buffers and solutions used in this study for DNA/RNA gel electrophoresis and DNA extraction.

buffer/solution	composition
50 x TAE used as 0.5 x	2 M Tris 100 mM EDTA 1 M sodium acetate adjust pH to 8.3 with acetic acid
100 mM sodium phosphate buffer (1L)	39 mL 1 M NaH_2PO_4 61 mL 1M Na_2HPO_4 adjust pH to 7 with NaH_2PO_4 or Na_2HPO_4
glyoxal loading buffer	50 % (v/v) glycerol 10 mM sodium phosphate buffer 0.25 % (w/v) bromophenol blue 0.25 % (w/v) xylene cyanol FF
RNA sample buffer	2 µL 100 mM sodium phosphate buffer 2.3 µL 8.8 M glyoxal 10 µL DMSO
DNA extraction buffer	0.5 % (w/v) SDS 0.2 M Tris HCl, pH 8 0.025 M EDTA, pH 8 0.25 M NaCl

Table 4.19: Buffers and solutions used in this study for transformation of *T. thermophilus*.

buffer/solution	composition
SMC	1.33 M sorbitol 50 mM $CaCl_2x2H_2O$ 20 mM MES buffer, pH 5.8
TC	50 mM $CaCl_2x2H_2O$ 10 mM Tris HCl, pH 7.5
STC	1.33 M sorbitol in TC
PEG 4000 buffer	60 % (w/v) PEG 4000 in TC
protoplastation solution	2 % (w/v) lysing enzymes from *T. harzianum* in SMC, pH 5.6, filter sterilized

Table 4.20: Buffers and solutions used in this study for Southern analysis.

buffer/solution	composition
denaturation buffer	0.5 M NaOH 1.5 M NaCl
neutralization buffer	1.5 M NaCl 0.5 M Tris adjust pH to 7 with HCl
20 x SSC	3 M NaCl 300 mM trisodium citrate adjust pH to 7 with HCl
hybridization buffer	1 M NaCl 1 % (w/v) SDS 10 % (w/v) dextrane sulfate
high stringency buffer	0.5 x SSC 0.1 % (w/v) SDS
low stringency buffer	2 x SSC 0.1 % (w/v) SDS
washing buffer	0.3 % (w/v) Tween 20 in maleic acid buffer
maleic acid buffer	100 mM maleic acid 150 mM NaCl adjust pH to 7.5 with NaOH

10 x blocking buffer	10 % (w/v) blocking reagent in maleic acid buffer
antibody solution	1 µL anti-digoxigenin-AP, Fab fragments in 10 mL 1 x blocking buffer
detection buffer	100 mM Tris HCl, pH 9.5 100 mM NaCl
CDP-star-solution	5 µL CDP-star in 495 µL detection buffer
stripping solution	200 mM NaOH 0.1 % (w/v) SDS

Table 4.21: Buffers and solutions used in this study for SDS-PAGE.

buffer/solution	**composition**
10 x laemmli running buffer	250 mM Tris 1.92 M glycine 1 % (w/v) SDS adjust pH to 8.3
reducing sample buffer	200 mM Tris HCl, pH 6.8 8 % (w/v) SDS 40 % (v/v) glycerol 4 % (v/v) 2-mercaptoethanol 50 mM EDTA 0.08 % bromophenol blue
staining solution	0.1 % (w/v) coomassie brilliant blue R250 10 % (v/v) acetic acid 25 % (v/v) methanol
destaining solution	10 % (v/v) methanol 10 % (v/v) acetic acid

Table 4.22: Buffers and solutions used in this study for the preparation of chemical competent *E. coli* cells.

buffer/solution	**composition**
Tfb I	30 mM KCl 100 mM RbCl 10 mM $CaCl_2x2H_2O$ 50 mM $MnCl_2x4H_2O$ 15 % (v/v) glycerol adjust pH to 5.8 with acetic acid
Tfb II	10 mM MOPS 10 mM RbCl 75 mM $CaCl_2x2H_2O$ 15 % (v/v) glycerin adjust pH to 6.5 with NaOH

Table 4.23: Buffers and solutions used in this study for storage of *T. thermophilus*.

buffer/solution	**composition**
freeze protection solution	10 % (w/v) glycerol 5 % (w/v) lactose

Table 4.24: Buffers and solutions used in this study for Gibson cloning.

buffer/solution	composition
5 x ISO-buffer	3 mL 1M Tris HCl, pH 7.5 150 µL 2 M $MgCl_2$ 60 µL 100 mM dGTP 60 µL 100 mM dTTP 60 µL 100 mM dCTP 60 µL 100 mM dATP 300 µL 1 M DTT 1.5 g PEG 8000 300 µL 100 mM NAD ad H_2O 6 mL
Gibson assembly mix	320 µL 5 x ISO-buffer 0.64 µL 10 U/µL T5 exonuclease 20 µL 2 U/µL Phusion polymerase 160 µL 40 U/µL Taq ligase ad H_2O 1.2 mL

Table 4.25: Buffers and solutions used in this study for CRISPR/Cas12a RNP application.

buffer/solution	composition
CRISPR reaction buffer	20 mM HEPES 150 mM KCl 0.1 mM EDTA 10 mM MgCl2x6H_2O adjust pH to 7.5 with KOH filter sterilize and add 0.5 mM DTT

4.1.9.5 Media

Supplements of the following media, shown in Table 4.26, were sterile filtrated and added to the medium after autoclaving. Stock solutions of glucose, xylose, ASP+/-N, $MgSO_4$, trace elements and $CaCl_2$ were autoclaved separately (except for $MgSO_4$ in the bioreactor medium) and added to the media after autoclaving. Cellulose and PPG 2000 were autoclaved separately and added to the medium after autoclaving. For the preparation of solid media 1.5 % agar-agar was added unless stated otherwise. All media and solutions shown in Table 4.27 and 4.28 were prepared with H_2O dest. unless stated otherwise.

Table 4.26: Supplements of media used in this study with stock and final concentrations.

supplement	stock concentration	final concentration
acetamide	1M	10 mM
ampicillin	50 mg/mL	50 µg/mL
biotin	200 mg/L	6 µg/L
$CsCl_2$	1.5 M	15 mM
kanamycin sulfate	50 mg/mL	50 µg/mL

Table 4.27: Media used for the cultivation of *E. coli* in this study.

medium	composition
LB	0.5 % (w/v) yeast extract 1 % (w/v) peptone 0.5 % (w/v) NaCl adjust pH to 7.2
Psi broth	0.5 % (w/v) yeast extract 2 % (w/v) peptone 0.5 % (w/v) $MgSO_4x7H_2O$ adjust pH to 7.6 with KOH

Table 4.28: Media used for the cultivation of *T. thermophilus* in this study.

medium/solution	composition
bioreactor medium	76 mM $(NH_4)_2SO_4$ 2 mM $MgSO_4x7H_2O$ 12 mM KH_2PO_4 7,5 mM KCl adjust pH to 6.7 with NaOH 0.27 mM $CaCl_2x2H_2O$ 0.025 mM biotin 1 x Tth trace elements 55.5 mM glucose/ 1% (w/w) cellulose/ 55.5 mM xylose
Tth trace elements (1000 x)	134 mM EDTA disodium salt dihydrate 70 mM $ZnSO_4x7H_2O$ 162 mM H_3BO_3 23 mM $MnSO_4xH_2O$ 16.4 mM $FeSO_4x7H_2O$ 6.5 mM $CoCl_2x6H_2O$ 5.8 mM $CuSO_4x5H_2O$ 5.7 mM $Na_2MoO_4x2H_2O$ solve the components separately in the given order, adjust pH to 6 with NaOH after addition of every component except $CuSO_4x5H_2O$
Asp trace elements (1000 x)	34 mM EDTA 15.3 mM $ZnSO_4x7H_2O$ 5 mM $MnCl_2x4H_2O$ 1.3 mM $CoCl_2x6H_2O$ 1.2 mM $CuSO_4x5H_2O$ 0.18 mM $(NH_4)6Mo_7O_{24}x4H_2O$ 10 mM $CaCl_2x2H_2O$ 3.6 mM $FeSO_4x7H_2O$ solve the components separately in the given order, adjust pH to 6 with NaOH after addition of every component, when all components are dissolved, adjust pH to 4 with HCl
Asp+N (50 x)	0.55 M KH_2PO_4 3.5 M $NaNO_3$ 0.35 M KCl adjust pH to 5.5 with KOH
Asp-N (50 x)	Asp+N without $NaNO_3$
MM	1 x Asp+/-N 1 % glucose 2 mM $MgSO_4x7H_2O$ 1 x Asp trace elements
MM-*amdS*	MM with 1 x Asp-N 10 mM acetamide 15 mM $CsCl_2$
MM-FAA	MM with 1 x Asp-N 10 mM urea 0.2 % fluoroacetamide

CM	1 x Asp+N 1 % glucose 2 mM $MgSO_4x7H_2O$ 1 x Asp trace elements 0.1 % casamino acids 0.5 % yeast extract adjust pH to 5.8
transformation plates (*amdS* as selection marker)	0.95 M sucrose 1.2 % oxoid agar 1 x Asp-N 1 x Asp trace elements 2 mM $MgSO_4x7H_2O$ 10 mM acetamide 15 mM $CsCl_2$
top agar (*amdS* as selection marker)	see transformation plates, but with 0.6 % oxoid agar

4.1.10 Software and packages

The following software was used in this study:

- ApE-A plasmid editor
- Carestream MI
- ImageJ
- LasX
- Image Lab
- Biocommand
- Biorad CFX Manager

The following software and packages were used for RNA Seq analysis:

- RStudio (circlize, complexheatmap, devtools, rafalib, rsamtools, BiocParallel, DESeq2, GenomicFeatures, GenomicAlignments, vsn, ggplot2, pheatmap, RColorBrewer, gplots)
- Linux (FastQC, STAR and BBTools)

4.2 Methods

4.2.1 Cultivation and storage of *E. coli* and *T. thermophilus*

4.2.1.1 Cultivation of *E. coli*

E. coli cells were always incubated over night at 37 °C in liquid or solid LB medium, unless stated otherwise. If necessary, antibiotics were added according to Table 4.26. In liquid cultures, shaking speed was set to 180 rpm. If necessary, *E. coli* cells were stored at -20 °C for isolation of plasmids and *E. coli* transformants on plates at 4 °C.

4.2.1.2 Cultivation of *T. thermophilus*

For the generation of spores and mycelium, *T. thermophilus* was cultivated on solid CM medium at 37 °C for 3 days. Cultivation after transformation was carried out on transformation plates with top agar and on MM-*amdS* for subcultivation of the transformants at 37 °C until colonies were visible. Special cultivation conditions, e.g. pre-culture for transformation or bioreactor cultivation will be mentioned later in the respective methods part.

4.2.1.3 Cryopreservation in freeze protection solution

For short-time storage (5-6 months) of *T. thermophilus* strains, approx. 2 cm^2 well grown mycelium of the respective strain were cut out with a scalpel, added to 1.5 mL freeze protection solution, and stored at -80 °C.

4.2.1.4 Generation of spore suspensions and cryopreservation

Spores were generated by cultivating *T. thermophilus* on solid CM medium for 3 days at 37 °C. Spores were harvested by rinsing the mycelium with 5-10 mL 0.9 % NaCl solution and removing the spores with a cotton stick. The resulting spore suspension was filtrated through miracloth to remove the mycelium and afterwards centrifuged for 10 min at 7000 rcf. Spore concentration was set according to the further usage with a cell counting chamber by resuspending the spores in the respective amount of 0.9 % NaCl solution.
To generate cryoprotected spore suspensions, spores were washed with 5 mL of 20 % glycerol and centrifuged as described above. To generate a WCB (working cell bank), a spore suspension which is only thawn once and afterwards used for 1-2 weeks, the spore concentration was set to ~1 x 10^7 spores/mL in 20 % glycerol. To generate a MCB (master cell bank), a spore suspension for long term storage which served as a backup for generating new MCBs, spores were resuspended in 1 mL of 20 % glycerol after the washing step. The resulting suspension was diluted 1/10 with 80 % of glycerol. WCB and MCB were frozen in liquid nitrogen and stored at -80 °C.

4.2.2 Microbiological and molecular biological methods

4.2.2.1 DNA isolation and analysis

4.2.2.1.1 Isolation of plasmid DNA from *E. coli*

Plasmid DNA from *E. coli* was isolated using innuPREP Plasmid Mini Kit 2.0 or PureYield™ Plasmid Midiprep System according to the manufacturer instructions.

4.2.2.1.2 Isolation of genomic DNA from *T. thermophilus*

Genomic DNA was isolated according to SOP_017 (based on Arentshorst et al. 2012 with small modifications). 3 mL of liquid CM medium were inoculated with mycelium or spores in a 15 mL reaction

tube and incubated horizontally at 37 °C for 2-3 days. Grown mycelium was harvested into a 2 mL reaction tube using a cotton stick. The reaction tube was sealed with a cap with small holes, frozen at-80 °C and freeze dried. The freeze-dried mycelium was ground to a fine powder (max. 500 µL volume) using a cotton stick, resuspended in 500 µL DNA extraction buffer and heated at 65 °C for 15 min. After that the suspension was cooled down on ice for 5 min. 100 µL 8 M potassium acetate was added, and the suspension was mixed by inverting the reaction tube 8-10 times. Thereafter a centrifugation at 13000 rpm for 15 min followed. The supernatant was transferred into a new reaction tube and the addition of potassium acetate as well as the centrifugation were repeated. DNA was precipitated by addition of 300 µL of isopropanol to the supernatant with following centrifugation at 13000 rpm for 15 min. The resulting pellet was washed with 1 mL 70 % ethanol (without resuspending) and centrifuged as described above. The supernatant was discarded, the pellet was dried at 42 °C for 1 h and resuspended in 50 µL of H_2O MQ containing 2 µL of 10 mg/mL RNAse A. After incubating this solution at 65 °C for 30 min, an agarose gel electrophoresis with 3 µL of DNA was performed to check the quality of the DNA isolation. Isolated DNA was stored at -20 °C.

4.2.2.1.3 Restriction analysis of DNA

Restriction analysis of genomic or plasmid DNA was performed using restriction enzymes from Thermo Fisher Scientific according to the manufacturer instructions.

4.2.2.1.4 Agarose gel electrophoresis

The separation of DNA fragments was performed according to SOP_006 using 0.7-1 % agarose gels (according to the expected size of the fragments). Agarose gels were produced by solving agarose in 0.5 % TAE buffer via heating in the microwave. 1 µL / 100 mL midori green was added for visualization of the DNA under UV radiation. 6 x DNA loading dye was added to the samples prior to loading on the gel (if necessary). Markers (Table 4.17) were used as a reference for DNA fragments sizes. The electrophoresis was performed in an electrophoresis chamber filled with 0.5 % TAE buffer using a voltage of 80-100 V. The visualization of DNA under UV radiation was performed using the Gel logic 212pro system.

4.2.2.1.5 Isolation of DNA from agarose gels and PCR reaction mixtures

DNA was isolated from agarose gels and PCR mixtures using innuPREP DOUBLEpure Kit according to the manufacturer instructions.

4.2.2.1.6 Determination of DNA concentration

DNA concentration was determined with the BioSpectrometer according to the manufacturer instructions using 3.5 µL sample volume.

4.2.2.1.7 DNA sequence analysis

DNA sequence analysis was performed externally by LGC Genomics. DNA preparation was performed according to recommendations from LGC.

4.2.2.2 Cloning experiments with bacteria, fungi, and plasmids

4.2.2.2.1 Polymerase chain reaction (PCR)

PCR was performed according to SOP_007 with modifications. According to the situation different polymerases were used. Taq polymerase was used for analysis of *E. coli* transformants and for amplification of DIG labelled probes for Southern analysis. Q5 polymerase was used for circular polymerase extension cloning. Phusion polymerase was used for all other PCR reactions, including, e.g. analysis of *T. thermophilus* transformants. In the following tables, detailed information about general PCR reaction mixes and PCR programs is provided. Special variants of the PCR will be mentioned later together with detailed information about their PCR reaction mixes and programs in the respective methods part. Annealing temperatures for PCR reactions with Taq or Phusion polymerase were determined using the T_m calculator from Thermo Fisher Scientific and for PCR reaction with Q5 polymerase via the T_m calculator from New England Biolabs.

Table 4.29: Composition of PCR reaction mix for analysis of *E. coli* transformants.

reagent	amount
10 x DreamTaq Green Buffer	2 µL
10 mM dNTPs (2.5 mM each)	1.6 µL
10 µM primer	1 µL each
Taq polymerase	0.1 µL
template	1 µL of *E. coli* colony transferred to 20 µL H_2O MQ
ad H_2O MQ 20 µL	

Table 4.30: Composition of PCR reaction mix for amplification of DIG labelled probes for Southern analysis.

reagent	amount
10 x DreamTaq Green Buffer	2.5 µL
DIG labelled dNTPs	2.5 µL
10 µM primer	1 µL each
Taq polymerase	0.25 µL
5-50 ng template	0.5 µL
ad H_2O MQ 25 µL	

Table 4.31: PCR program for reactions with Taq polymerase. Annealing temperature (x) was determined using T_m calculator from Thermo Fisher Scientific.

step	temperature	time	cycles
initial denaturation	94 °C	3 min	1
denaturation	94 °C	20 sec	30
annealing	x	20 sec	
extension	72 °C	1 min/kb	
final extension	72 °C	5 min	1
storage	4 °C	∞	

Table 4.32: Composition of PCR reaction mix for reactions with Phusion polymerase.

reagent	amount
5 x Phusion GC buffer	10 µL
10 mM dNTPs (2.5 mM each)	4 µL
10 µM primer	2.5 µL each
Phusion polymerase	0.5 µL
5-50 ng template	0.5 µL
ad H_2O MQ 50 µL	

Table 4.33: PCR program for reactions with Phusion polymerase. Annealing temperature (x) was determined using T_m calculator from Thermo Fisher Scientific.

step	temperature	time	cycles
initial denaturation	98 °C	5 min	1
denaturation	98 °C	30 sec	30
annealing	x	30 sec	
extension	72 °C	30 sec/kb	
final extension	72 °C	10 min	1
storage	4 °C	∞	

4.2.2.2.2 Circular polymerase extension cloning (CPEC)

Circular polymerase extension cloning (Quan and Tian 2011) was used for cloning without introducing additional DNA sequences (seamless cloning). Prior to CPEC, the different fragments that should be assembled had to be amplified with ~20 bp overlapping sequences by using the Phusion polymerase. To reduce the number of fragments, fusion PCRs according to Table 4.32 and Table 4.33 with 50 ng of each fragment as template were performed. The resulting fragments were used for CPEC. The PCR reaction mix and PCR program used to assemble the fragments with CPEC are shown in Table 4.34 and Table 4.35. 200 ng of the largest fragment and inserts with a molar ratio of 1:1 compared to the largest fragment were used. The CPEC product was digested with 1 µL DpnI for selective digestion of methylated and hemimethylated DNA and 5 µL of the product was transformed into chemical competent *E. coli* cells.

Table 4.34: Composition of CPEC reaction mix using Q5 polymerase.

reagent	amount
5 x Q5 reaction buffer	10 µL
5 x GC Enhancer	10 µL
10 mM dNTPs (2.5 mM each)	4 µL
Q5 polymerase	1 µL
fragments	200 ng BB+5´flank A
	molar ratio 1:1 for other fragments
ad H_2O MQ 50 µL	

Table 4.35: PCR program for CPEC reaction with Q5 polymerase. Annealing temperature (x) was determined using T_m calculator from New England Biolabs.

step	temperature	time	cycles
initial denaturation	98 °C	5 min	1
denaturation	98 °C	30 sec	30
annealing	x	30 sec	
extension	72 °C	10 min	
final extension	72 °C	10 min	1
storage	4 °C	∞	

4.2.2.2.3 Gibson cloning

Gibson cloning was performed according to SOP_013 and Gibson et al. 2009. It was used for cloning without introducing additional DNA sequences (seamless cloning) as an alternative to CPEC. Prior to Gibson cloning, the different fragments that should be assembled had to be amplified with ~20 bp overlapping sequences by using the Phusion polymerase. The resulting fragments were used for Gibson cloning. 12 ng/ 1000 bp of the backbone fragment and inserts with a molar ratio of 4:1 for fragments of 1/2-1/4 size and 8:1 for fragments of < 1/4 size in relation to the longest fragment were mixed in a total volume of 5 µL. Afterwards 15 µL of the Gibson assembly mix were added, followed by a 60 min incubation at 50 °C. Finally, after cooling down the reaction mix at RT for 3 min and 3 min on ice, 5 µL of the assembly reaction were transformed into chemical competent *E. coli* cells.

4.2.2.3 Transformation experiments with *E. coli* and *T. thermophilus*

4.2.2.3.1 Preparation of chemical competent *E. coli* cells

Preparation of chemical competent *E. coli* cells was performed according to SOP_004. As a preculture, 20 mL of Psi broth were inoculated with *E. coli* Top 10 cells in a 100 mL Erlenmeyer flask and incubated overnight at 250 rpm and 37 °C. The next day, 100 mL of Psi broth were inoculated with 1 mL of the preculture in a 500 mL Erlenmeyer flask and incubated at 250 rpm and 37 °C until the OD_{600} (determined via BioSpectrometer) reached 0.48. The culture was incubated on ice for 15 min and centrifuged at 4 °C and 5000 rpm for 5 min. The supernatant was removed, and the pellet carefully resuspended in 40 mL ice cold TfbI and incubated on ice for 15 min. Centrifugation was repeated as described above. Supernatant was removed and the pellet was carefully resuspended in 3 mL ice cold TfbII. Finally, 150 µL aliquots were prepared and immediately frozen in liquid nitrogen. The aliquots were stored at -80 °C.

4.2.2.3.2 Transformation of chemically competent *E. coli* cells

Transformation of chemically competent *E. coli* cells was performed according to SOP_005. Up to 5 µL of plasmid DNA (total amount 1-10 ng) were added to 50 µL of chemically competent *E. coli* cells. As a negative control, H_2O was used instead of plasmid DNA. The suspension was incubated on ice for 30 min, heat shocked at 42 °C for 40 sec, and incubated on ice for 1 min. After that, 1 mL LB medium was added and the suspension was incubated at 800 rpm and 37 °C for 1 h. 100 µL of undiluted as well as 100 µL of 1/10 diluted suspension were plated on the respective selective LB agar. For the negative

control, all cells were plated on LB plates containing the respective antibiotic. The cells were grown over night at 37 °C and analyzed afterwards.

4.2.2.3.3 Analysis of *E. coli* clones

At first, *E. coli* transformants gained from cloning experiments were analyzed for the integration of the desired plasmid via colony PCR as described in Table 4.29 and Table 4.31. Afterwards, plasmid DNA was isolated from putative positive clones and analyzed via restriction analysis. If restriction analysis was positive, the plasmid DNA was sequenced by LGC Genomics (Berlin). Primers used for the sequencing of the generated plasmids are listed in Table 4.8 and Table 4.9. The cloning was considered successful when sequencing results were congruent with the expected sequence. *E. coli* clones gained from the re-transformation of an existing plasmid were not additionally analyzed. Growth on LB plates containing the respective antibiotic and no growth of the negative control was considered as sufficient.

4.2.2.3.4 Transformation of *T. thermophilus* via split marker approach

T. thermophilus was transformed according to SOP_003 with modifications, using an adapted protocol for PEG mediated protoplast transformation (Arentshorst et al. 2012; Kwon et al. 2019). At first, 250 mL of CM were inoculated with 2.5 x 10^8 spores and incubated for 12-13 h at 37 °C and 120 rpm. After the incubation, mycelium was harvested by filtration through miracloth and washed with SMC. To protoplast the mycelium, 0.5–1 g (wet weight) of mycelium was transferred to protoplastation solution and incubated at 37 °C for 1.5–3 h. Protoplast formation was checked under the microscope and considered sufficient when protoplasts were formed from 70–80 % of mycelium. The protoplasts were collected in a 50 mL reaction tube through miracloth and rinsed with STC up to 50 mL. The resulting suspension was centrifuged at 2000 g for 10 min at 4 °C. The supernatant was discarded, and the pellet was carefully resuspended in 1 mL of STC and centrifuged for 5 min at 3000 g and 4 °C. This washing step was repeated twice. After washing, the protoplasts were resuspended in STC (max. 1 mL) according to the amount of transformations planned and stored on ice until the transformation started. For each transformation the following components were added to a 50 mL reaction tube in the following order: 100 µL protoplasts, 1 µL 0.5 M ATA, 3-5 µg of each split marker fragment (max. volume 10 µL) and 25 µL of PEG 4000 buffer. As a negative control H_2O and as a positive control 3-5 µg of MT28 was used instead of split marker fragments. Thereafter, 1 mL of PEG 4000 buffer was added. After exactly 5 min of incubation at room temperature, 2 mL of STC was added to the transformation reaction mix and gently mixed by tipping. Top agar was added to a total volume of 30 mL and mixed by inverting the reaction tube several times. The suspension was poured onto transformation plates. The plates were incubated at 37 °C for 4-10 days until colonies were visible. Primary transformants were subcultured twice on MM-*amdS* via dilution streaking. The resulting clones were analyzed via PCR and Southern analysis.

4.2.2.3.5 Transformation of *T. thermophilus* via CRISPR/Cas12a

Transformation of *T. thermophilus* via CRISPR/Cas12a was performed as described above with the following modifications according to Kwon et al. 2019. The DNA template for the sgRNA was amplified

via Phusion polymerase with two oligonucleotides as template (listed in Table 4.7) and transcribed to RNA using the MegaScript Kit according to the manufacturer instructions. Prior to the transformation, formation of the RNP complex was started by mixing 5 µL Cas12a (~20-30 µg), 2 µL CRISPR reaction buffer, 1 µL of sgRNA, and 12 µL of nuclease free H_2O. The reaction mix was incubated at 37 °C for 15 min and stored at room temperature (1-2 h) before starting the transformation. For each transformation the following components were added to a 50 mL reaction tube in the following order: 100 µL protoplasts, 3-5 µg whole plasmid DNA (max. volume 20 µL), 20 µL RNP complex, 30 µL 2 x STC, 25 µL PEG 4000 buffer, and 20 µL CRISPR reaction buffer. All previous and following steps were performed as described in 4.2.2.3.4.

4.2.2.3.6 Analysis of clones via PCR

Two different types of PCR were used to analyze transformants of *T. thermophilus*. The first type is a standard PCR with the Phusion polymerase (see Table 4.32 and Table 4.33) and gDNA of the respective clones as template. The second type of PCR used, was a colony PCR with template gained by using QuickExtract™ Plant DNA Extraction Solution as follows: the template was prepared by resuspending small amounts of mycelium (should be hardly visible) in 25 µL of QuickExtract™ Plant DNA Extraction Solution and incubating the suspension for 6 min at 65 °C and 2 min at 98 °C. After repeating the incubation step, the suspension was stored at 12 °C (1-2 h prior to the PCR reaction). In the following tables, detailed information about the PCR reaction mix and the PCR program used is provided. Primers used for both types of PCR are listed in Table 4.5.

Table 4.36: Composition of colony PCR reaction mix using Phusion polymerase.

reagent	amount
5 x Phusion GC buffer	6 µL
10 mM dNTPs (2.5 mM each)	2.4 µL
10 µM primer	1 µL each
Phusion polymerase	0.3 µL
4 M betaine monohydrate	7.5 µL
template (as described in the text)	5 µL
ad H_2O MQ 30 µL	

Table 4.37: PCR program for colony PCR reactions with Phusion polymerase. Annealing temperature was determined using T_m calculator from ThermoFisher Scientific.

step	temperature	time	cycles
initial denaturation	98 °C	5 min	1
denaturation	98 °C	30 sec	20
annealing	67 °C*	20 sec	
extension	72 °C	1 min/kb	
denaturation	98 °C	30 sec	20
annealing	61 °C*	20 sec	
extension	72 °C	1 min/kb	
final extension	72 °C	10 min	1
storage	4 °C	∞	

* depends on the annealing temperatures of the primers, used primers were designed for the usage with the described temperatures.

4.2.2.3.7 Analysis of clones via Southern blot

Southern blot analysis was performed according to SOP_010. For Southern analysis gDNA of the strain to be examined was first digested via restriction enzymes. The next day, digested DNA was separated via agarose gel electrophoresis using a 0.7 % agarose gel. Markers used as a reference for fragment sizes are listed in Table 4.17. After electrophoresis, the gel was irradiated by UV light for 5 min to induce double strand breaks to facilitate transfer of big DNA fragments. After that, the gel was incubated in denaturation buffer for 20 min under continuous agitation, changing denaturation buffer after 10 min. The denaturation solution was poured off and the gel was rinsed with H_2O dest. The above-mentioned incubation was repeated with neutralization buffer. The Southern blot was assembled in the following order: clingfilm, three sheets of Whatman paper soaked with 10 x SSC, gel, nylon membrane soaked with 10 x SSC, three sheets of dry Whatman paper, paper towels, and weights to exert pressure. The DNA was transferred to the nylon membrane overnight using capillary forces. The next day the blot was disassembled and the membrane dried at room temperature for 5 min. 10 ng of a positive control (plasmid DNA containing a sequence that is complementary to the used probe) was dropped to the right upper edge of the membrane and the DNA was crosslinked via irradiating the membrane by UV light for 2 min. The membrane was placed into a hybridization tube and prehybridized by adding 25 mL of preheated (65 °C) hybridization buffer. The incubation was performed rotating in a hybridization oven for 1.5 h at 65 °C. The DIG labelled probes were prepared according to Table 4.30 and Table 4.31 with primers listed in Table 4.6. After isolating the probe from the PCR reaction mix, 200 ng were added to 50 µL H_2O dest., boiled at 95 °C for 5 min, chilled on ice and added to 20 mL hybridization buffer (65 °C). The old hybridization buffer used for prehybridization was exchanged with the hybridization buffer containing the probe and incubation was continued over night at 65 °C. The next day the hybridization buffer containing the probe was poured into a 50 mL reaction tube and stored at -20 °C in case Southern analysis needs to be repeated (buffer can be reused 3-5 times). 20 mL of high stringency buffer (65 °C) was added to the membrane and discarded after inverting the hybridization tube several times. Again, 20 mL of high stringency buffer (65 °C) was added followed by a rotating incubation at 65 °C for 20 min. This step was repeated with low stringency buffer (65 °C). After that, the membrane was washed with 20 mL washing buffer for 2 min at room temperature in the hybridization oven. Washing buffer was replaced with 10 mL 1 x blocking buffer and incubation continued for 1 h at room temperature. The blocking buffer was exchanged with 10 mL antibody solution and incubation continued for 30 min at room temperature. Thereafter, the membrane was washed twice with 20 mL washing buffer for 15 min rotating at room temperature. After washing, the membrane was equilibrated in 10 mL detection buffer for 3 min while rotating at room temperature. The membrane was placed into a plastic bag, CDP star solution was applied to the membrane and distributed over the whole surface by quickly running the hand over the plastic bag on a smooth surface. Excess liquid was squeezed out and removed with paper towels. Detection and analysis of chemiluminescence was performed using the ChemiDoc MP system and Image Lab program. Clones obtained via split marker transformation were analyzed using one restriction enzyme and two different probes (5´flank and 3´flank probe), each probe complementary to one of the two single split marker fragments used. After detection of chemiluminescence of the first probe, the membrane was stripped by incubating it at 37 °C for 20 min rotating in stripping solution. This

step was repeated with 15 min of incubation. After stripping, the membrane was washed with 2 x SSC by inverting the hybridization tube several times and Southern analysis with the second probe was started with the prehybridization step. Clones obtained via transformation assisted by CRISPR/Cas12a were analyzed using one probe (5´ flank) and two different restriction enzymes. Results of Southern analysis performed in this study can be found in the appendix.

4.2.2.3.8 Marker removal via FAA

After proving the successful deletion of the GOI, the integrated selection marker cassette was removed via counterselection on MM-FAA. Loss of *amdS* prevents the conversion of FAA into a toxic compound, allowing the fungus to grow on the medium. To achieve this, ~2 * 10^7 spores were plated on MM-FAA and incubated at 37 °C for 3-6 days (Arentshorst et al. 2012). Single colonies were subcultured for a second time on MM-FAA and analyzed afterwards via PCR and Southern analysis for the absence of the deletion marker cassette as mentioned above.

4.2.2.3.9 Separation of mixed strains

In some cases, wildtype genotype as well as mutant genotypes were observed in the same clone, meaning this clone is a mixture of wildtype and the desired mutant. To separate these mixed clones, a defined number of spores (10-100) was plated onto the appropriate medium and the resulting colonies were analyzed via PCR and Southern analysis as described above. This procedure was repeated until pure strains were obtained.

4.2.2.4 Quantitative polymerase chain reaction (qPCR)

Quantitative polymerase chain reaction was performed according to SOP_011. First, isolated, purified, and DNAse treated RNA (as described above) was transcribed into cDNA using the RevertAid H Minus First Strand cDNA Synthesis Kit. The reaction mix and the PCR program used for the following qPCR are listed in the following tables. As controls a no template control (NTC) and a no reverse transcriptase control (NRT) were used. The qPCR reaction was performed in a Biorad CFX96 cycler, using the Biozym Blue S´Green qPCR Kit and analyzed with the help of Biorad CFX Manager.

Table 4.38: Composition of a qPCR reaction mix.

reagent	amount
Blue S´Green qPCR 2x Mix	10 µL
100 µM primer	0.1 µL each
cDNA template (1-5 ng)	4 µL
ad H_2O MQ 20 µL	

Table 4.39: Program used in this study for qPCR.

step	temperature	time	cycles/detection
initial denaturation	98 °C	2 min	1
denaturation	95 °C	15 sec	40(detection)
annealing/extension	60 °C	20 sec	
melt curve	55-95 °C (0.5 °C increments)	5 sec per increment	1(detection)
storage	12 °C	∞	1

4.2.2.5 Growth assay on different carbon sources

To investigate the effects of the regulator knockout mutants generated, a growth assay using different carbon sources was performed. The medium used was MM with a spatula tip of bromophenolblue as a pH indicator and different carbon sources. For mono- and disaccharides a final concentration of 25 mM and for polysaccharides a final concentration of 1 % in the medium were used. Mono- and disaccharides were sterile filtrated and added to the medium after autoclaving. Galacturonic acid, polygalacturonic acid, and pectin were added prior to autoclaving and pH was adjusted to ~ pH 5.5. Cellulose and xylan were added to the medium before autoclaving without adjusting the pH value. 1000 spores (in 10 µL) of each strain were spotted on the plate and incubated at 37 °C for 4 days.

4.2.3 Biochemical methods

4.2.3.1 Determination of protein concentration (Bradford assay)

Protein concentration was determined according to SOP_025 via the Bio-Rad Protein Assay Kit II. As a standard, a BSA concentration of 10–1000 µg/mL was used. 100 µL of 1/5 diluted reagent were added to 10 µL of the undiluted sample or standard. After 5 min of incubation at room temperature, absorption was measured at 595 nm.

4.2.3.2 SDS PAGE

Gels for SDS PAGE were prepared as described in Table 4.40, using a 12.5 % resolving gel and a 5 % stacking gel. If necessary, gels were stored at 4 °C upon usage. Samples were prepared via using a defined volume (20 µL) or a defined amount (2 µg) of protein. This was achieved by freeze drying the respective volume and amounts of the sample and adding 10 µL of H_2O MQ. Afterwards, 4 µL of reducing sample buffer was added, the solution was incubated at 95 °C for 5 min and loaded on the gel. As a marker for determination of protein sizes, 6 µL of PageRuler unstained protein ladder was used. Proteins were separated in 1 x laemmli running buffer using 200 V for 30–60 min. After that, the gel was incubated for 1-2 h in staining solution and destained for 2-3 h in destaining solution. The process of staining and destaining can be accelerated via heating in the microwave.

Table 4.40: Composition of the resolving gel (12.5 %) and the stacking gel (5 %) used for SDS PAGE. Components were added in the following order, APS and TEMED were added after the previous components have been thoroughly mixed. The given amounts are sufficient for 2 gels.

resolving gel (12.5 %)	component	stacking gel (5 %)
3.12 mL	40 % acrylamide/bisacrylamide	500 µL
2.5 mL	1.5 M Tris pH 8.8	
	0.5 M Tris pH 6.8	1 mL
0.1 mL	10 % SDS	40 µL
4.28 mL	H_2O	2.56 mL
40 µL	10 % APS	40 µL
10 µL	TEMED	10 µL

4.2.3.3 Determination of glucose/xylose concentration

Glucose concentration was determined using the Glucose Fluid GOD-PAP Kit. As a standard, a glucose concentration of 0.1–1 g/L was used. 100 µL of the reagent were added to 10 µL of the sample or standard. Samples were diluted where necessary. After 30 min of incubation at room temperature, absorption was measured at 505 nm. Xylose concentration was determined using the D-Xylose Assay Kit according to the manufacturer's instructions.

4.2.4 Bioreactor cultivation

4.2.4.1 Bioreactor preparation

After assembling the Bioflo & Celligen 310 bioreactor according to the manufacturer instructions, the pH probe was calibrated with pH reference solutions (pH 4 and 7). After autoclaving, all parts of the reactor were attached to the control and analytical units and filled with 5 kg of bioreactor medium. Analytical components used are listed in Table 4.41. Functionality of the reactor was tested, and the DO probe was calibrated with air and nitrogen. pH regulation was achieved via 25 % ammonia solution and 20 % phosphoric acid. Before inoculation, a sample was taken to compare the internal pH with an external pH meter. The pH value was adjusted if necessary. Spore suspensions for the inoculation of the bioreactor were prepared as previously described, adjusting spore concentration to a value that enables inoculation of the reactor (final spore concentration: $1*10^9$ spores/L) with a maximum volume of 10-20 mL via a syringe. To prevent foaming PPG 2000 was added with a rate of 20 mg/h starting from the end of the batch phase.

Table 4.41: Analytical components used in all bioreactor experiments with their respective purpose.

component	purpose
off-gas analyzer/sequencer	measuring O_2 and CO_2
temperature sensor	measuring temperature
pH probe	measuring pH value
DO probe	measuring dissolved oxygen
balances	measuring weight of bioreactor and amount of base/acid added
bioreactor control unit	monitor and control all fermentation parameters, e.g. aeration, stirring, temperature, pH value (via pumps), weight

4.2.4.2 Batch cultivation

Prior to inoculation, the temperature was set to 45 °C, pH value to 6.7 (if necessary), stirring to 100 rpm, and aeration to 0.01 slpm (after reaching 100 % DOT). After inoculation, a time-based profile for aeration and stirring was started. Aeration and stirring were constantly increased within 10 h to 1 slpm and 750 rpm, respectively. Samples were taken for biomass, protein and glucose determination, RNA extraction, and microscopy. Batch cultivation was continued until glucose was consumed.

4.2.4.3 Chemostat cultivation

Prior to chemostat cultivation a batch cultivation had to be performed as described above. At the end of the exponential phase when glucose concentration was close to 0-1 g/L (~13 g base addition) chemostat cultivation was started by addition of medium at a dilution rate of 0.1 1/h and maintaining culture broth weight at 5 kg by releasing excess culture broth. Release of excess culture broth was achieved via a magnetic valve and a constant bottom aeration of 1 L/min to prevent clogging by biomass accumulation. During the whole chemostat cultivation, glucose was the limiting component and thereby, not detectable. Chemostat cultivation was continued until steady state was achieved. A steady state was defined as a condition with constant base addition, biomass, off-gas data as well as morphology (no pellets, no spores, filamentous growth) without major changes over a timespan of at least one retention time. Samples were taken for biomass, protein and glucose determination, RNA extraction, and microscopy.

4.2.4.4 Cellulose spiking experiment

The cellulose spiking experiment was started after achieving steady state during chemostat cultivation. The chemostat program and the addition of new medium were stopped, and 500 g culture broth was released. After that, 450 g new bioreactor medium (without glucose) containing 50 g microcrystalline cellulose (autoclaved separately and mixed before the addition to the bioreactor) was added to the culture broth resulting in 5 kg total weight of the culture broth and the cultivation was continued as a batch cultivation. Samples for biomass, protein and glucose determination, RNA extraction, and microscopy were taken before and after (0.5 h, 1 h, 2 h, 4 h) the addition of the medium containing cellulose. Cultivation was continued for up to 24 h.

4.2.4.5 pH shift experiment

For the pH shift experiment, a modified batch cultivation was run. In mid-exponential phase at a certain biomass value (determined via biomass to base addition relations of previous bioreactor runs) the t= 0 sample was taken, the pH was lowered to pH 5 with 1M HCl, and pH regulation was set to pH 5. Additional samples were taken 0.5 h, 1 h and 2 h after sample t= 0 was taken.

4.2.4.6 Sampling, analysis, and storage

Culture broth samples were taken using the sampler. Every sample taken was split into different aliquotes according to the planned analysis. For microscopy, a small amount of the culture broth was

transferred to a reaction tube. The biomass sample was gained by filtrating the culture broth sample via vacuum filtration, collecting mycelium on a filter. This filter was wrapped in tinfoil, freeze dried, and weighted before and after filtration. Via comparison of the weight of the filter wrapped in tinfoil before and after filtration and the weight of the culture broth sample that was filtrated, dry weight could be determined. The supernatant of the culture broth sample, separated from the mycelium via the above-mentioned filtration, was used for protein and glucose determination. Samples for RNA isolation were taken as described for the biomass sample. In contrast to the biomass sample, the filter carrying the mycelium wrapped in tinfoil did not have to be weighted and was immediately frozen in liquid nitrogen after taking the sample. All samples taken were stored at -80 °C.

4.2.5 RNA isolation and analysis

4.2.5.1 RNA Isolation

RNA was isolated according to SOP_011. One quarter of a sample for RNA isolation as described above was ground to a fine powder with a precooled mortar and pestle in liquid nitrogen. The powder (200-400 µL) was transferred to a precooled 2 mL reaction tube and 800 µL of Trizol (4 °C) were added immediately. After that the suspension was vortexed until the whole powder was resuspended. If necessary, storage on ice was possible until all samples were ready for further extraction. The samples were vortexed again for 15 sec and incubated at room temperature for 5 min. Thereafter, 150 µL of chloroform were added and the sample was shaken by hand for 15 sec. An incubation at room temperature for 3 min and a centrifugation at 4 °C for 15 min at max. speed followed. The resulting colorless upper phase was transferred to a 1.5 mL reaction tube. 500 µL of isopropanol were added, an incubation at room temperature for 10 min, vortexing, and a centrifugation at 4 °C for 10 min at max. speed followed. The supernatant was removed with a pipette and the pellet was washed with 150 µL of 75 % (v/v) ethanol without redissolving the pellet. After the following centrifugation at 4 °C for 5 min at max. speed, the supernatant was removed with a pipette and the pellet was dried at 37 °C for 10 min. The pellet was dissolved in 100 µL of H_2O MQ at 60 °C for 10 min and aliquots of 10 µL were prepared to avoid multiple freeze and thaw cycles. The samples were stored at -80 °C prior to their usage.

4.2.5.2 RNA gel electrophoresis

RNA gel electrophoresis was performed according to SOP_011. For RNA gel electrophoresis an 0.6 % agarose gel was prepared with 10 mM sodium phosphate buffer. After solving the agarose by heating in the microwave, 1 µL/100 mL of midori green was added. Samples were prepared by adding 5.7 µL of the isolated RNA to RNA sample buffer and incubating at 50 °C for 60 min. 2 µL of glyoxal loading buffer were added after cooling down the samples on ice. Finally, 15 µL of the sample were loaded onto the agarose gel. The electrophoresis was performed in a cooled (10 °C) electrophoresis chamber filled with 10 mM sodium phosphate buffer using a circulation pump and 0.5 A for 30-45 min. The visualization of RNA under UV radiation was performed using the Gel logic 212pro system.

4.2.6 RNA sequencing and analysis

4.2.6.1 Preparations of RNA samples for RNA sequencing

Prior to sending the samples for RNA sequencing the quality of the isolated RNA was checked via RNA gel electrophoresis. RNA quality was considered as good when two sharp bands (28S rRNA and 18S rRNA) were visible without smears indicating degradation. After proving a good quality, the samples were purified using the innuPREP RNA Mini Kit 2.0 according to the manufacturer instructions. Possible remaining DNA was removed via DNA-free™ DNA Removal Kit according to the manufacturer instructions. After purification and DNAse treatment, samples were ready for RNA sequencing. Prior to RNA sequencing at GenomeScan (Leiden), no purification and DNAse treatment was performed.

4.2.6.2 RNA sequencing

RNA sequencing was done at Microsynth AG (Balgach) and at GenomeScan (Leiden) using an Illumina platform with 150 bp reads paired end, polyA enrichment, and > 5 million reads per sample. A quality check of the RNA samples prior to sequencing was included according to the guidelines of the company.

4.2.6.3 RNA sequencing analysis

Obtained read data were first quality controlled via FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and if necessary, trimmed with BBTools (https://jgi.doe.gov/data-and-tools/bbtools/). STAR (Dobin et al. 2013) was used for mapping the reads to the *T. thermophilus* ATCC 42464 genome (assembly ASM22609v1, downloaded from NCBI (https://www.ncbi.nlm.nih.gov/). Data normalization and differential gene expression analysis was performed with DEseq2 (Love et al. 2014). Differential gene expression was evaluated with Wald test and Benjamini and Hochberg False Discovery Rate (FDR) with a threshold of 0.05 (Benjamini and Hochberg 1995). Enrichment analyses were performed with DAVID (Huang et al. 2009a; Huang et al. 2009b). Annotations were obtained from DAVID (see above), NCBI (see above), JGI (https://jgi.doe.gov/), and publications that are named in the respective section. The most important source for the annotations were the publications of Berka et al. 2011 and Karnaouri et al. 2014a. Venn analysis was done with the help of http://bioinformatics.psb.ugent.be/webtools/Venn/. Packages used for analysis, creating plots, and diagrams in R via RStudio (https://rstudio.com/) are listed in chapter 4.1.10.

5. Results

In the following the results acquired in this thesis are presented. At firs,t the establishment of molecular biological methods and its application to obtain deletion mutants are shown. Afterwards, establishment and performance of bioreactor cultivations under various conditions with following biochemical analysis are described. Finally, the results of the transcriptomic analysis of the samples obtained during the above-mentioned bioreactor cultivations, especially regarding protease, CAZY, and transcription factor expression are presented.

5.1. Establishment of molecular biological methods, deletion of cellulase regulators, and effects of the deletion on growth using different carbon sources

5.1.1 Establishment of molecular biological methods for working with *T. thermophilus*

In the following the establishment is briefly described, since this organism was never used before in this laboratory. Data for these processes will not be shown. First, growth of *T. thermophilus* was tested, using recipes of growth media received from BASF and growth media used for the cultivation of *A. niger* in this laboratory. Finally, growth media used for the cultivation of *A. niger* were chosen due to logistic reasons. Growth temperature was chosen according to the cultivation of industrial isolates of *T. thermophilus*.

Before starting the genetic modification of *T. thermophilus*, experimental methods for the targeted modification of the *T. thermophilus* genome had to be tested and optimized. The established transformation protocol for the industrial *T. thermophilus* strains that was received from the company BASF, was combined with the existing *A. niger* transformation protocol and resulted, after several optimizations done by Min Jin Kwon, in the final protocol that is described in the publication of Kwon et al. 2019 and this thesis. The general deletion strategy based on a split marker approach via homologous recombination of a removable *amdS* integration cassette (Kelly and Hynes 1985; Nielsen et al. 2006) and experiences with the industrial isolate. In this selection marker cassette, *amdS* from *A. nidulans* is expressed constitutively via the *gpdA* promoter from *A. nidulans*. After successful deletion of the GOI via integration of *amdS* (enables growth on acetamide), the marker was removed via FAA plating (growth on FAA only in absence of *amdS*). A scheme of the split marker deletion strategy is shown in Figure 5.1 A. During this thesis, a new CRISPR/Cas12a based strategy for genome editing was established and optimized by Min Jin Kwon (Kwon et al. 2019). This strategy was used for gene deletions besides the split marker approach. The deletion principle with *amdS* a selection marker and marker recycling via FAA plating remained identical to the split marker approach. A scheme of the deletion strategy using the CRISPR/ Cas12a approach is shown in Figure 5.1 B. Which strain was generated with which strategy will be described in the next chapter. The first gene that was deleted in *T. thermophilus* was *ku80* (performed by Dr. Min Jin Kwon), a part of the non-homologous end joining (NHEJ) machinery. Deletion of this gene drastically increases the efficiency of targeted homologous

recombination and therefore, the success of the following planned deletions. Subsequently, the resulting strain MJK20.3 was used as the background strain for the following deletions and is from now on referred to as the reference strain. Via mutation of *pyr5*, a further strain was generated within this work that enables the usage of *pyr5* as an alternative marker in future transformations (data not shown).

Finally, during this thesis the functionality of the tunable expression system Tet-On/Off system from *A. niger* according to Wanka et al. 2016 was analyzed. In the Tet-On system gene transcription is turned on by the addition of doxycycline, which forms a complex with the via P *gpdA* constitutive expressed transcription factor rtTA2S-M2, thereby inducing association of rtTA2S-M2 protein to its operator binding site *tetO7*. Located behind the *tetO7* binding site, is the minimal promoter of *fraA* which itself controls the expression of the reporter gene firefly luciferase. The Tet-Off system works in the opposite manner, where addition of Dox reversibly turns off gene transcription, due to the antibiotic induced dissociation of tTA2S from *tetO7* (Wanka et al. 2016). To test these systems in *T. thermophilus*, the whole system together with the backbone was amplified using the plasmids pFW21.8 and pFW22.1 (Wanka et al. 2016) and connected to a truncated and mutated *T. thermophilus pyr5* gene instead of the mutated *A. niger pyrG* gene. The resulting plasmid was integrated into the *T. thermophilus pyr5* mutant according to the mechanisms described in (Arentshorst et al. 2015). In total, approximately 25/200 transformants for each system were analyzed via a luciferase assay. For none of the transformants the expected luciferase signals were obtained and the results of a diagnostic PCR on 10 clones (for each system) suggested that the transformants are mixed clones where an integration band as well as the wildtype band were visible. Taken together, the results suggest that the integration of the Tet-systems did not work although the transformation itself worked, since the transformants did have an intact *pyr5* gene. This project could unfortunately not be continued due to time limitations. All further molecular biological methods that were used in this work were performed according to the common procedures for work with *A. niger* in this laboratory.

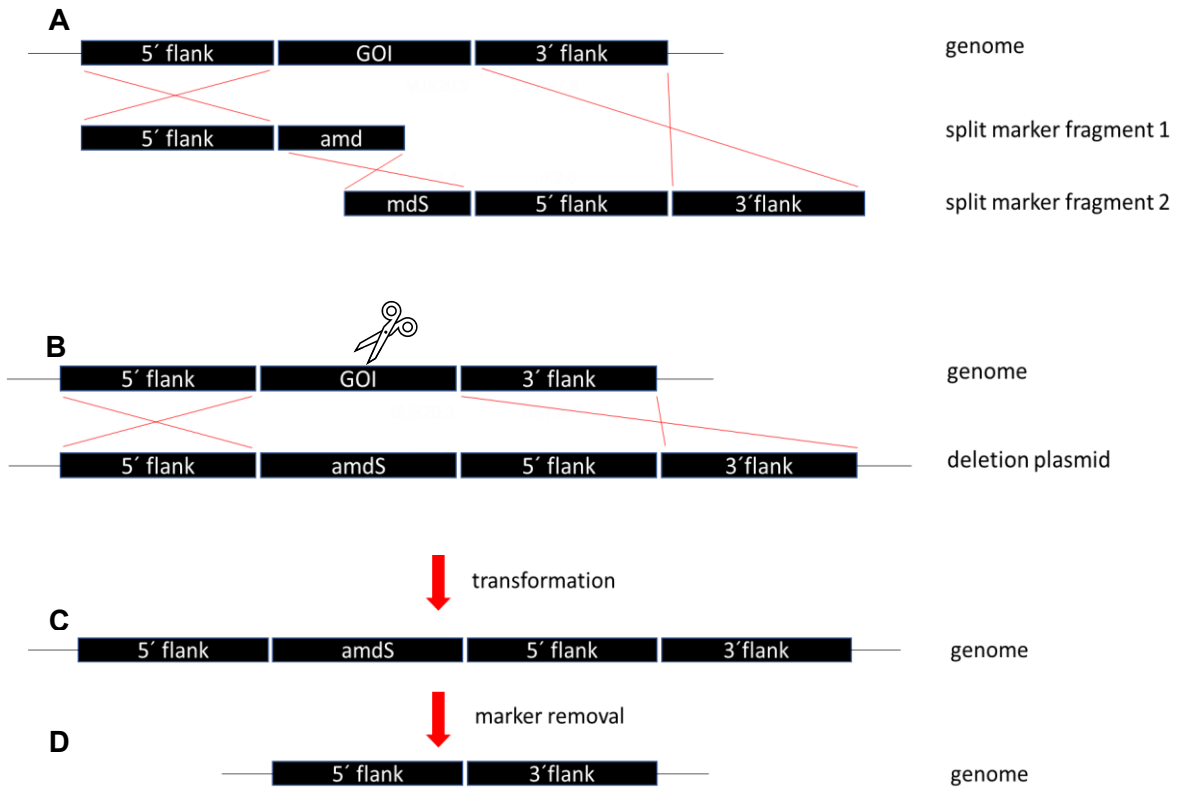

Figure 5.1: Schematic illustration of gene deletion strategies used in this work. Homologous recombination sites are shown as red lines. (A) Split marker approach. Two split marker fragments were transformed in *T. thermophilus* to enhance the possibility of correct homologous recombination, resulting in (C). (B) CRISPR/Cas12a approach. The whole plasmid carrying the deletion cassette was transformed in *T thermophilus*, where CRISPR/Cas12a (depicted by scissors) was used to cut in the gene of interest (GOI) to enhance the possibility of correct homologous recombination. This approach also resulted in (C). After the gene was successfully deleted (C), the marker (*amdS*) was removed via FAA plating, resulting in (D).

5.1.2 Generation of *T. thermophilus* regulator deletion strains

In order to study the function of different regulators in *T. thermophilus* the following genes were deleted in the reference strain MJK20.3. These genes included MYCTH_2298863 (*clr1*) and MYCTH_38704 (*clr2*), which are known to be the main regulators of cellulase expression in many filamentous fungi, especially in *T. thermophilus* (Haefner et al. 2017a; Haefner et al. 2017b; Benocci et al. 2017), as well as MYCTH_2296492 (*clr4*) which is known to be involved in cellulase expression in *T. thermophilus* and *N. crassa* (Liu et al. 2019a). Furthermore, MYCTH_46530 (*vib1*) whose ortholog is involved in carbon catabolite repression in *N. crassa* and therefore, also in regulation of cellulase expression (Xiong et al. 2014) was chosen for deletion as well as the so far unknown and uncharacterized putative regulator MYCTH_108157 which was selected based on its homology to *vib1* and therefore, named *vib2*. The general strategies used for the generation of deletion mutants was described in the previous chapter. The genes *clr1*, *clr2*, and *vib2* were deleted the via split marker approach (positive clones/ analyzed clones: *clr1*: 0/5; *clr2*: 6/11; *vib2*: 4/12; data not shown). The deletion of *clr1* was only successful after additional usage of the CRISPR/Cas12a system (positive clones/ analyzed clones: *clr1*: 5/5; data not shown). The fragments used for transformation were generated via PCR using plasmids (see Tables 4.3 and 4.4 for templates and primers used) received from BASF, as a template. The genes *clr4* and

vib1 were deleted using a CRISPR/Cas12a approach and transformation of the whole deletion plasmid (positive clones/ analyzed clones: *clr4*: 9/9; *vib1*: 5/6, data not shown). The deletion plasmids were constructed via CPEC (*clr4*) and Gibson cloning (*vib1*). Information about the fragments and templates used for CPEC and Gibson cloning can be found in Table 8.1 and Table 8.2 (appendix), respectively. As an example, the plasmid used for the deletion of *clr4* is shown in Figure 8.2 (appendix). The structure of the *vib1* deletion plasmid was identical except the 5´ and 3´ flanks used for homologous recombination. Successful deletion and marker recycling were analyzed via PCR and Southern blot. A schematic overview of the PCR reactions used as well as the results of the Southern analysis of all strains after the marker recycling can be found in the appendix. The resulting strains are listed in Table 4.2. For each deletion mutant one strain was chosen for the further experiments (Table 5.1), additional strains were stored at -80 °C.

Table 5.1: Generated strains that were used in the following experiments.

strain	**genotype**
MJK20.3	Δ*ku80*
BS4.1	Δ*ku80*, Δ*vib2*
BS6.4	Δ*ku80*, Δ*clr2*
BS7.8	Δ*ku80*, Δ*clr1*
JK2.8	Δ*ku80*, Δ*clr4*
BS9.3	Δ*ku80*, Δ*vib1*

5.1.3 Plate growth assay of *T. thermophilus* strains using different carbon sources

To test the consequences of the deletion of the different regulators regarding the ability to use various saccharides as a carbon source, a plate growth assay with 16 different carbon sources was conducted as described in chapter 4.2.2.5. All generated deletion strains and the reference strain MJK20.3 were used for inoculation of the agar plates. The chosen saccharides represent components and degradation products of plant biomass that are found in cellulose, hemicellulose, and pectin as well as components and degradation products of starch. The results after 4 days of growth at 37 °C are shown in Figure 5.2. The experiment was carried out in duplicate but since the results were similar only results for one replicate are shown.

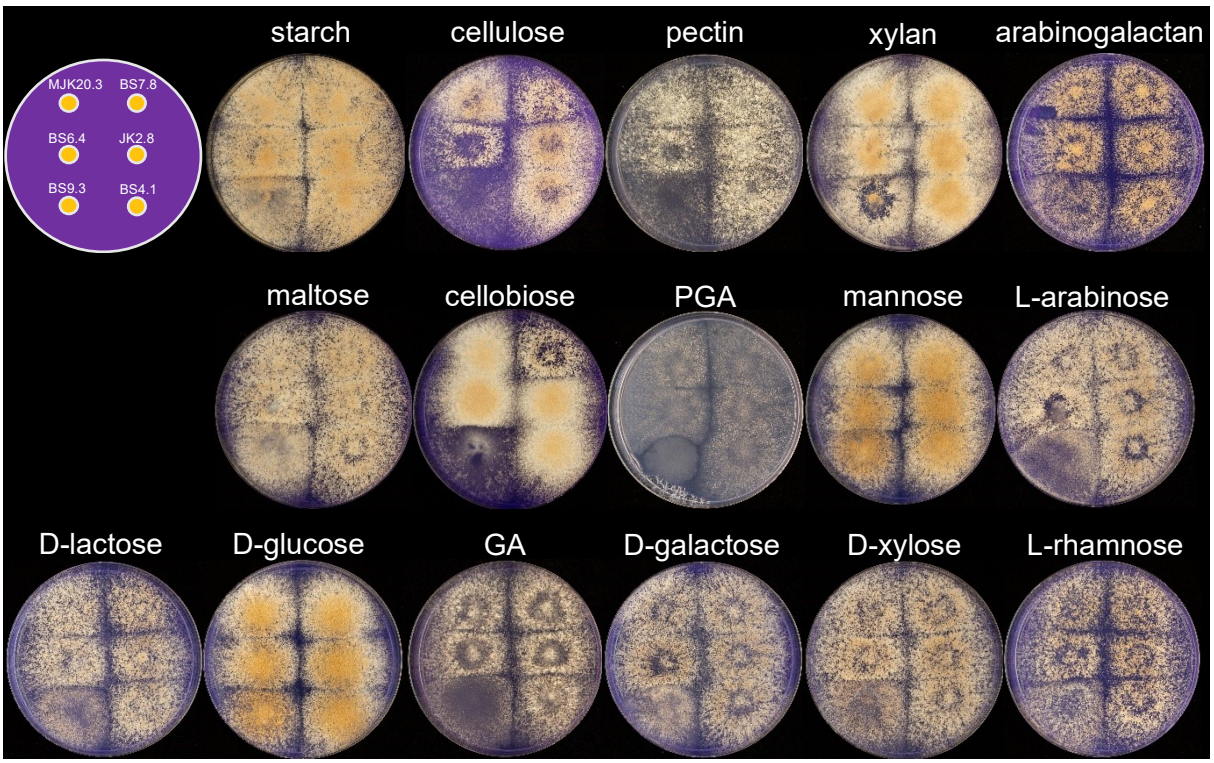

Figure 5.2: Plate growth assay of *T. thermophilus* strains on different carbon sources. The plates were inoculated for 4 days at 37 °C with 1000 spores each according to the scheme shown in the upper left corner (MJK20.3: Δ*ku80*, BS7.8: Δ*ku80*, Δ*clr1*, BS6.4: Δ*ku80*, Δ*clr2*, JK2.8: Δ*ku80*, Δ*clr4*, BS9.3: Δ*ku80*, Δ*vib1*, BS4.1: Δ*ku80*, Δ*vib2*). PGA= polygalacturonic acid, GA= galacturonic acid.

Growth is similar for all strains on all carbon sources, except for BS9.3 and BS6.4. It is clearly visible, that BS9.3 shows a growth defect on every carbon source used. This growth defect is much stronger when growing on plant cell wall components. For starch, maltose, and glucose only minor differences compared to all other strains can be observed. BS6.4 has a growth defect when growing on cellulose and pectin, while BS7.8 has a growth defect when growing on cellulose and cellobiose. Besides the above-mentioned differences, preferred carbon sources are glucose, cellobiose, xylan, mannose, starch, and cellulose.

5.2 Establishment and performance of bioreactor cultivations of different *T. thermophilus* strains

At first, a protocol for a stable cultivation of *T. thermophilus* in the bioreactor had to be established. The hurdles are briefly described in the following. Data for these processes will not be shown. In the beginning, a batch cultivation using bioreactor medium (see Table 4.28, with 2.7 mM $CaCl_2x2H_2O$) as well as cultivation conditions provided by BASF (described in chapter 4.2.4, with 37 °C instead of 45 °C) was performed. Here, the *T. thermophilus* wildtype strain started sporulation immediately after germination of the spores, which was regarded as a signal that cultivation conditions or the composition of the medium are not suitable. Therefore, several further bioreactor cultivations and shaking flask experiments were performed together with varying medium composition or cultivation conditions. Finally, it was discovered, that a cultivation temperature of 45 °C in combination with complete absence of $CaCl_2x2H_2O$ in the medium was required to cultivate the *T. thermophilus* wildtype strain in a batch experiment without sporulation. The sporulation of a culture was always accompanied by pink medium color, enabling a fast readout of the tested conditions in shaking flask experiments. Figure 5.3 illustrates the color and morphology of a sporulating and a non-sporulating *T. thermophilus* culture. After establishing the batch cultivation, growth behavior during chemostat cultivation was investigated. After an initial sporulation as soon as starting the chemostat program, the fungus was growing filamentously and the sporulation stopped. A steady state was achieved after approximately 12-17 residence times and cultivation in chemostat was regarded as successful. With the same conditions that were successfully used for cultivation of the wildtype strain in chemostat, a steady state could not be reached using MJK20.3. Therefore, several further shaking flask experiments were performed. The discovery was made that the addition of $CaCl_2x2H_2O$ to the medium is a critical factor. If the medium contains no $CaCl_2x2H_2O$, MJK20.3 is not able to maintain growth in chemostat and if the $CaCl_2x2H_2O$ concentration is too high a premature sporulation of the fungus can be observed. Therefore, bioreactor medium with 10-fold lower $CaCl_2x2H_2O$ concentration (=bioreactor medium listed in Table 4.28) compared to the original recipe (0.27 mM vs. 2.7 mM) that was received from BASF, was used in a following bioreactor experiment. In this experiment, batch cultivation and chemostat cultivation proceeded identical to the wildtype cultivation without $CaCl_2x2H_2O$ and finally the establishment of stable bioreactor cultivations was completed.

After the successful establishment of stable bioreactor cultivations, the strains were cultivated using various conditions. An overview of the strains and the used cultivation conditions can be found in Table 5.2. A detailed description of the cultivation parameters, sampling, and the media used can be found in Table 4.28 and chapter 4.2.4. The following figures (Figures 5.4-5.12) are showing the results of the bioreactor cultivations.

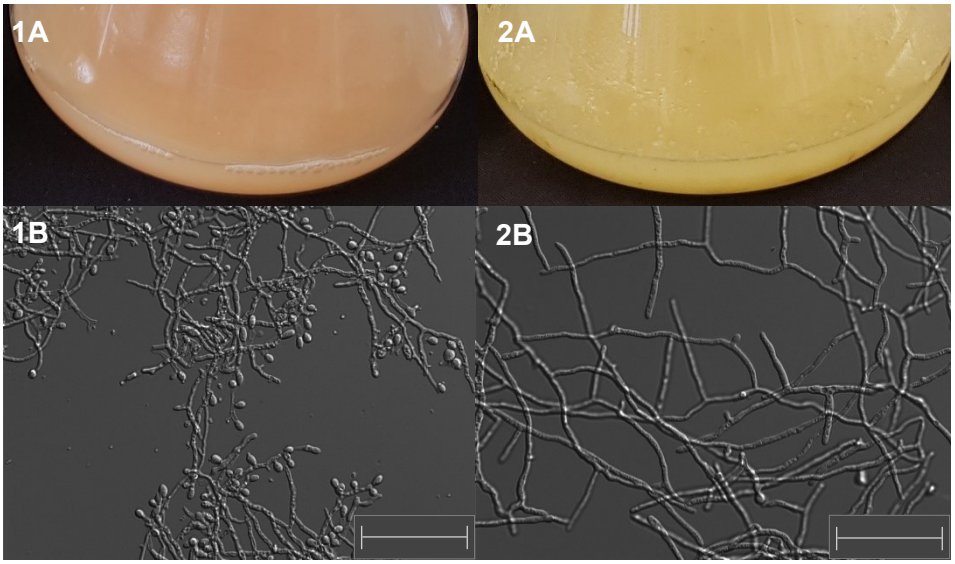

Fig 5.3: Color and morphology of a sporulating and a non-sporulating *T. thermophilus* culture. Pictures are showing the color of the culture broth (A) as well as the microscopic morphology (B) of a sporulating (1) and a non-sporulating (2) *T. thermophilus* culture. Scale bar= 50 µm.

Table 5.2: Overview of the strains used, their genotypes, the cultivation conditions of the performed bioreactor experiments, and the number of the figure in which the results of the respective bioreactor cultivation can be found.

strain	genotype	cultivation conditions	Figure
MJK20.3	Δ*ku80*	chemostat cultivation with glucose as carbon source, spiking of cellulose in steady state	5.4
		chemostat cultivation with xylose as carbon source	5.5
		batch cultivation with pH shift experiment	5.6, 5.7
BS4.1	Δ*ku80*, Δ*vib2*	chemostat cultivation with glucose as carbon source, spiking of cellulose in steady state	5.8
BS6.4	Δ*ku80*, Δ*clr2*		5.9
BS7.8	Δ*ku80*, Δ*clr1*		5.10
JK2.8	Δ*ku80*, Δ*clr4*		5.11
BS9.3	Δ*ku80*, Δ*vib1*		5.12

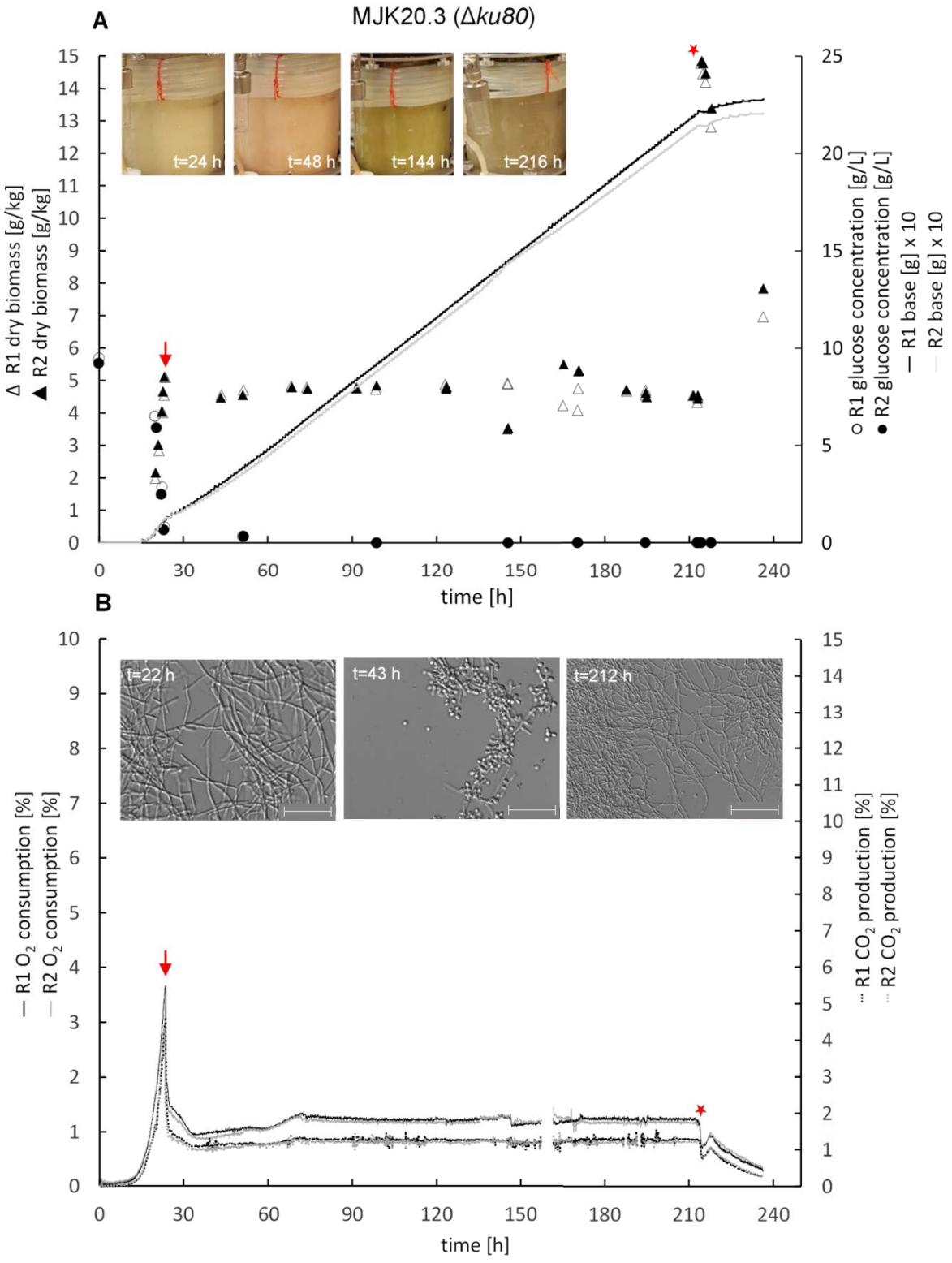

Figure 5.4: Physiology of *T. thermophilus* during controlled chemostat bioreactor cultivations of strain MJK20.3 (Δ*ku80*). Dry biomass, glucose concentration, base addition (A), and off gas data (B) during cultivation of MJK20.3. Pictures indicate color development during cultivation and morphology at different points in time. Each cultivation was run in duplicate (R1, R2). Chemostat cultivation (dilution rate 0.1 1/h) in bioreactor medium with 55.5 mM glucose was started at the end of the batch phase (red arrow). During steady state, 1 % cellulose was spiked (red star) and chemostat cultivation was switched to batch cultivation. Due to technical problems some datapoints in (B) are missing. Scale bar= 50 µm.

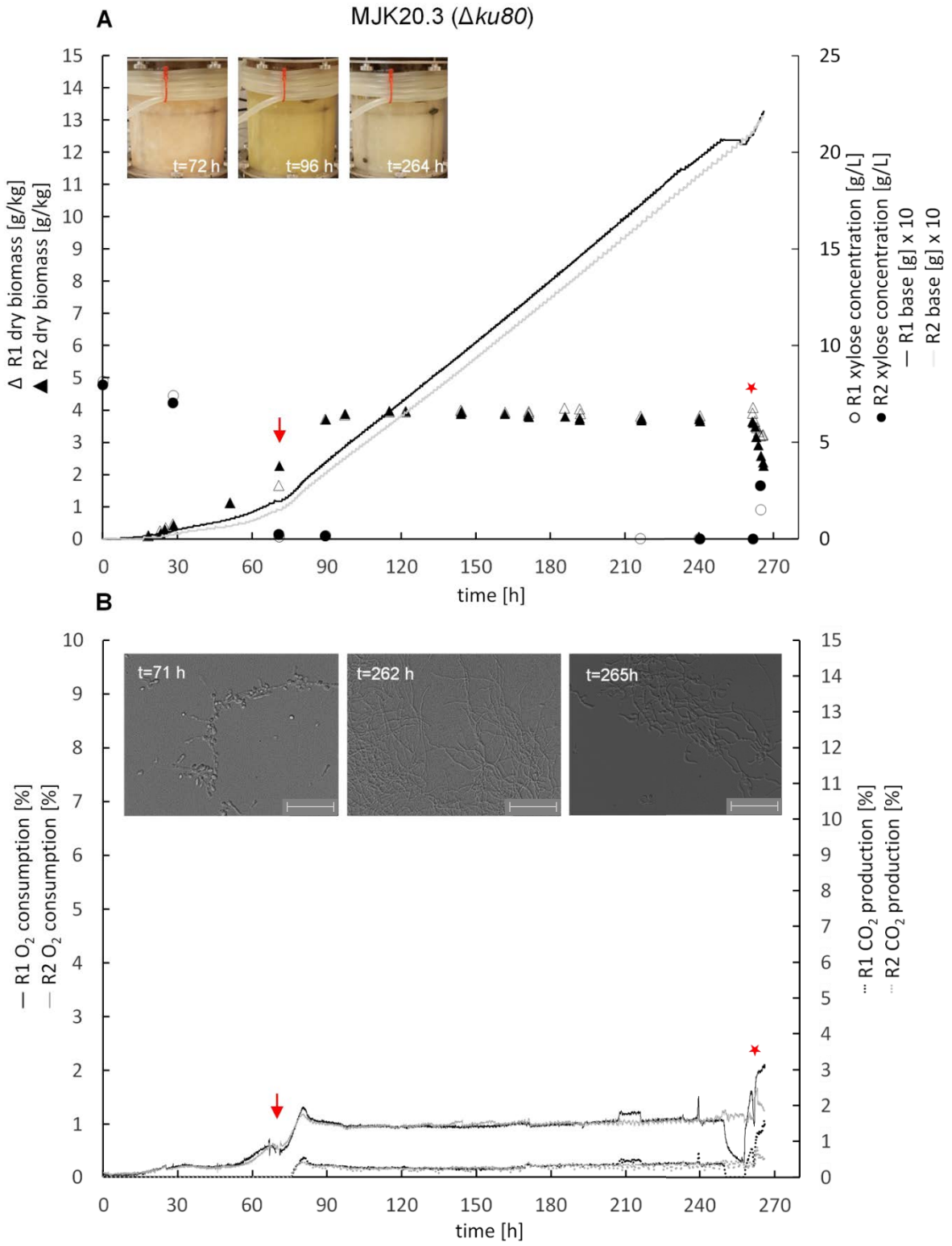

Figure 5.5: Physiology of *T. thermophilus* during controlled chemostat bioreactor cultivations of strain MJK20.3 (Δ*ku80*). Dry biomass, xylose concentration, base addition (A), and off gas data (B) during cultivation of MJK20.3. Pictures indicate color development during cultivation and morphology at different points in time. Each cultivation was run in duplicate (R1, R2). Chemostat cultivation (dilution rate 0.1 1/h) in bioreactor medium with 55.5 mM xylose was started at the end of the batch phase (red arrow). During steady state, a wash out with the maximum dilution rate (0.366 1/h) was performed (red star). Scale bar= 50 µm.

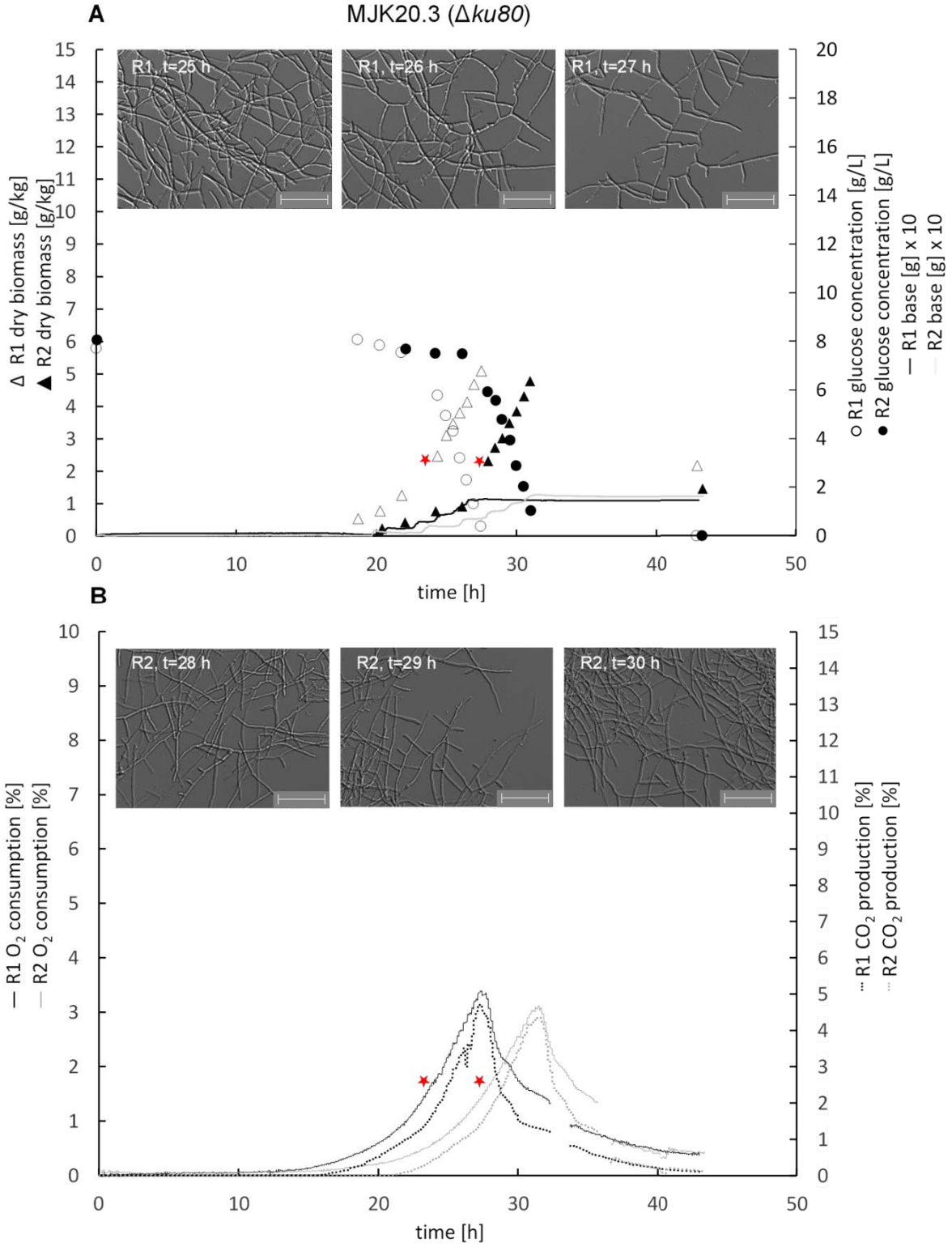

Figure 5.6: Physiology of *T. thermophilus* during pH shift experiment of strain MJK20.3 (Δ*ku80*). Dry biomass, glucose concentration, base addition (A), and off gas data (B) during cultivation of MJK20.3 without pH shift. Each cultivation was run in duplicate (R1, R2). Pictures indicate morphology at different points in time (R1 in (A), R2 in (B)). Since the color of the broth was identical to the color in exponential phase of MJK20.3 cultivated in chemostat, no pictures are shown here. Samples for transcriptome analysis were taken starting from a dry biomass value of 2.2 to 2.5 g/kg as t0= 0 h (red stars). The following samples were taken t1= 0.5 h, t2= 1 h, and t3= 2 h after sampling started. Due to technical problems some datapoints in (B) are missing. Scale bar= 50 µm.

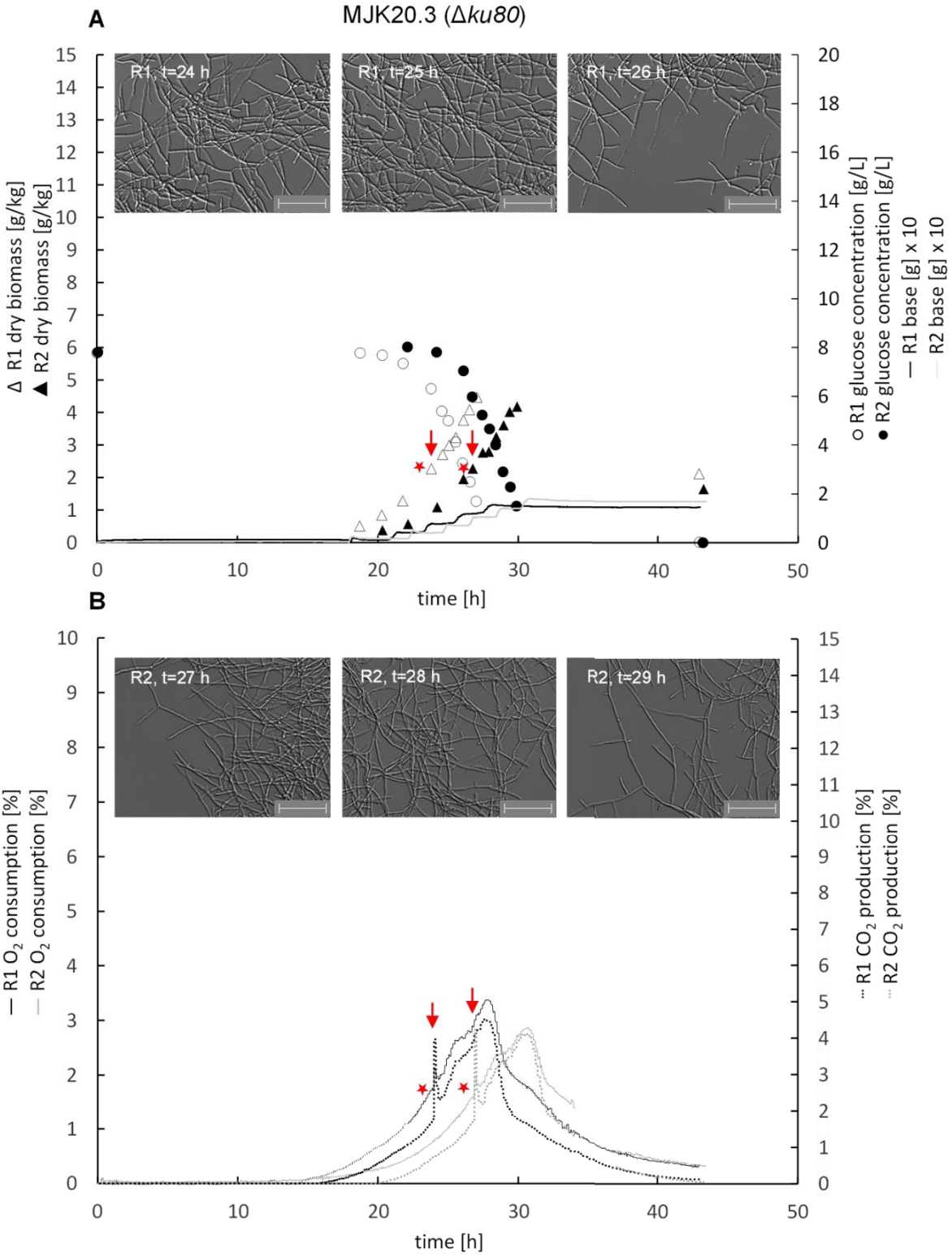

Figure 5.7: Physiology of *T. thermophilus* during pH shift experiment of strain MJK20.3 (Δ*ku80*). Dry biomass, glucose concentration, base addition (A), and off gas data (B) during cultivation of MJK20.3 with pH shift to 5.0 (red arrow). Each cultivation was run in duplicate (R1, R2). Pictures indicate morphology at different points in time (R1 in (A), R2 in (B)). Since the color of the broth was identical to the color in exponential phase of MJK20.3 cultivated in chemostat, no pictures are shown here. Samples for transcriptome analysis were taken starting from a dry biomass value of 2.2 to 2.5 g/kg as t0= 0 h prior to the pH shift (red stars). The following samples were taken t1= 0.5 h, t2= 1 h and t3= 2 h after the pH shift. Due to technical problems some datapoints in (B) are missing. Scale bar= 50 µm.

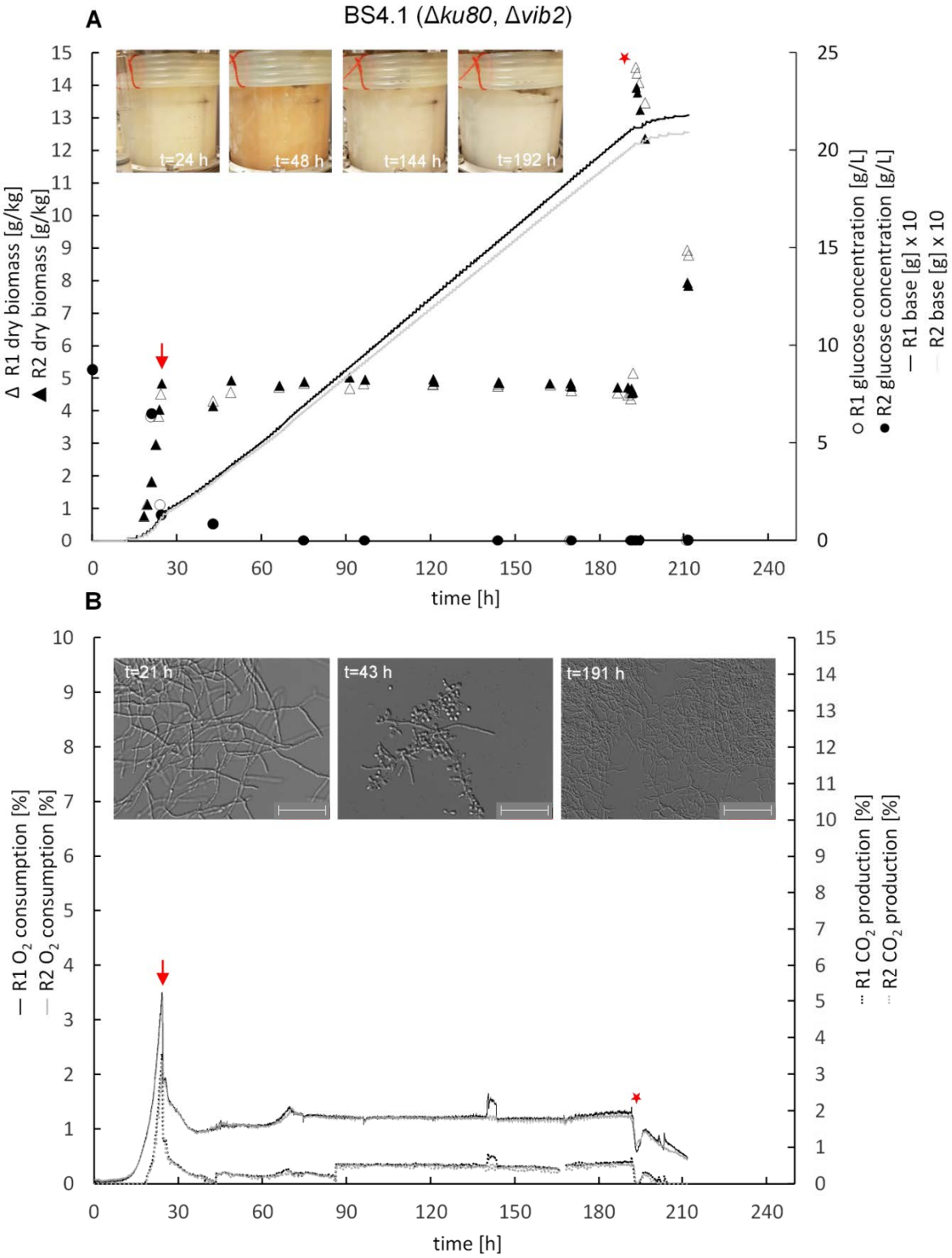

Figure 5.8: Physiology of *T. thermophilus* during controlled chemostat bioreactor cultivations of strain BS4.1 (Δ*ku80*, Δ*vib2*). Dry biomass, glucose concentration, base addition (A), and off gas data (B) during cultivation of BS4.1. Pictures indicate color development during cultivation and morphology at different points in time. Each cultivation was run in duplicate (R1, R2). Chemostat cultivation (dilution rate 0.1 1/h) in bioreactor medium with 55.5 mM glucose was started at the end of the batch phase (red arrow). During steady state, 1 % cellulose was spiked (red star) and chemostat cultivation was switched to batch cultivation. Scale bar= 50 µm.

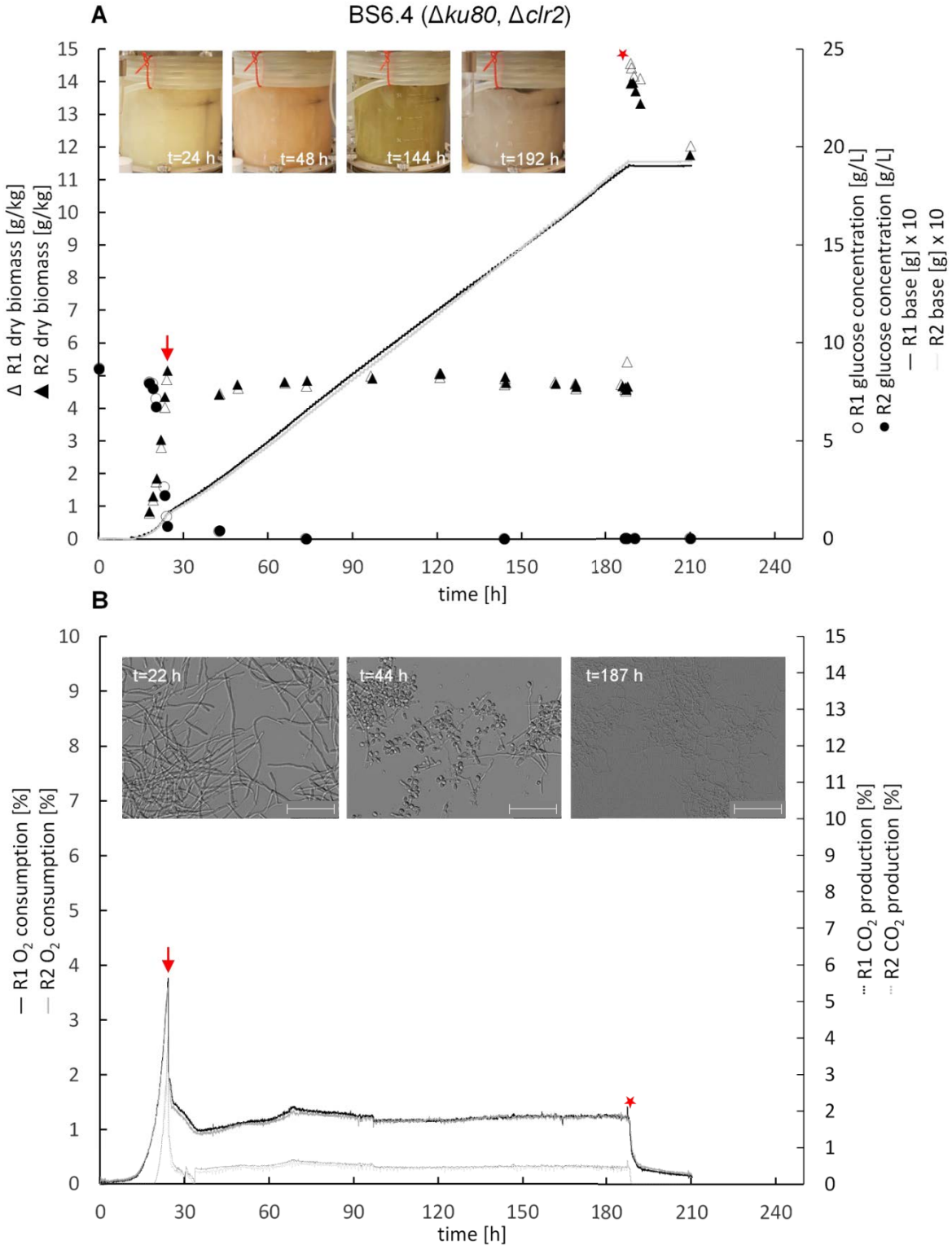

Figure 5.9: Physiology of *T. thermophilus* during controlled chemostat bioreactor cultivations of strain BS6.4 (Δ*ku80*, Δ*clr2*). Dry biomass, glucose concentration, base addition (A), and off gas data (B) during cultivation of BS6.4. Pictures indicate color development during cultivation and morphology at different points in time. Each cultivation was run in duplicate (R1, R2). Chemostat cultivation (dilution rate 0.1 1/h) in bioreactor medium with 55.5 mM glucose was started at the end of the batch phase (red arrow). During steady state, 1 % cellulose was spiked (red star) and chemostat cultivation was switched to batch cultivation. Scale bar= 50 µm.

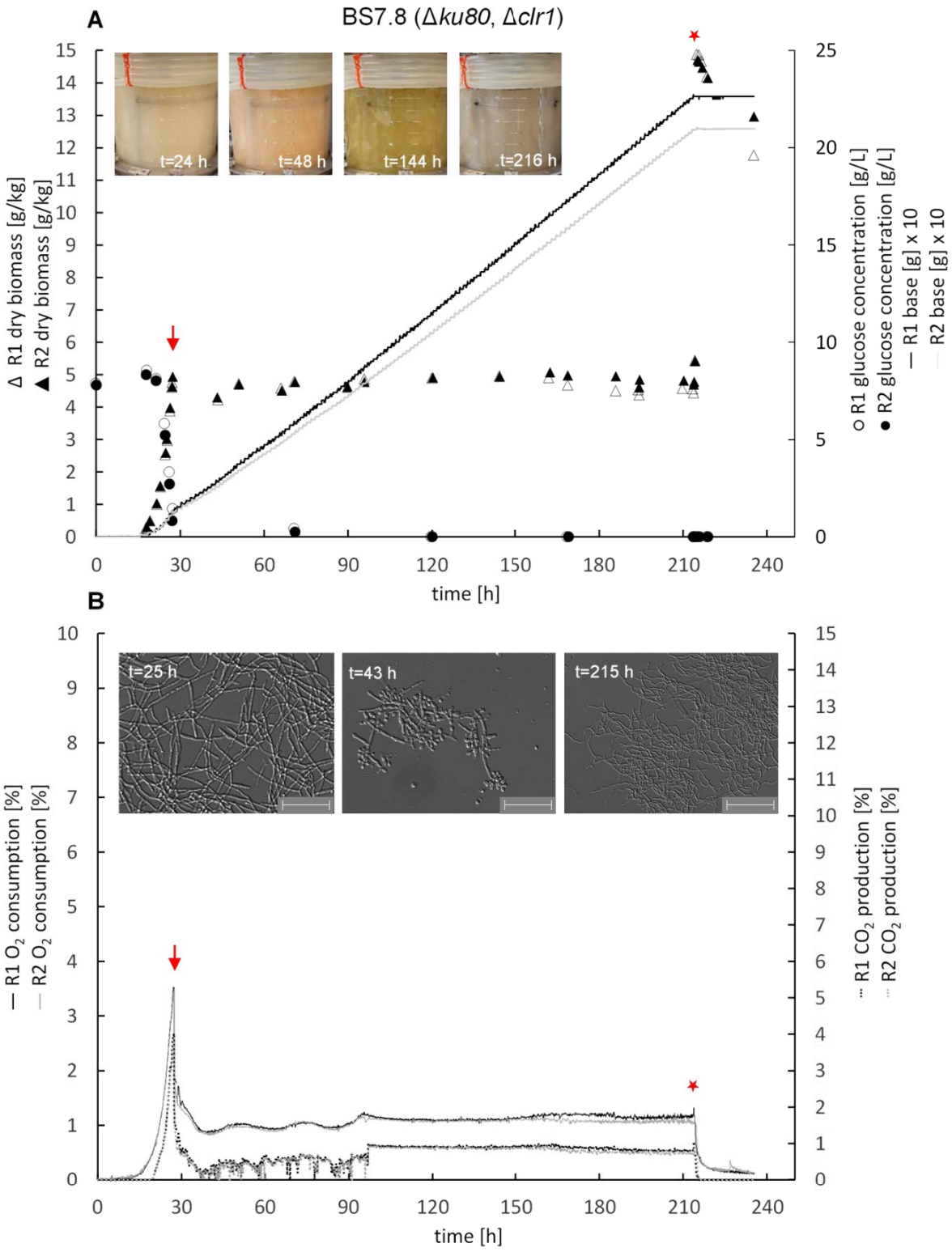

Figure 5.10: Physiology of *T. thermophilus* during controlled chemostat bioreactor cultivations of strain BS7.8 (Δ*ku80*, Δ*clr1*). Dry biomass, glucose concentration, base addition (A), and off gas data (B) during cultivation of BS7.8. Pictures indicate color development during cultivation and morphology at different points in time. Each cultivation was run in duplicate (R1, R2). Chemostat cultivation (dilution rate 0.1 1/h) in bioreactor medium with 55.5 mM glucose was started at the end of the batch phase (red arrow). During steady state, 1 % cellulose was spiked (red star) and chemostat cultivation was switched to batch cultivation. Scale bar= 50 µm.

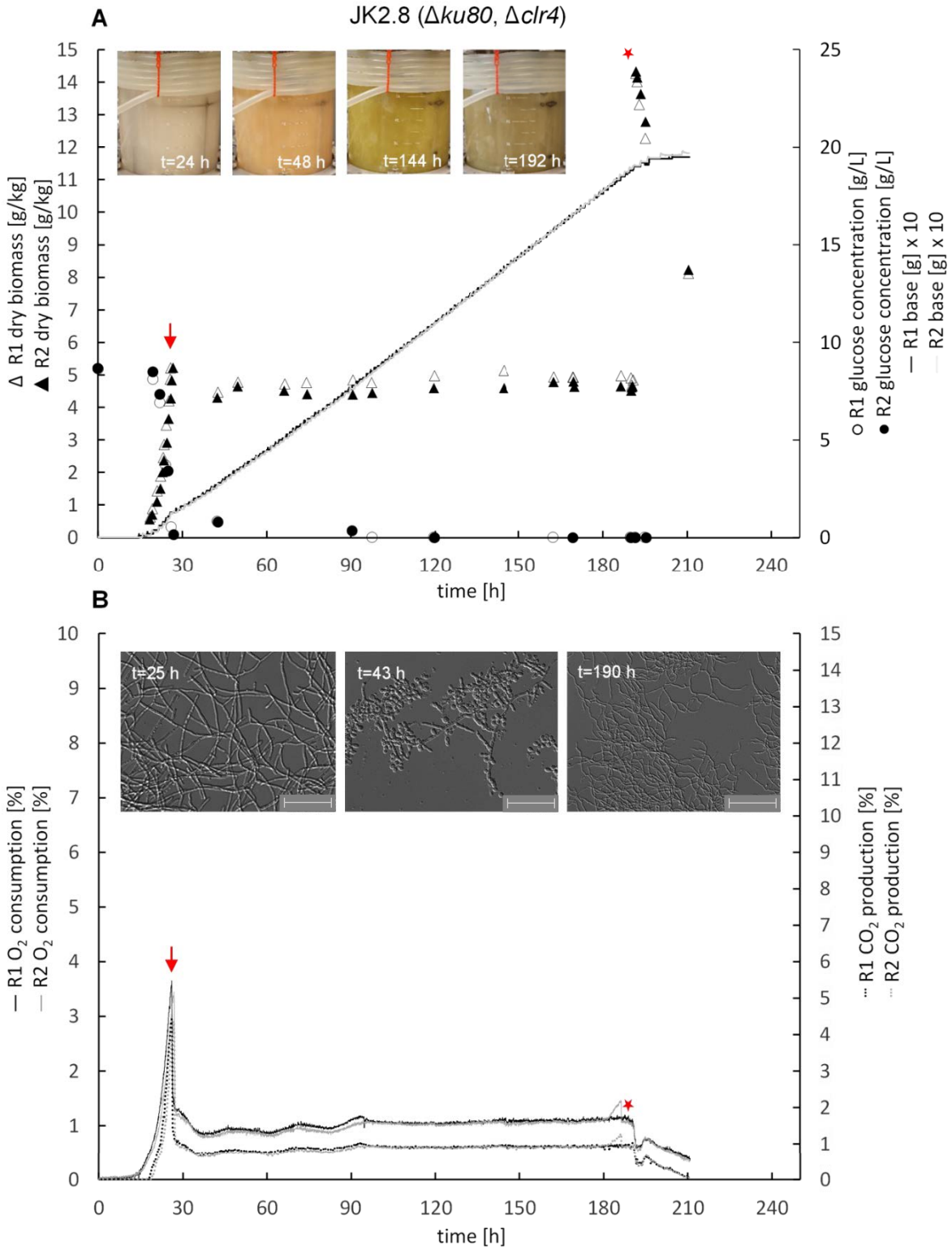

Figure 5.11: Physiology of *T. thermophilus* during controlled chemostat bioreactor cultivations of strain JK2.8 (Δ*ku80*, Δ*clr4*). Dry biomass, glucose concentration, base addition (A), and off gas data (B) during cultivation of JK2.8. Pictures indicate color development during cultivation and morphology at different points in time. Each cultivation was run in duplicate (R1, R2). Chemostat cultivation (dilution rate 0.1 1/h) in bioreactor medium with 55.5 mM glucose was started at the end of the batch phase (red arrow). During steady state, 1 % cellulose was spiked (red star) and chemostat cultivation was switched to batch cultivation. Scale bar= 50 µm.

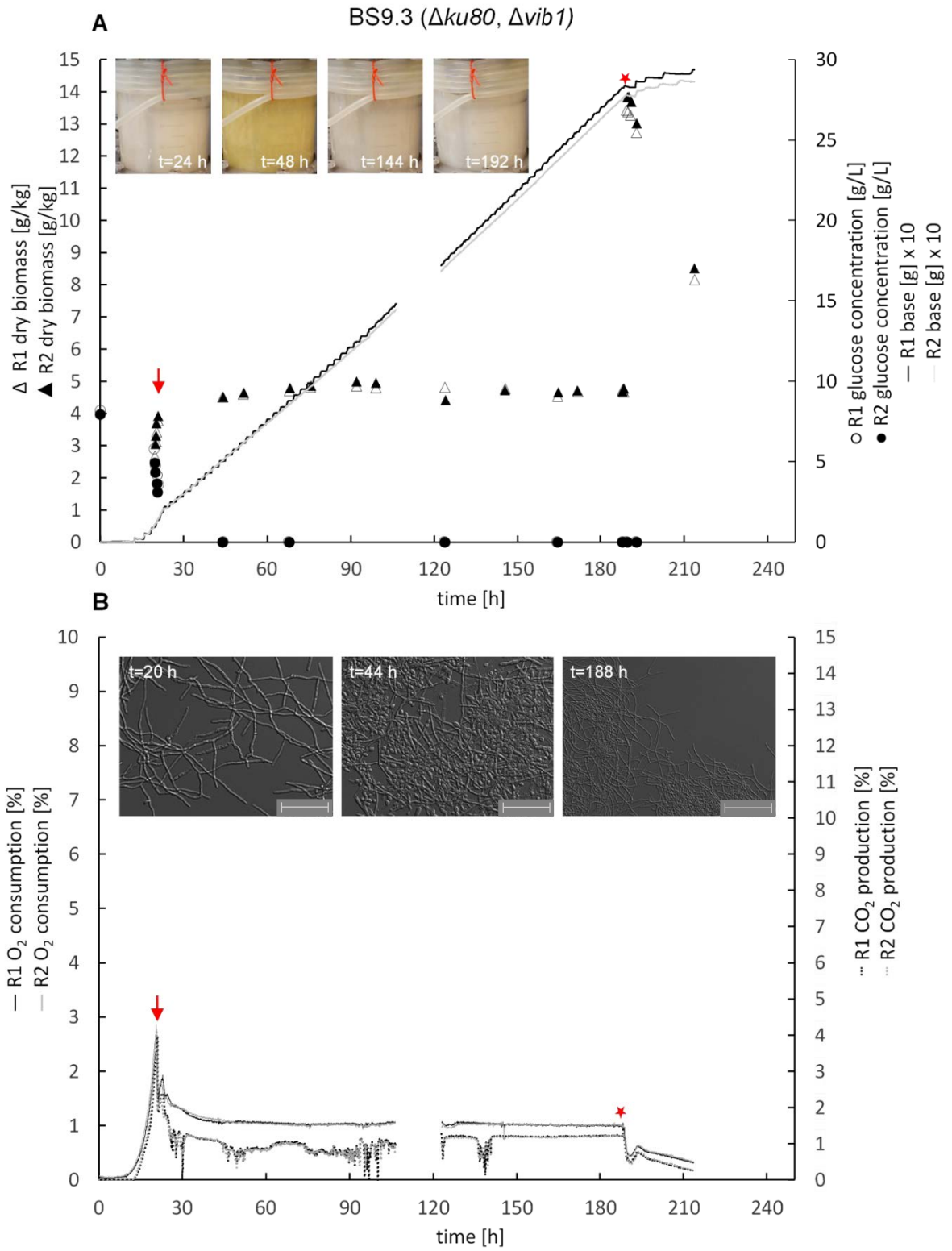

Figure 5.12: Physiology of *T. thermophilus* during controlled chemostat bioreactor cultivations of strain BS9.3 (Δ*ku80*, Δ*vib1*). Dry biomass, glucose concentration, base addition (A), and off gas data (B) during cultivation of BS9.3. Pictures indicate color development during cultivation and morphology at different points in time. Each cultivation was run in duplicate (R1, R2). Chemostat cultivation (dilution rate 0.1 1/h) in bioreactor medium with 55.5 mM glucose was started at the end of the batch phase (red arrow). During steady state, 1 % cellulose was spiked (red star) and chemostat cultivation was switched to batch cultivation. Due to technical problems some datapoints in (A) and (B) are missing. Scale bar= 50 µm.

Growth rates of the strains while growing under different conditions were determined using the dry biomass values of the exponential state. The results are shown in Figure 5.13.
For further evaluation of the morphology, hyphal diameters using the microscopic pictures from exponential state and steady state were determined via ImageJ, as shown in Figure 5.14.
To investigate the differences in protein secretion, Bradford assays and SDS-PAGEs were performed. The results are shown in Figures 5.15 and 5.16, respectively. SDS PAGEs of the exponential state samples were not showing any protein and are therefore, not shown. Since the band pattern for the steady state samples was almost identical to the first points in time after spiking cellulose (t1, t2) and sample t3 almost identical to sample t4, only the results of the SDS PAGEs for the steady state samples and t4 are shown. All other results of the SDS-PAGEs can be found in the appendix.

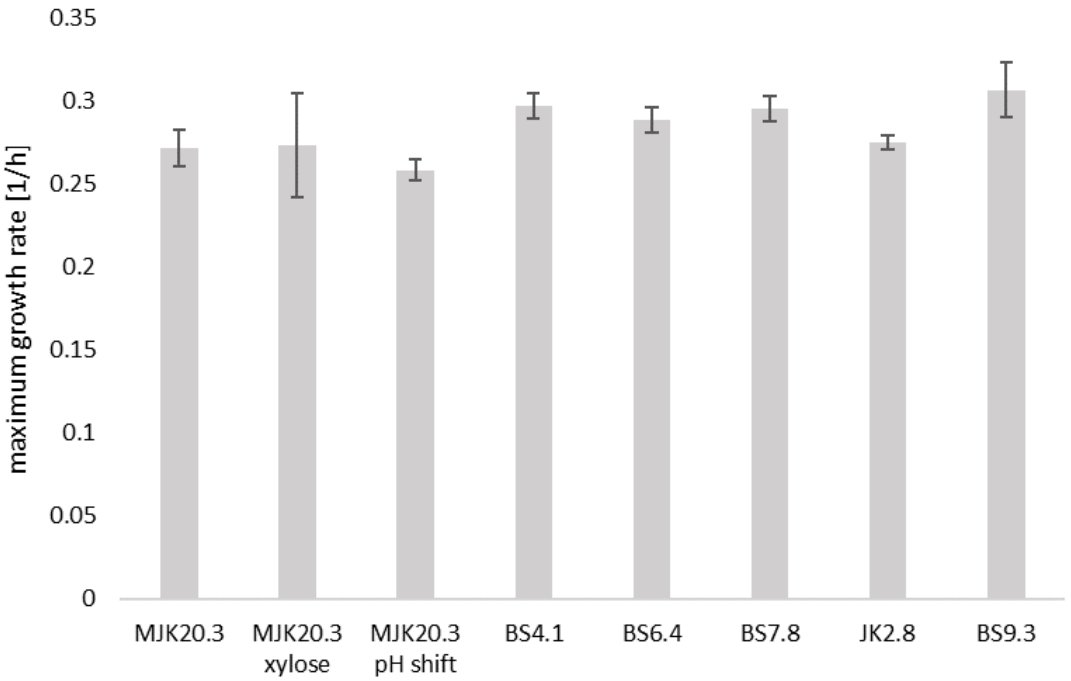

Figure 5.13: Growth rate comparison between different strains and conditions. The growth rates [1/h] were determined using dry biomass values from the respective exponential states. Each bar represents data from two replicates, except for MJK20.3 (four replicates). For each replicate growth rate was determined using three different amounts of datapoints. The error bar represents the standard deviation of the mean value, deriving from the six different values for each strain/condition. If not mentioned separately, glucose was used as carbon source in the cultivations. Strains: MJK20.3: Δ*ku80*, BS7.8: Δ*ku80*, Δ*clr1*, BS6.4: Δ*ku80*, Δ*clr2*, JK2.8: Δ*ku80*, Δ*clr4*, BS9.3: Δ*ku80*, Δ*vib1*, BS4.1: Δ*ku80*, Δ*vib2*.

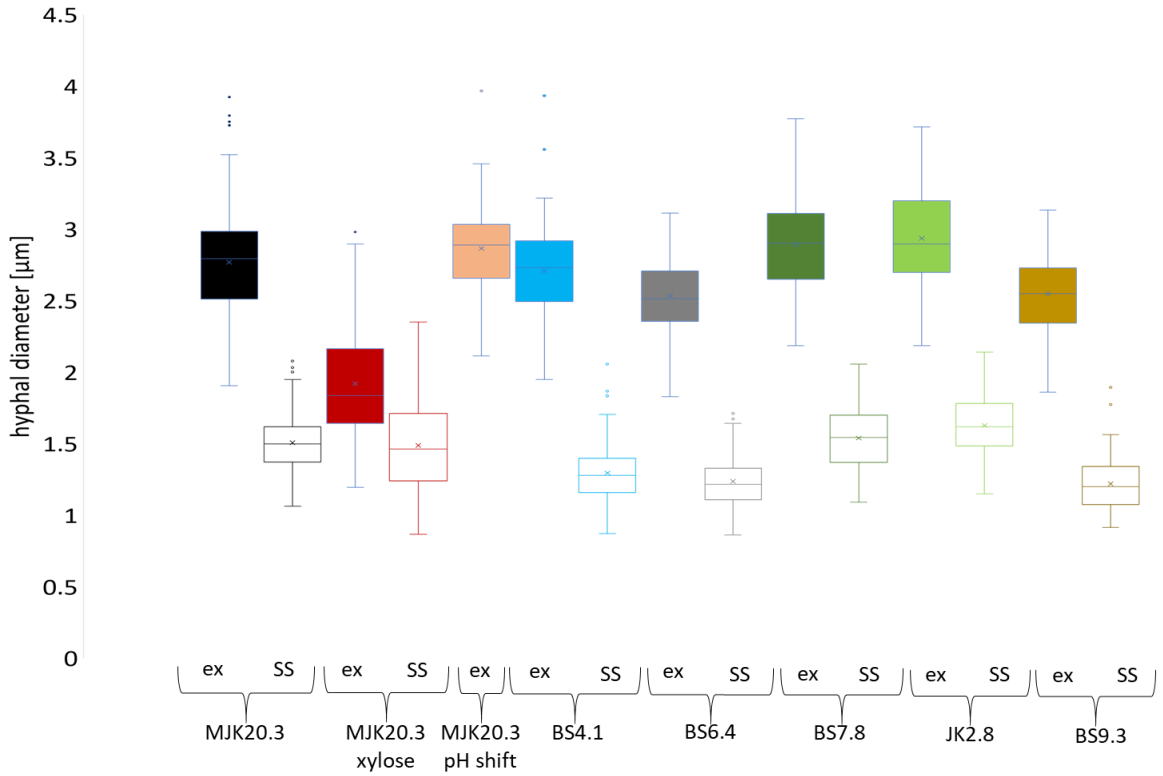

Figure 5.14: Comparison of the hyphal diameters between different strains and conditions. The hyphal diameters [µm] were determined via ImageJ using microscopic pictures from exponential state (ex) and steady state (SS). Each bar represents data from two replicates, except for MJK20.3 ex (four replicates) In total 100 hyphae for each condition/strain were analyzed using at least six different microscopic pictures. This boxplot is showing the median (horizontal line), mean (cross), interquartile range (box), minimum and maximum (whiskers) as well as outliers (dots) for each condition/strain. If not mentioned separately, glucose was used as carbon source in the cultivations. Strains: MJK20.3: Δ*ku80*, BS7.8: Δ*ku80*, Δ*clr1*, BS6.4: Δ*ku80*, Δ*clr2*, JK2.8: Δ*ku80*, Δ*clr4*, BS9.3: Δ*ku80*, Δ*vib1*, BS4.1: Δ*ku80*, Δ*vib2.*

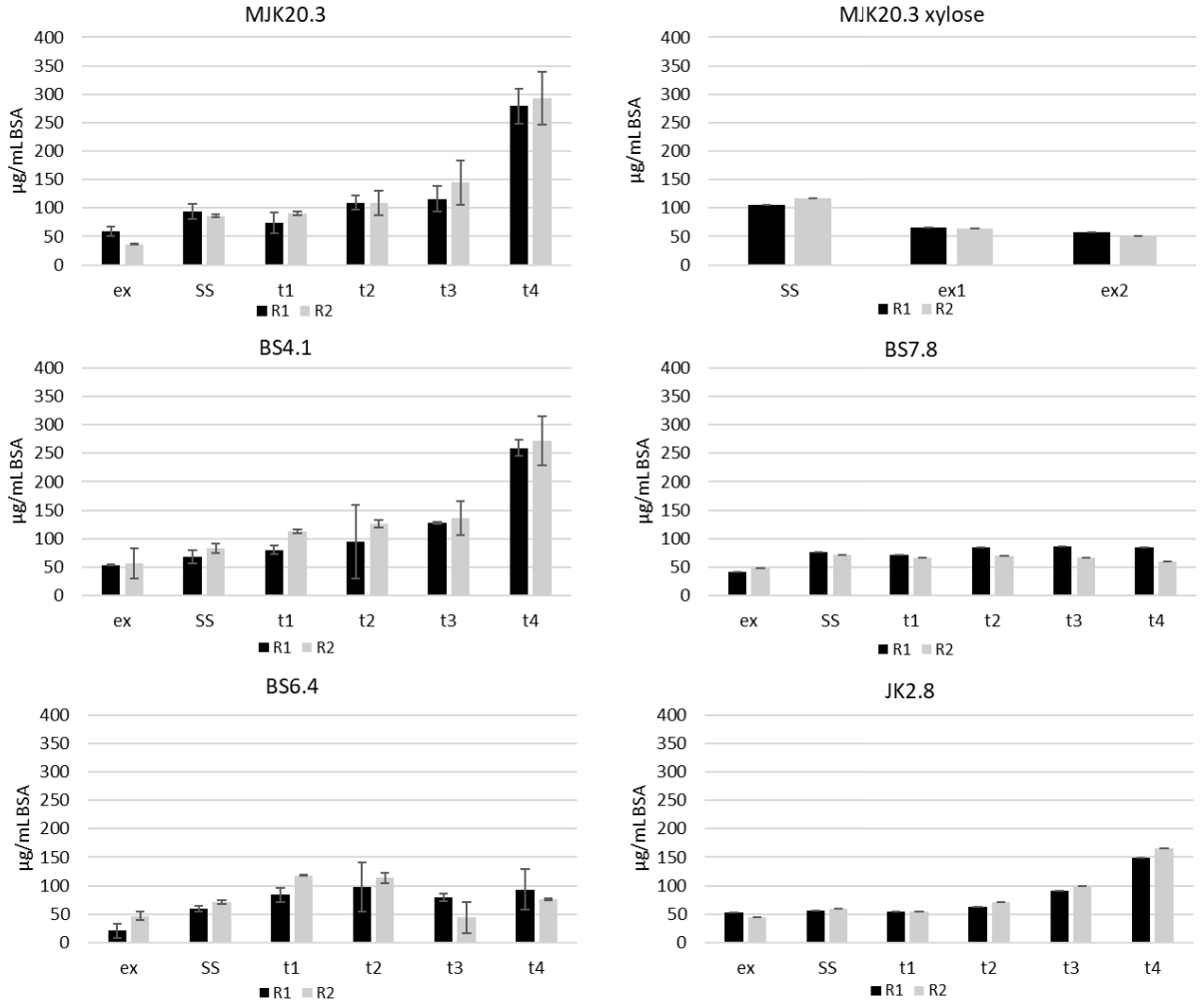

Figure 5.15: Protein analysis of culture supernatants via Bradford assay. Shown is the amount of protein [µg/mL BSA] present in the culture supernatants of the samples taken in exponential state (ex), steady state (SS), as well as 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after spiking cellulose (t4) for the two replicates (R1, R2). For MJK20.3 xylose two samples from exponential state were analyzed (ex1, ex2). The error bar represents the standard deviation of the mean value, deriving from three technical replicates for each sample. If not mentioned separately, glucose was used as carbon source in the cultivations. Strains: MJK20.3: Δ*ku80*, BS7.8: Δ*ku80*, Δ*clr1*, BS6.4: Δ*ku80*, Δ*clr2*, JK2.8: Δ*ku80*, Δ*clr4*, BS9.3: Δ*ku80*, Δ*vib1*, BS4.1: Δ*ku80*, Δ*vib2.*

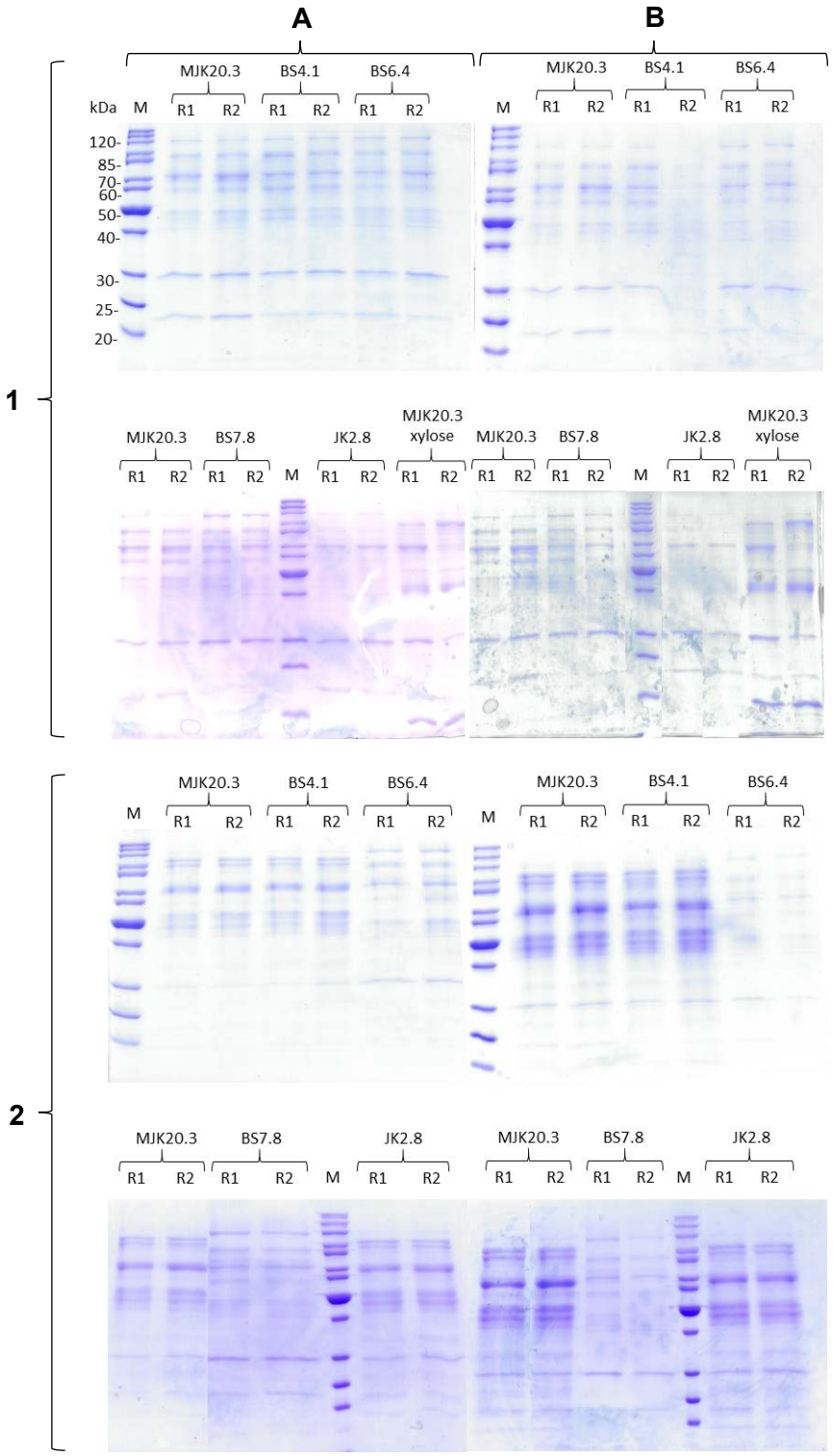

Figure 5.16: Protein analysis of culture supernatants via SDS-PAGE. Shown are the results of the SDS PAGEs with (A) 2 µg protein (determined via Bradford assay) and (B) 20 µL supernatant for the strains/conditions MJK20.3 (Δ*ku80*) on glucose or xylose, BS4.1 (Δ*ku80,* Δ*vib2*), BS6.4 (Δ*ku80,* Δ*clr2*), BS7.8 (Δ*ku80,* Δ*clr1*), and JK2.8 (Δ*ku80,* Δ*clr4*) at steady state (1) and 4 h after spiking cellulose (2) in duplicate (R1, R2). For a better visual presentation, the SDS PAGEs from (2) were cut to size within the whole gel. As a marker for protein size determination [kDa], PageRuler Unstained Protein Ladder (M) was used. If not mentioned separately, glucose was used as carbon source in the cultivations.

5.2.1 Reference strain MJK20.3 (Δ*ku80*); cellulose spiking experiment

The chemostat cultivation of *T. thermophilus* MJK20.3 (Figure 5.4) was successful. Except some foaming and overflow problems at approximately 140 h and 170 h runtime that are reflected in the respective biomass values, no technical problems occurred during the cultivation. After the batch phase (maximum growth rate ~0.27 1/h), continuous cultivation was initiated, and the fungus started sporulation together with fragmentation of the hyphae. This is reflected in the physiology, the respective microscopic pictures, and the color of the culture broth. The pinkish color here also indicated sporulation as mentioned in the previous chapter. Generally, the color of the culture broth during the cultivation shifted from white-yellowish over pinkish and yellow to white-brownish. After around 17 retention times the steady state condition reached regarding morphology (no remaining spores, filamentous growth), biomass, base addition, and offgas values (O_2 consumed, CO_2 produced). Morphological differences between the exponential state and steady state can be seen in the respective microscopic pictures and are quantified in Figure 5.14. Here, hyphal diameters are clearly smaller in steady state compared to the exponential state. The spiking of cellulose to the medium was leading to a continuing increase of base addition and raising offgas values (after an initial spiking related drop), indicating the ability to use cellulose as a carbon source. Protein analysis (Figures 5.15 and 5.16) showed that the fungus started to secrete proteins after the cellulose spike, resulting in an increase of approximately 200 µg/mL protein and an altered protein secretion (see band pattern SDS PAGEs) 4 h after the spike compared to the steady state condition.

5.2.2 Cellulase regulator deletion mutants BS6.4 (Δ*ku80,* Δ*clr2*), BS7.8 (Δ*ku80,* Δ*clr1*), and JK2.8 (Δ*ku80,* Δ*clr4*); cellulose spiking experiment

The chemostat cultivations of *T. thermophilus* BS6.4 (Figure 5.9), BS7.8 (Figure 5.10), and JK2.8 (Figure 5.11) were also successful. For JK2.8 clogging of a tube at approximately 190 h runtime for one of the duplicates caused an increase in the measured offgas values. However, physiology was not further affected, and no other technical problems occurred during the cultivations. After chemostat cultivation was initiated, all strains started sporulation and hyphal fragmentation, which can be seen when looking at the physiology, the respective microscopic pictures, and the color of the culture broth. Generally, the color development of the culture broth during the cultivation was identical to the reference strain (MJK20.3). The maximum growth rates of the strains (Figure 5.13) are almost identical compared to the reference strain, ranging from roughly 0.27 to 0.3 1/h. Note that the growth rate data is only available for two replicates of each strain except for MJK20.3. After around 15 (BS6.4, JK2.8) and 17 (BS7.8) retention times, the steady state condition was reached regarding morphology (no remaining spores, filamentous growth), biomass, base addition, and offgas values. Morphological differences between the exponential state and steady state can be seen in the respective microscopic pictures and are quantified in Figure 5.14. Here, hyphal diameters are clearly smaller in steady state compared to the respective exponential state, but no differences between the respective states compared to the reference strain can be observed. The spiking of cellulose to the medium was leading to a continuing increase of base addition and raising offgas values (after an initial spiking related drop) in JK2.8 only,

indicating the ability to use cellulose as a carbon source. In BS6.4 and BS7.8 base addition stopped and offgas values were constantly decreasing, indicating that both strains were not able to use cellulose as a carbon source and therefore, started to die. Protein analysis (Figures 5.15 and 5.16) showed that only JK2.8 started to secrete proteins after the cellulose spike, resulting in an increase of approximately 100 µg/mL protein and an altered protein secretion (see band pattern SDS PAGEs) 4 h after the spike compared to the steady state condition. Compared to the reference strain JK2.8 secreted a lower amount of protein in steady state (~50 µg/mL) and also difference in protein content from t4 to steady state was roughly 100 µg/mL lower compared to the reference strain (200 µg/mL vs. 100 µg/mL). The band pattern of the SDS PAGEs for JK2.8 in steady state and 4 h after the cellulose spike is identical but in steady state weaker compared to the reference strain confirming the protein assay data. Regarding BS6.4 and BS7.8, in steady state no differences in protein secretion compared to the reference strain can be observed, but 4 h after spiking cellulose the amount of secreted protein (~200 µg/mL lower) and the protein band pattern are completely different compared to the reference strain but identical among the two strains.

5.2.3 Cellulase regulator deletion mutant BS9.3 (Δ*ku80,* Δ*vib1*) and putative cellulase regulator deletion mutant BS4.1 (Δ*ku80,* Δ*vib2*); cellulose spiking experiment

The chemostat cultivations of *T. thermophilus* BS4.1 (Figure 5.8) and BS9.3 (Figure 5.12) were also successful. For BS4.1 clogging of a tube at approximately 140 h runtime for one of the duplicates caused an increase in the offgas values. However, physiology was not further affected, and no other technical problems occurred during the cultivations. During cultivation of BS9.3 a bigger technical problem occurred: the computer was shutting down at around 100 h until 120 h runtime. Together with the computer shut down also the chemostat program stopped, meaning there was no mass regulation but medium addition by an external pump, leading to an overflow. Nevertheless, growth stabilized quickly after re-starting the programs. After chemostat cultivation was initiated, BS4.1 started sporulation and fragmentation identical to the other strains. In contrast, weaker sporulation was observed for BS9.3 and fragmentation of the hyphae was absent, which can be seen when looking at the physiology, the respective microscopic pictures, and the color of the culture broth. No pinkish color was observed after turning on the chemostat program. Note that due to a higher base consumption (which served as a reference for biomass and thus for starting chemostat cultivation) of BS9.3 compared to all other strains (final consumption after ~17 retention times ~300 g vs. ~200 g), the chemostat program was started at a lower biomass value (biomass: ~4 g/kg vs. ~5 g/kg) and higher glucose concentration (3.8 g/L vs. ~0.8 g/L) compared to all other strains. Except for the color of BS9.3 after starting the chemostat program, the color development of the culture broth during the cultivation was almost identical to the reference strain. One exception for both strains was that the yellow color shifted directly to white instead of shifting from yellow to brown to brownish-white. The maximum growth rates of the strains (Figure 5.13) are almost identical compared to the reference strain, ranging from roughly 0.28 to 0.3 1/h. Note that maximum growth rate data is only available for two replicates of each strain except for MJK20.3. After around 190 h runtime, the steady state condition for both strains was reached regarding morphology

(no remaining spores, filamentous growth), biomass, base addition, and offgas values. Morphological differences between the exponential state and steady state can be seen in the respective microscopic pictures and are quantified in Figure 5.14. Here, hyphal diameters are clearly smaller in steady state compared to the respective exponential state, but no differences between the respective states compared to the reference strain can be observed. The spiking of cellulose to the medium was leading to a physiological response in both strains, indicating the ability to use cellulose as a carbon source. Protein analysis (Figures 5.15 and 5.16) showed that BS4.1 started to secrete proteins after the cellulose spike, resulting in an increase of approximately 200 µg/mL protein and an altered protein secretion (see band pattern SDS PAGEs) 4 h after the spike compared to the steady state condition. Compared to the reference strain, BS4.1 secreted the same amount of protein and showed the same band pattern as well as signal intensity across all conditions. Timewise, protein analysis was not performed for samples taken from the cultivation of BS9.3.

5.2.4 Reference strain MJK20.3 (Δ*ku80*); xylose as carbon source

The chemostat cultivation of *T. thermophilus* MJK20.3 with xylose as carbon source (Figure 5.5) was successful. During the batch phase, sporulation was already detected at 18 h runtime and a dry biomass value of 0.1 g/kg which was not observed during cultivation on glucose. Therefore, the batch phase lasted much longer compared to the cultivation on glucose (~70 h vs 25 h). The color development of the culture broth during the cultivation was identical to the cultivation on glucose except an earlier pinkish color due to the early sporulation when grown on xylose. Close to the end of the experiment, the feed tubing clogged in R1, therefore, 1 kg of medium was removed, and 1 kg of fresh medium was added. To get reliable data for the exponential phase it was decided to perform a wash out with the maximum dilution rate (0.366 1/h) at the end of the run. The growth rates on glucose and xylose are identical. After around 17 retention times, the steady state condition was reached regarding morphology (no remaining spores, filamentous growth), biomass, base addition, and offgas values. One day after taking the steady state sample, the wash out experiment was started to force growth at the maximum growth rate. This state is identical to the exponential growth phase and therefore, the samples are named accordingly. Morphological differences between the exponential state and steady state can be seen in the respective microscopic pictures and are quantified in Figure 5.14. The morphology in steady state was identical to the morphology of the strain when grown on glucose. In exponential state (after starting the wash out experiment) the morphology was completely different. Here, short hyphal fragments with swellings and thickenings could be observed (see Figure 5.5). Hyphal diameters in steady state on xylose are identical to the ones obtained from a glucose cultivation. Due to the thickenings and swellings of the hyphae no significant differences (high deviation) to the diameters in steady state during growth on glucose and xylose as well as exponential state during growth on glucose can be detected. Protein analysis (Figures 5.15 and 5.16) showed that the amount of secreted proteins is similar in both steady state and exponential state when grown on glucose or xylose. However, SDS-PAGE analysis showed a completely different band pattern for the steady state condition on glucose compared to xylose.

5.2.5 Reference strain MJK20.3 (Δ*ku80*); pH shift experiment

The batch cultivation of *T. thermophilus* MJK20.3 without (Figure 5.6) and with pH shift (Figure 5.7) was successful. Except for some technical problems with the offgas analyzer starting at roughly 32 h runtime (false values were removed in the figures) no other problems occurred during the cultivation. The color development of the culture broth during the cultivation with and without the pH shift was identical to the previous cultivation of MJK20.3 and is therefore, not shown. The growth rate of the strain with pH shift is lower compared to the strain without pH shift. Morphological differences (see Figures 5.6 and 5.7) between the two conditions could not be observed and hyphal diameters (Figure 5.14) were identical. The physiological response (offgas data) to the pH shift was due to the spiking of 1 M HCl with the syringe where, offgas was transferred together with the acid and therefore, interfered with the measurements. Timewise, protein analysis was not performed for samples taken from the cultivation of the pH shift experiment.

5.3 Transcriptomic analysis of the different *T. thermophilus* strains, cultivated using different carbon sources and conditions

After successful cultivation of the different strains, a transcriptomic analysis via RNA sequencing using the isolated RNA of the samples that were taken in the exponential states (ex), the steady states (SS), as well as 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h (t4) after spiking cellulose (t4) was performed. Data for the cultivation of BS9.4 and the pH shift experiment were timewise not analyzed and will therefore not be a part of the following chapters. RNA sequencing was done at two different companies: Microsynth AG, Balgach (MJK20.3, BS4.1, BS6.4) and GenomeScan, Leiden (BS7.8, JK2.8, MJK20.3 xylose) for RNA sequencing. The analysis proceeded as described in chapter 4.2.6.3. All samples were normalized together and those normalized raw count values as well as log2 fold change (L2FC) values of differentially expressed genes with an adjusted p-value (padj.) ≤ 0.05 were used for the following analysis.

In the following chapters the results for the regulator deletion mutants BS6.4 (Δ*ku80*, Δ*clr2*), BS7.8 (Δ*ku80,* Δ*clr1*), and JK2.8 (Δ*ku80,* Δ*clr4*), the putative regulator deletion mutant BS4.1 (Δ*ku80,* Δ*vib2*) and the cultivation on xylose (MJK20.3 xylose) will be discussed separately and compared to each other and the reference strain. Each of those chapters will start with an overview and general analysis including Venn diagrams and GO enrichment analysis. The general analysis will be followed by specific analyses, where the expression of predicted hydrolases (proteases and CAZYs) as well as predicted transcription factors and orthologs of regulators that are known to be involved in plant biomass degradation are investigated in detail.

An overview of the results and the relations of the expression profiles of the analyzed samples are shown in the following principal component analysis (PCA) plot (Figure 5.17). A clustering in the PCA plot always indicates, how similar the expression profiles of the samples are, compared to each other. Here, all exponential state and steady state samples, except the respective samples of MJK20.3 cultivated on xylose cluster together. Furthermore, a clustering between all samples after spiking cellulose except for the strains BS6.4 and BS7.8 can be detected. Nevertheless, the samples of these two strains BS6.4 and BS7.8 are clustering together.

To prove the quality of the RNA sequencing analysis results, an exemplary quantitative real-time PCR (qPCR) on four different genes was performed using one sample that was sent to RNA sequencing (Figure 8.13, appendix) and the results were compared with the RNA seq. raw count values of those genes for this sample (Table 8.4, appendix). The primers used are listed in Table 4.10. The results of the qPCR fit to the generated RNA seq. data: The genes with the highest count numbers have higher RFU (relative fluorescence units) values at an earlier cycle (lower Cq values) and the genes with lower count numbers have the same RFU values at later cycles (higher Cq values). The order of count number and Cq values of the genes is identical according to their relative expression level.

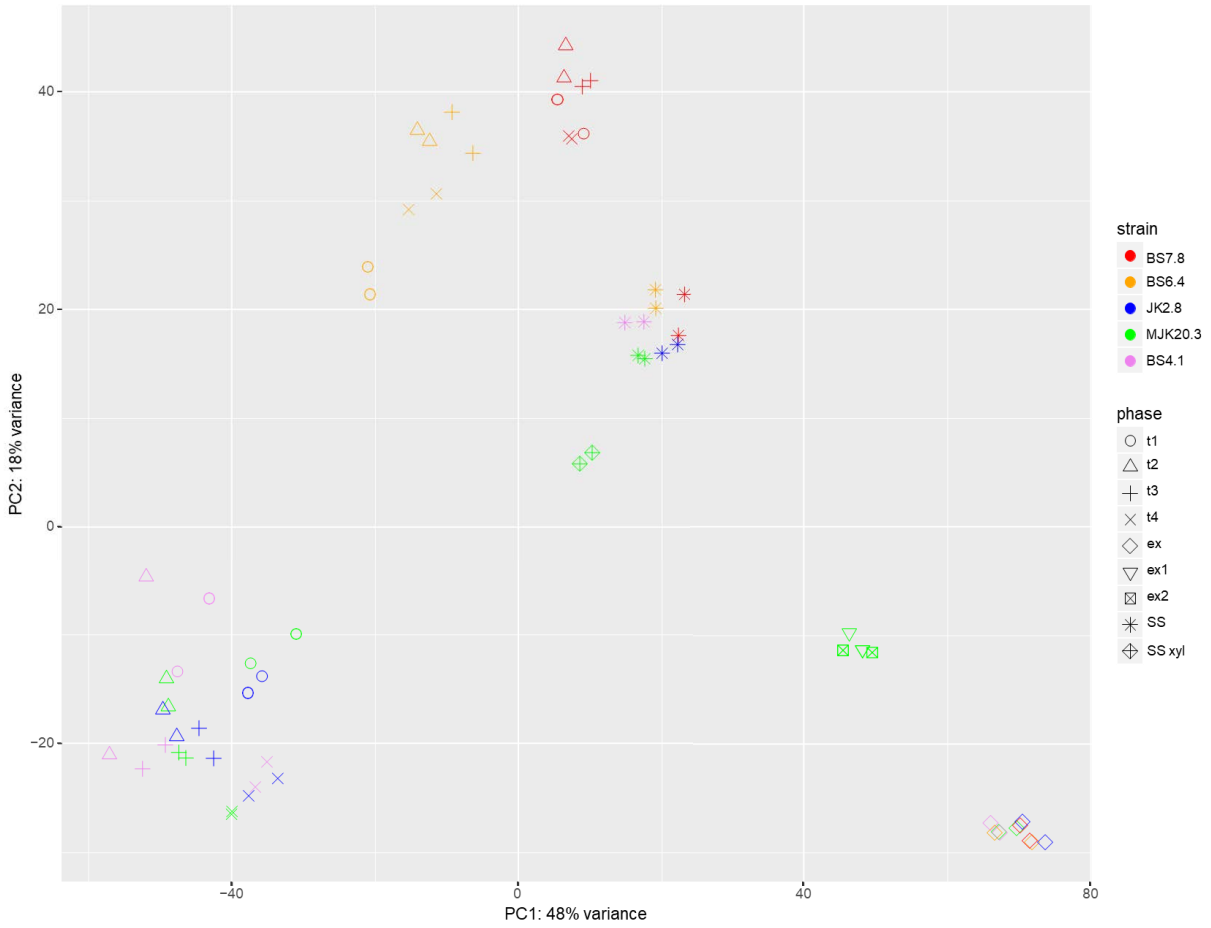

Figure 5.17: Principal component analysis (PCA) plot using the results of the RNA seq. analysis. Shown are the similarities/relations in expression profile of all samples that were analyzed. The closer the samples cluster, the more similar is their expression profile, whereas a higher distance in principal component 1 (PC1; 48 % variance) means a higher variance between the samples compared to the same distance in principal component 2 (PC2; 18 % variance). The analyzed strains include MJK20.3 (Δ*ku80*), BS6.4 (Δ*ku80*, Δ*clr2*), BS7.8 (Δ*ku80,* Δ*clr1*), JK2.8 (Δ*ku80,* Δ*clr4*) and BS4.1 (Δ*ku80,* Δ*vib2*). The analyzed conditions (phase) include exponential state (ex, (ex1, ex2→MJK20.3 xylose)), steady state (SS, (SS xyl→MJK20.3 xylose)) as well as 0.5 h (t1), 1 h (t2), 2 h (t3) and 4 h (t4) after cellulose spike.

5.3.1 Reference strain MJK20.3 (Δ*ku80*) and putative cellulase regulator deletion mutant BS4.1 (Δ*ku80*, Δ*vib2*)

5.3.1.1 General and enrichment analysis

For differential gene expression analysis, the samples taken after the cellulose spike were always compared to the respective steady state condition. The number of the resulting differentially expressed genes with padj. ≤ 0.05 is shown in Table 5.3. Especially at the earlier points in time after the spike more genes are differentially expressed in strain BS4.1 compared to MJK20.3. At t4 the number of up- and downregulated genes is very similar between the two strains. The highest number of differentially expressed genes can be observed 1 h after the cellulose spike (t2) for both strains.

Table 5.3: Number of differentially expressed genes with padj. ≤ 0.05 of the strains MJK20.3 and BS4.1 at different points in time after the cellulose spike in relation to the respective steady state condition.

strain	condition	upregulated genes	downregulated genes
MJK20.3	t1	1419	1376
	t2	1797	1721
	t3	1297	1308
	t4	604	455
BS4.1	t1	1932	1870
	t2	2266	2055
	t3	1556	1508
	t4	615	589

With those numbers, Venn diagrams were generated to access the number of differentially expressed genes that are exclusive for one point in time as well as intersections between the four points in time. The diagrams are shown in Figure 5.18. In addition to a higher number of up- and downregulated genes regarding single points in time (especially the earlier), also the total number of up- and downregulated genes is higher in BS4.1. All in all, in both strains more than 50 % of all genes (total number of genes: 9292) are differentially expressed.

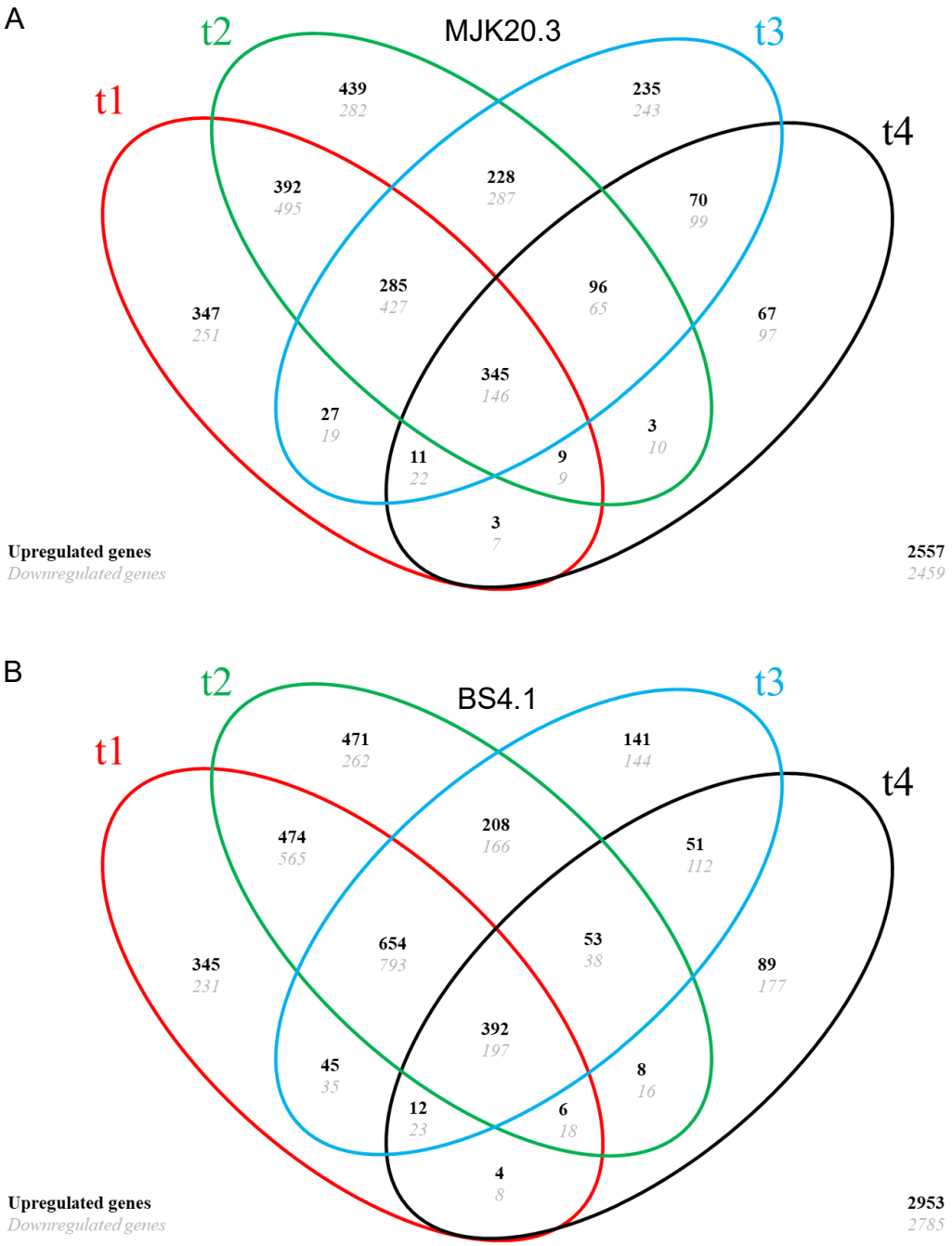

Figure 5.18: Venn diagrams of differentially expressed genes after cellulose spike in relation to the respective steady state condition in strains MJK20.3 and BS4.1. Black (grey) numbers refer to upregulated (down-) genes in strain MJK20.3 (A) and BS4.1 (B). The total number of upregulated and downregulated genes is shown in the bottom right corner of (A) and (B). Differential expression (padj. ≤ 0 05) was determined via comparison of expression profiles of t1= 0.5 h, t2 =1 h, t3= 2 h and t4= 4 h after spiking cellulose with the respective steady state condition.

In many of the following analyses it is referred to the t1-t4 intersection. This intersection covers the genes that are up- or downregulated at every point in time after the spike and are slightly higher in BS4.1 (see Figure 5.18). Those genes are of special interest due to their significant differential expression across all points in time, meaning those genes are presumably the most important genes that are

involved in the transcriptomic response to cellulose. To get an idea of the overall transcriptomic response of the strains MJK20.3 and BS4.1 to cellulose, a GO enrichment analysis using the DAVID database (Huang et al. 2009a; Huang et al. 2009b) with the respective t1-t4 intersection gene set was performed using a p-value cutoff of 0.05. The results of the GO enrichment analysis are shown in Tables 5.4-5.7. Note that due to the limited research done on *T. thermophilus*, only approximately 3150 to 4100 genes (~30-45 % of all genes), depending on the respective GO term category, have a GO term annotation. Due to this fact the fold enrichment (for a detailed description on how this value is calculated see Mancuso 2010) is sometimes very high despite the low gene numbers that belong to the respective categories. Therefore, gene number and fold enrichment should always be analyzed together. For MJK20.3 and BS4.1 the results of the GO enrichment analysis are very similar.

Table 5.4: GO term enrichment analysis of the upregulated genes in the t1-t4 intersection gene set of strain MJK20.3. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	carbohydrate metabolic process	7	42
	carbohydrate transport	9	5
	cellulose catabolic process	16	9
	ER-associated ubiquitin-dependent protein catabolic process	6	4
	metal ion transport	12	3
	polysaccharide catabolic process	16	9
	xylan catabolic process	11	4
CC	extracellular region	23	26
	integral component of endoplasmic reticulum membrane	9	5
	integral component of Golgi membrane	9	3
	integral component of membrane	1	60
MF	beta-glucosidase activity	10	3
	carbohydrate binding	9	8
	cellulase activity	21	5
	cellulose binding	19	15
	endo-1,4-beta-xylanase activity	10	5
	hydrolase activity	3	22
	hydrolase activity, hydrolyzing O-glycosyl compounds	8	28
	mannan endo-1,4-beta-mannosidase activity	21	3
	pectate lyase activity	21	6
	protein disulfide isomerase activity	16	3
	scopolin beta-glucosidase activity	10	3
	xylan 1,4-beta-xylosidase activity	21	3

Table 5.5: GO term enrichment analysis of the downregulated genes in the t1-t4 intersection gene set of strain MJK20.3. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	one-carbon metabolic process	26	3
CC	glycine cleavage complex	54	2
	integral component of membrane	1	40
MF	starch binding	48	2
	transferase activity	4	5

Table 5.6: GO term enrichment analysis of the upregulated genes in the t1-t4 intersection gene set of strain BS4.1. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	carbohydrate metabolic process	8	45
	carbohydrate transport	11	6
	cellulose catabolic process	18	10
	polysaccharide catabolic process	13	7
	sphingolipid metabolic process	12	3
	xylan catabolic process	11	4
CC	extracellular region	19	24
	integral component of membrane	1	65
MF	beta-glucosidase activity	10	3
	carbohydrate binding	9	8
	cellobiose dehydrogenase (acceptor) activity	21	3
	cellulase activity	21	5
	cellulose binding	19	15
	endo-1,4-beta-xylanase activity	10	5
	glucosylceramidase activity	21	3
	hydrolase activity	3	19
	hydrolase activity, hydrolyzing O-glycosyl compounds	8	30
	mannan endo-1,4-beta-mannosidase activity	21	3
	pectate lyase activity	14	4
	RNA polymerase II transcription factor activity, sequence-specific DNA binding	2	12
	scopolin beta-glucosidase activity	10	3
	sequence-specific DNA binding	3	6

Table 5.7: GO term enrichment analysis of the downregulated genes in the t1-t4 intersection gene set of strain BS4.1. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	glycolytic process	10	3
CC	integral component of membrane	1	51

Biological processes that are enriched among the upregulated genes are mainly polysaccharide catabolic processes like “cellulose catabolic process” and “xylan catabolic process” as well as “carbohydrate transporters”. The main cellular components that are enriched are integral components of membranes like the ER membrane or the Golgi membrane in MJK20.3. Regarding the molecular function, mainly lignocellulolytic enzyme activity is enriched, including, e.g. “cellulose binding” and “cellulase activity” as well as xylanase, mannanase, and pectinase activities. Among the downregulated genes, only a few GO terms are enriched, including “integral components of the membrane”, “starch binding”, and “transferase activity” in MJK20.3 and “integral component of the membrane” and “glycolytic process” in BS4.1. Note, that the number of annotated genes is very low for “glycolytic process”, therefore, the high fold enrichment is not very significant. In summary, no big differences between both strains can be detected.

5.3.1.2 Differential expression of proteases

Knowledge about the expression of proteases under various industrial conditions, as during the bioreactor experiments of this work, enables direct targeting of genes for deletion that are highly expressed under the respective conditions. The deletion of those proteases, as already mentioned in the introduction, can lead to higher yields of secreted proteins of interest via preventing or reducing protein degradation. Nevertheless, it is also possible to examine the conditions under which certain proteases are highly expressed to improve production conditions of this protease for industrial applications. The annotations of the predicted proteases of *T. thermophilus* were retrieved using the JGI (https://jgi.doe.gov/), NCBI (https://www.ncbi.nlm.nih.gov/) and MEROPS (https://www.ebi.ac.uk/merops/) databases as well as the publication of Li et al. 2020. Here, also the sequences of the predicted proteases were investigated for the presence of signal peptides to determine presumably secreted proteases. In Figure 5.19 the number and classes of predicted *T. thermophilus* proteases and the percentage of proteases with a signal peptide is shown.

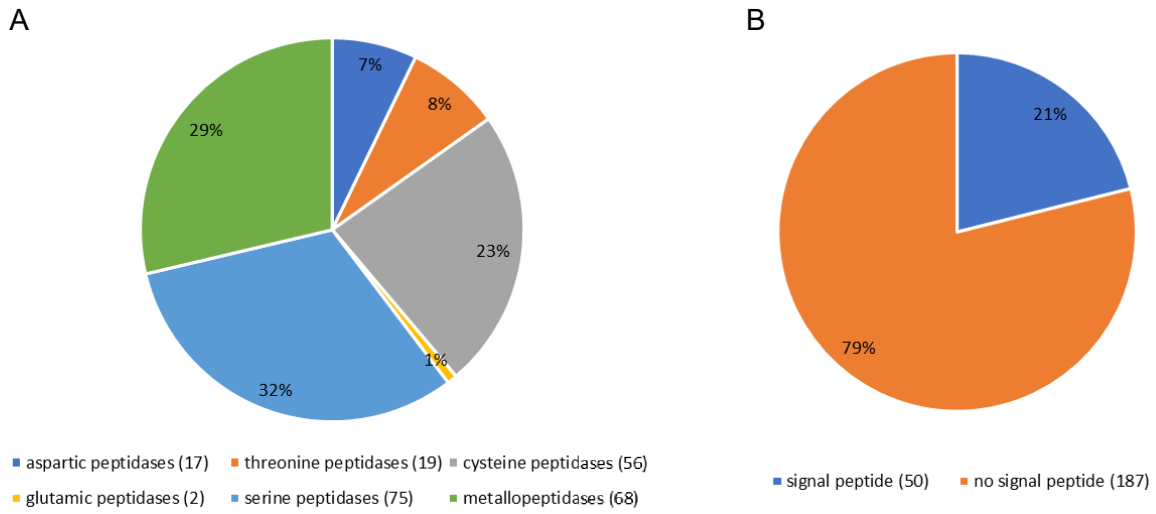

Figure 5.19: Predicted proteases in *T. thermophilus*. Shown are the number and percentage of proteases of each protease class (A) as well as the number and percentage of proteases with a signal peptide (B) in relation to the total number of proteases.

The total number of predicted protease genes that were differentially expressed in the two strains MJK20.3 and BS4.1 at the different points in time after the cellulose spike, is shown in Figure 5.20. A higher number of up- as well as downregulated genes encoding predicted proteases in BS4.1 for points in time t1-t3 can be detected. At t4, the number of downregulated genes is higher, and the number of upregulated genes lower in BS4.1. In total up to roughly 40 % of the predicted protease genes are differentially expressed in MJK20.3 and more than 50 % in BS4.1.

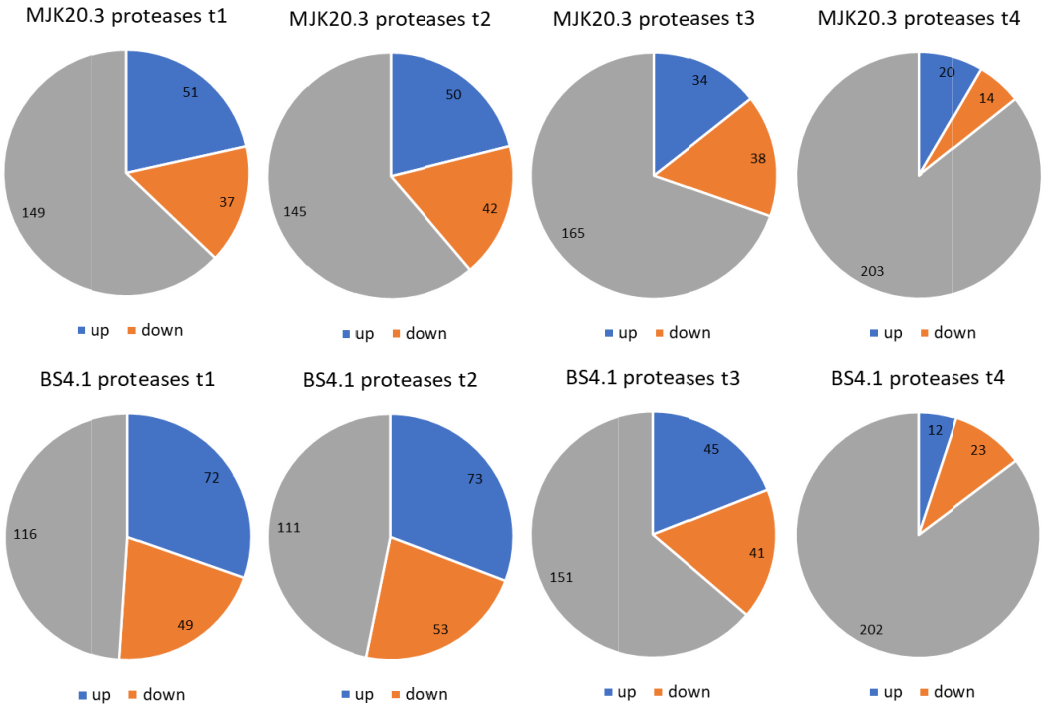

Figure 5.20: Number of differentially expressed predicted protease genes of strains MJK20.3 and BS4.1 at different points in time after the cellulose spike. Shown are the numbers of up- (blue) and downregulated (orange) genes as well as genes with no differential expression (grey) at: 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after cellulose spike (t4) compared to the respective steady state condition.

To get a more detailed overview of the expression profiles of the two strains regarding protease gene expression, a heatmap (Figure 5.21) was created, showing the differential expression levels of genes belonging to the different classes of proteases among the two strains. The expression profile of both strains is very similar but a stronger up- and downregulation of the single genes especially for t1-t3 is observable for BS4.1 as well as a higher total number of differentially expressed predicted protease genes at the respective points in time as already seen in Figure 5.20. At t4 a very low number of predicted protease genes is differentially expressed in both strains. Differential expression is evenly distributed among the different protease classes.

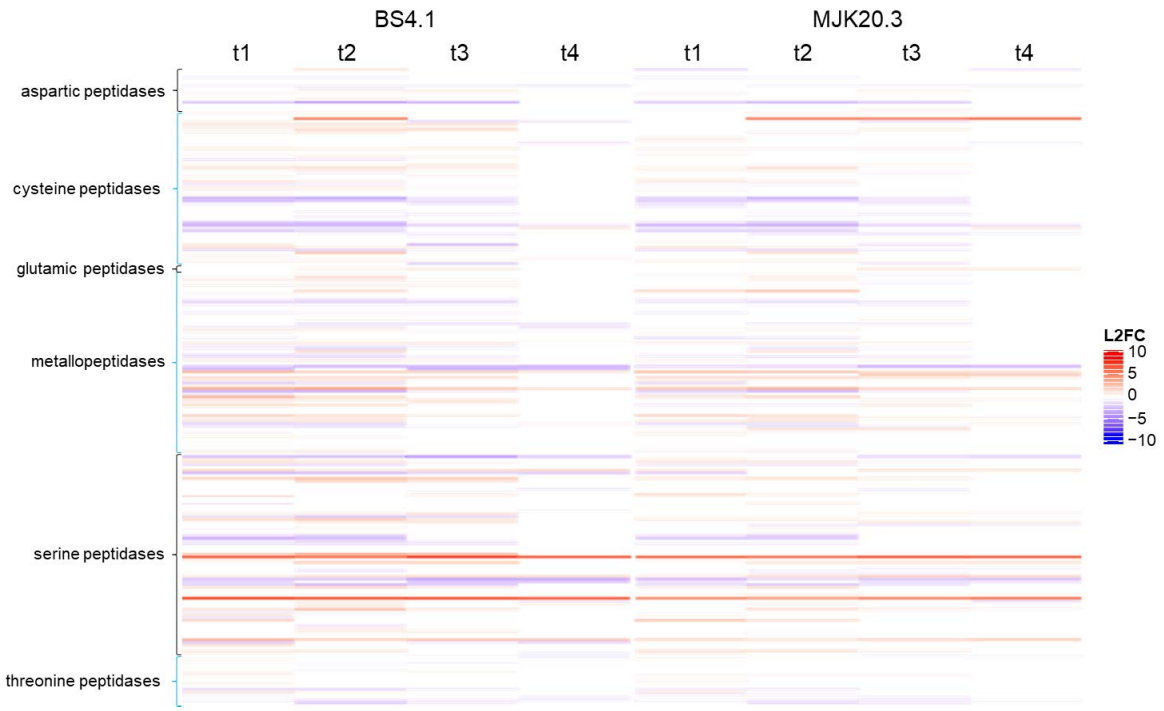

Figure 5.21: Heatmap with differential expression values of all predicted protease genes in strains MJK20.3 and BS4.1. Shown are the log2 fold change values (L2FC) of single protease genes belonging to the different protease classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

To narrow down the number of differentially expressed genes and filter out the most important differentially expressed genes, the t1-t4 intersection was analyzed next. A gene was included, when differential expression over all points in time after spiking cellulose in at least one of the two strains was detected. In Table 5.8, the numbers of genes that are differentially expressed at all points in time after the spike are shown for the two strains. A further heatmap (Figure 5.22) visualizes the differential expression levels of the respective t1-t4 intersection genes.

Table 5.8: Number of genes belonging to different protease classes that are differentially expressed at all points in time after the spike in strains MJK20.3 and BS4.1.

		MJK20.3		**BS4.1**	
class	genes	t1-t4 up	t1-t4 down	t1-t4 up	t1-t4 down
aspartic peptidases	17	0	0	0	1
threonine peptidases	19	0	1	0	1
cysteine peptidases	56	0	1	0	1
glutamic peptidases	2	0	0	0	0
serine peptidases	75	5	1	5	3
metallopeptidases	68	3	1	2	1
total	237	8	4	7	7

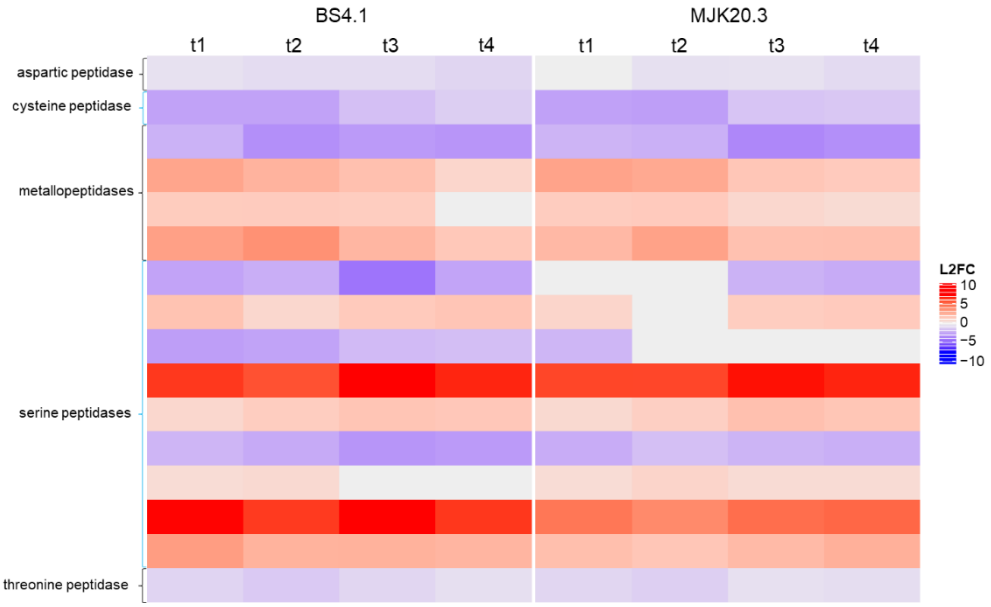

Figure 5.22: Heatmap with expression values of predicted protease genes that are differentially expressed at all points in time after the cellulose spike in strains MJK20.3 and BS4.1. Shown are the log2 fold change values (L2FC) of single protease genes belonging to the different protease classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

The heatmap of the t1-t4 intersection genes confirms the results of the previous graphs. The expression pattern between both strains is very similar and only few predicted protease genes have slightly higher or lower expression levels in BS4.1. This observation can also be made while analyzing Figure 5.21, where expression levels regarding L2FC and normalized raw counts are almost identical between both strains.

To investigate and compare gene expression in detail, graphs showing L2FC values as well as graphs showing the mean of the normalized raw count values of the two replicates of the single genes at every point in time after the spike, were created (Figure 5.23). Those two graphs must always be analyzed together, to identify, whether expression levels differ between the strains/conditions. Sometimes, due to an insufficient padj. value (> 0.05), the L2FC values are higher for one condition/strain compared to another (when padj. > 0.05, L2FC= 0), although the normalized raw counts are identical. In addition, low normalized raw counts in steady state lead to much higher L2FC values when differential expression occurs. This leads to false interpretations, which can be prevented by analyzing both graphs next to each other. When discussing those figures, the terms medium expressed and highly expressed are mentioned very often. As an internal threshold, to describe the most relevant genes, a normalized count value of 1000 was applied. Medium expressed describes values that are very close to this threshold (~1000-1500), whereas highly expressed describes values that are above this threshold (>1500).

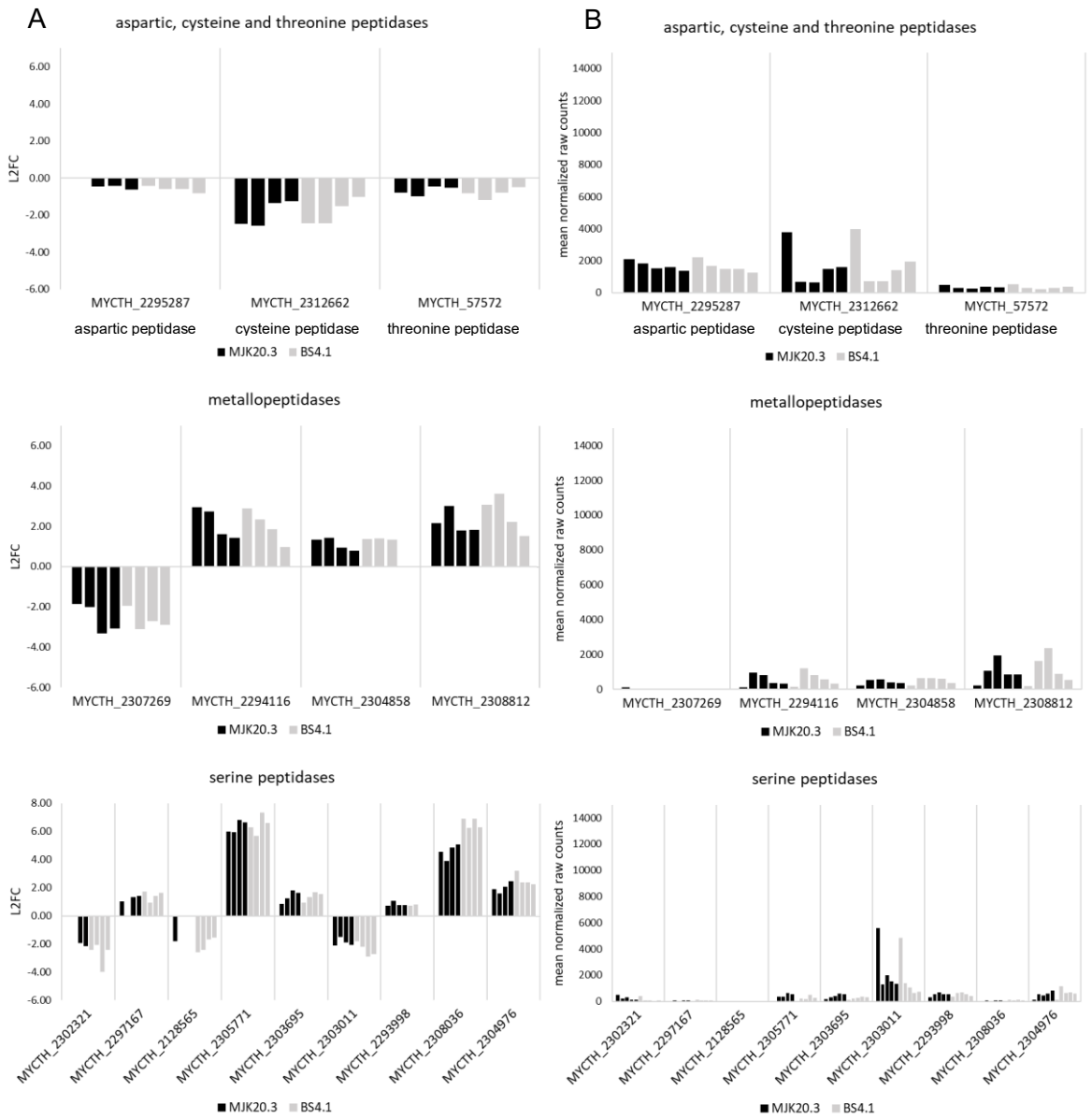

Figure 5.23: Differentially expressed predicted protease genes in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

Amongst the downregulated t1-t4 intersection genes, the predicted protease genes with the highest remaining expression after the cellulose spike are MYCTH_2295287, MYCTH_2303011, and MYCTH_2312662. The upregulated predicted protease genes that possess the highest expression levels, are MYCTH_2308812 and MYCTH_2294116.

To identify further interesting highly expressed proteases that were not covered by the t1-t4 intersection genes analysis, the same graphs, as seen in Figure 5.23, were created with all genes encoding for predicted proteases that possess a signal peptide. The results are shown in Figure 5.24.

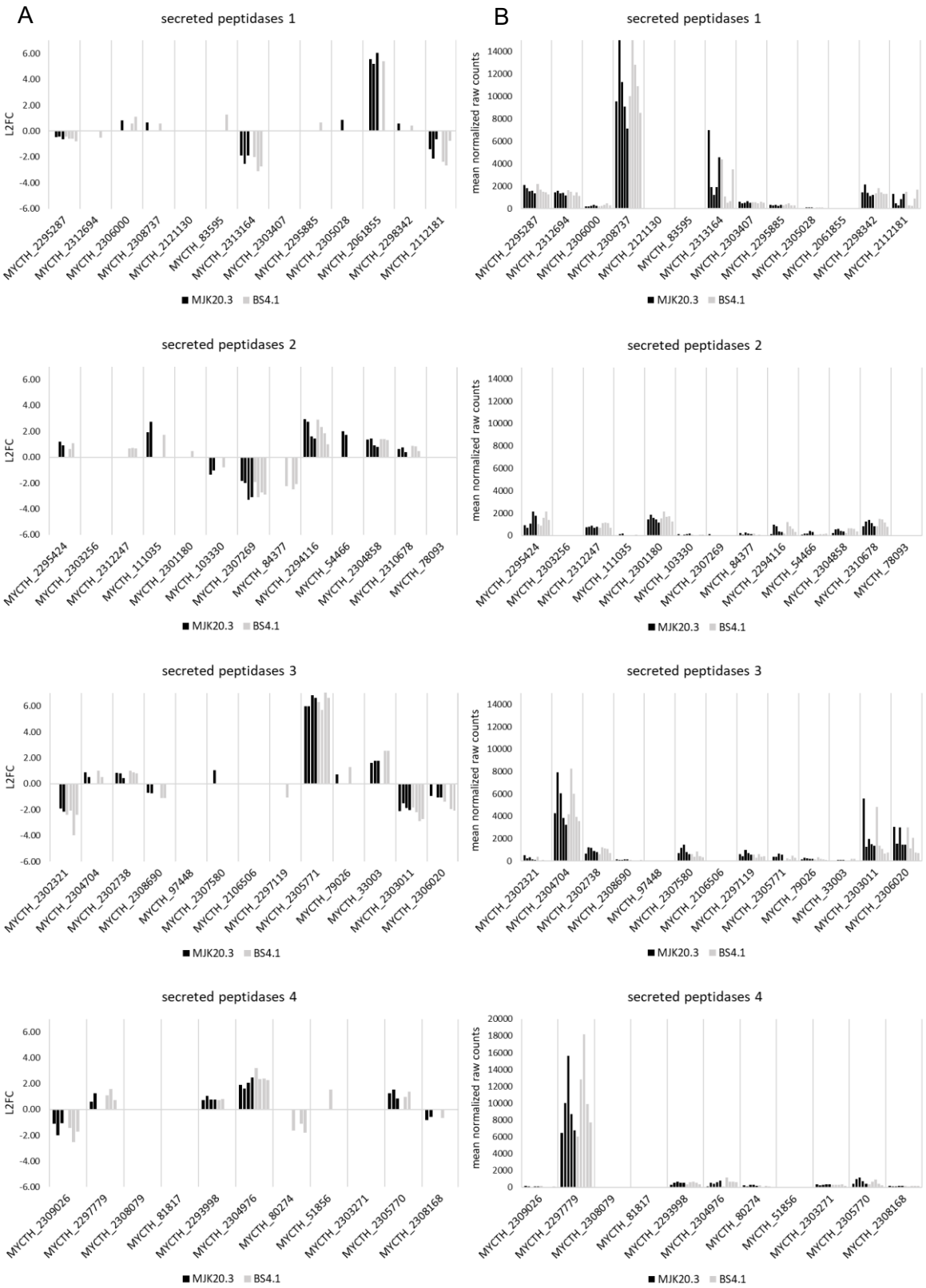

Figure 5.24: Differentially expressed genes encoding for predicted proteases with a signal peptide in strains MJK20.3 and BS4.1 (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

Regarding expression levels of genes encoding for predicted proteases with a signal peptide (Figure 5.24), there are also no major differences between MJK20.3 and BS4.1, except some only slightly higher/lower expression levels. Nevertheless, the most interesting results are the highly expressed protease genes. Amongst them the top three are MYCTH_2308737, MYCTH_2304704, and MYCTH_2297779. All of those predicted protease genes also show, besides a high expression level, a significant upregulation after spiking cellulose. Further highly expressed predicted protease genes can be found in Figure 5.24.

In summary, the expression of predicted protease genes between both strains is almost identical.

5.3.1.3 Differential expression of carbohydrate-active enzymes (CAZYs)

Knowledge about the expression of CAZYs under various industrial conditions, as in the bioreactor experiments of this work, enables direct targeting of genes for deletion that are highly expressed under the respective conditions. The deletion of those enzymes, as already mentioned in the introduction, can lead to higher yields of secreted proteins of interest via reducing the secretory burden and facilitating downstream processing. On the other hand, it is also possible to examine the conditions under which certain CAZYs are highly expressed to improve production conditions of those enzymes for lignocellulosic biorefineries applications. Detailed knowledge about CAZY expression also enables the fine tuning of the composition of enzymes cocktails used in plant biomass degradation processes. The annotations of the predicted CAZYs of *T. thermophilus* were retrieved via combining information from the JGI (https://jgi.doe.gov/), NCBI (https://www.ncbi.nlm.nih.gov/), and CAZY (http://www.cazy.org/) databases as well as the publications of Berka et al. 2011 and Karnaouri et al. 2014a. In Figure 5.25, the number and classes of predicted *T. thermophilus* CAZYs are shown. Genes that could not be assigned to a class are grouped in the class “other”.

class	type	genes
cellulases	endoglucanases	9
	cellobiohydrolases	7
	ß-glucosidases	9
	LPMOs	24
hemicellulases	xylanases	12
	xylosidases	4
	endoarabinases	3
	exoarabinases/arabinofuranosidases	11
	mixed-linked glucanase	5
	mannanases	10
	mannosidases	11
	galactanases	2
	galactosidases	7
pectinases	polygalacturonases	2
	rhamnosidases	1
	pectin lyases	8
esterases	feruloyl esterases	4
	acetyl esterases	9
	pectin esterases	4
	glycuronoyl esterases	2
starch metabolism	alpha amylases	4
	alpha glucosidases	4
	glucoamylases	2
	glycogen debranching enzymes	2
cell wall remodelling	glucanases	12
	transglucosylases	4
	chitosanases	2
	diacetylmuramidase	1
	glucosaminidase	3
	chitinases	8
	crosslinking transglycosidase	3
glycosyltransferases	glycosyltransferases	83
other	other	124
total		396

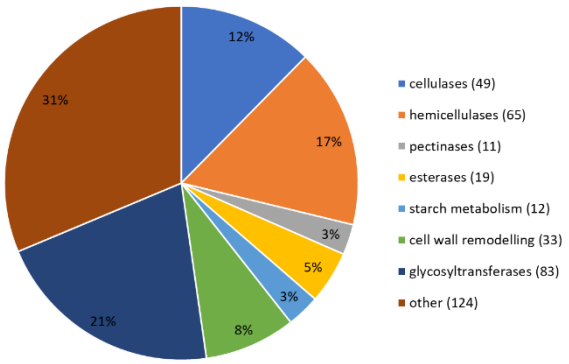

Figure 5.25: CAZYs in *T. thermophilus*. Shown are all predicted *T. thermophilus* CAZYs with the respective class and type as well as an overview of the percentage and number of those classes in relation to the total number of predicted CAZYs.

The total number of predicted CAZY genes that were differentially expressed in the two strains MJK20.3 and BS4.1 at the different points in time after the cellulose spike, is shown in Figure 5.26. Here, a higher number of upregulated genes in BS4.1 for points in time t1-t3 can be detected. The number of downregulated genes is almost identical between both strains. The strongest differential expression can

be detected at t2 in both strains. At t4, the number of up- and downregulated genes is almost identical. In total more than 50 % of all predicted CAZY genes are differentially expressed in MJK20.3 and BS4.1.

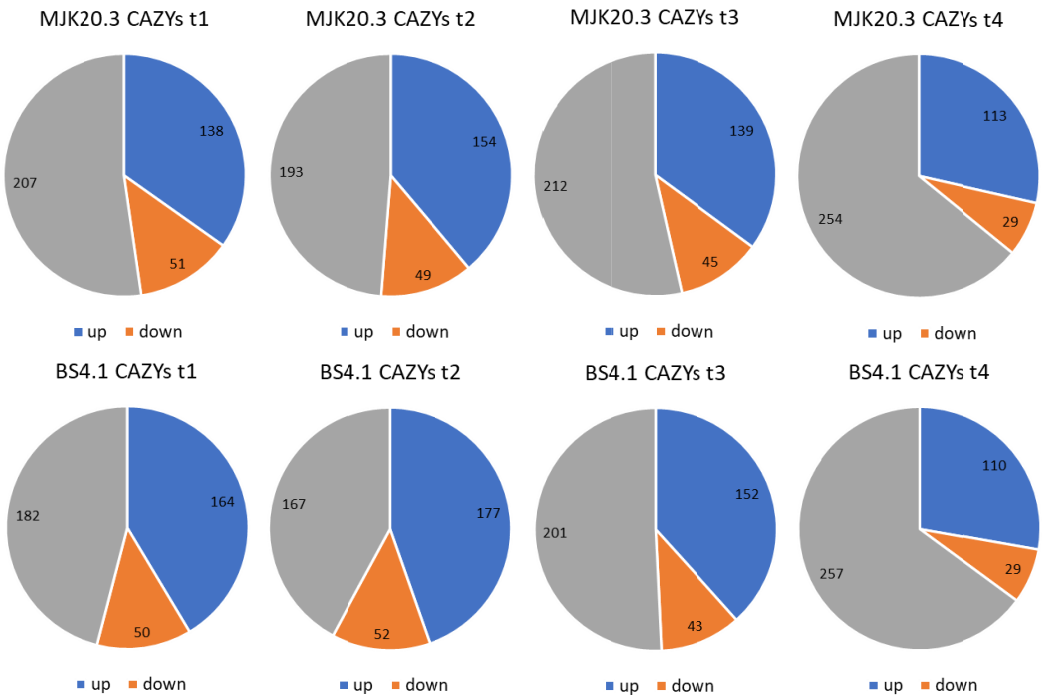

Figure 5.26: Number of differentially expressed predicted CAZY genes of strains MJK20.3 and BS4.1 at different points in time after the cellulose spike. Shown are the numbers of up- (blue) and downregulated (orange) genes as well as genes with no differential expression (grey) at: 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after cellulose spike (t4) compared to the respective steady state condition.

For a detailed overview of the expression profiles of the two strains regarding predicted CAZY gene expression, a heatmap (Figure 5.27) with differential expression levels of genes belonging to the different classes of CAZYs among the two strains was created. The expression profile of both strains is almost identical but a slightly stronger up- and downregulation of single genes especially for t1-t3 is observable for BS4.1 as well as a higher total number of differentially expressed predicted CAZY genes at the respective points in time as already seen in Figure 5.26. At t4, a lower number of predicted CAZY genes is differentially expressed in both strains compared to the earlier points in time. The genes that are showing the strongest upregulation and a high expression even at t4 in both strains belong to the classes: cellulases, hemicellulases, pectinases, and esterases as well as single genes that belong to the “other” class. The genes that are showing the strongest downregulation belong to the starch metabolism class. Genes that belong to all other classes show mostly low differential expression values.

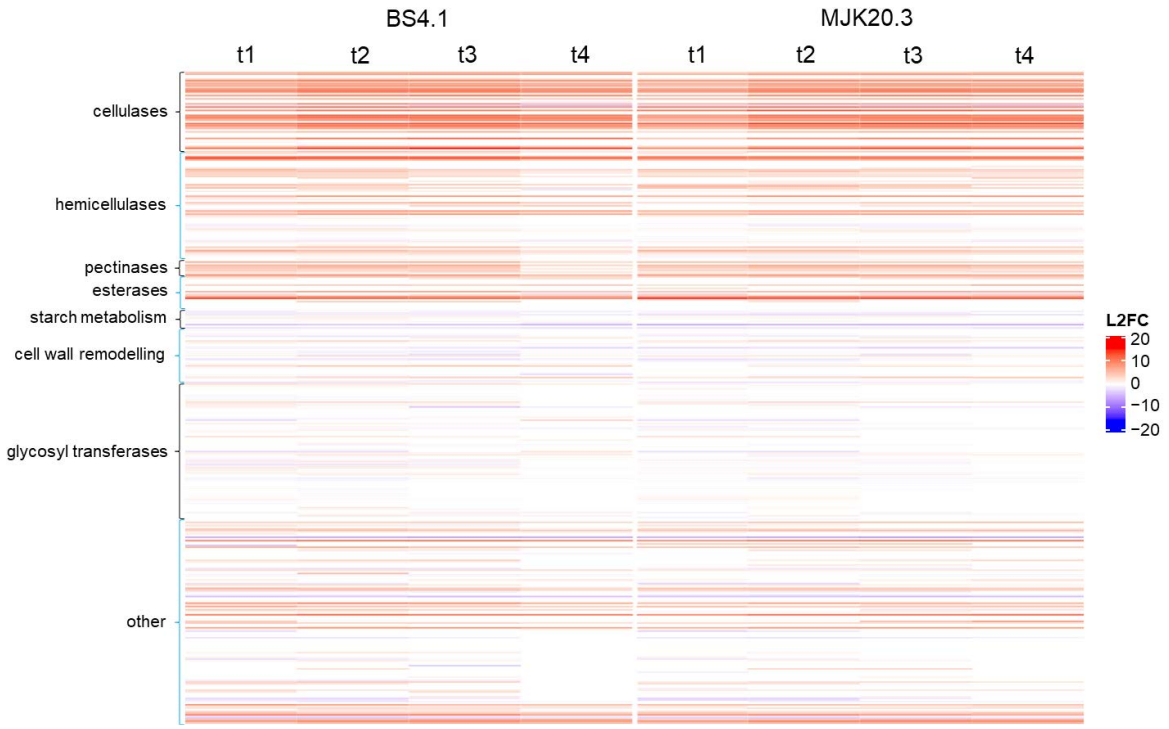

Figure 5.27: Heatmap with differential expression values of all predicted CAZY genes in strains MJK20.3 and BS4.1. Shown are the log2 fold change values (L2FC) of single predicted CAZY genes belonging to the different classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

To narrow down the number of differentially expressed genes and filter out the most important differentially expressed genes, the t1-t4 intersection was analyzed next. A gene was included, when differential expression over all points in time after spike in at least one of the two strains was detected. In Table 5.9, the numbers of genes that are differentially expressed at all points in time after the spike are shown for the two strains. A further heatmap (Figure 5.28) visualizes the differential expression levels of the respective t1-t4 intersection genes.

Table 5.9: Number of genes belonging to different CAZY classes that are differentially expressed at all points in time after the spike in strains MJK20.3 and BS4.1.

			MJK20.3		**BS4.1**	
class	type	genes	t1-t4 up	t1-t4 down	t1-t4 up	t1-t4 down
cellulases	endoglucanases	9	7	0	7	0
	cellobiohydrolases	7	5	0	6	0
	ß-glucosidases	9	4	1	4	1
	LPMOs	24	14	0	13	0
hemicellulases	xylanases	12	5	0	5	0
	xylosidases	4	3	0	2	0
	endoarabinases	3	0	0	0	0
	exoarabinases/ arabinofuranosidases	11	3	0	3	0
	mixed-linked glucanase	5	1	0	2	0
	mannanases	10	3	0	3	1
	mannosidases	11	3	1	3	1
	galactanases	2	1	0	1	0
	galactosidases	7	3	0	3	0
pectinases	polygalacturonases	2	1	0	1	0

pectinases	rhamnosidases	1	0	0	0	0
	pectin lyases	8	7	0	5	0
esterases	feruloyl esterases	4	0	0	0	0
	acetyl esterases	9	5	0	5	0
	pectin esterases	4	0	0	0	0
	glycuronoyl esterases	2	0	0	0	0
starch metabolism	alpha amylases	4	0	2	0	1
	alpha glucosidases	4	0	0	1	0
	glucoamylases	2	0	1	0	1
	glycogen debranching enzymes	2	0	1	0	1
cell wall remodeling	glucanases	12	1	1	0	1
	transglucosylases	4	0	1	0	1
	chitosanases	2	1	0	1	0
	diacetylmuramidase	1	0	0	0	1
	glucosaminidase	3	0	0	0	0
	chitinases	8	1	0	2	0
	crosslinking transglycosidase	3	0	1	0	0
glycosyltransferases	glycosyltransferases	83	4	2	4	1
other	other	124	24	6	24	6
	total	396	96	17	95	16

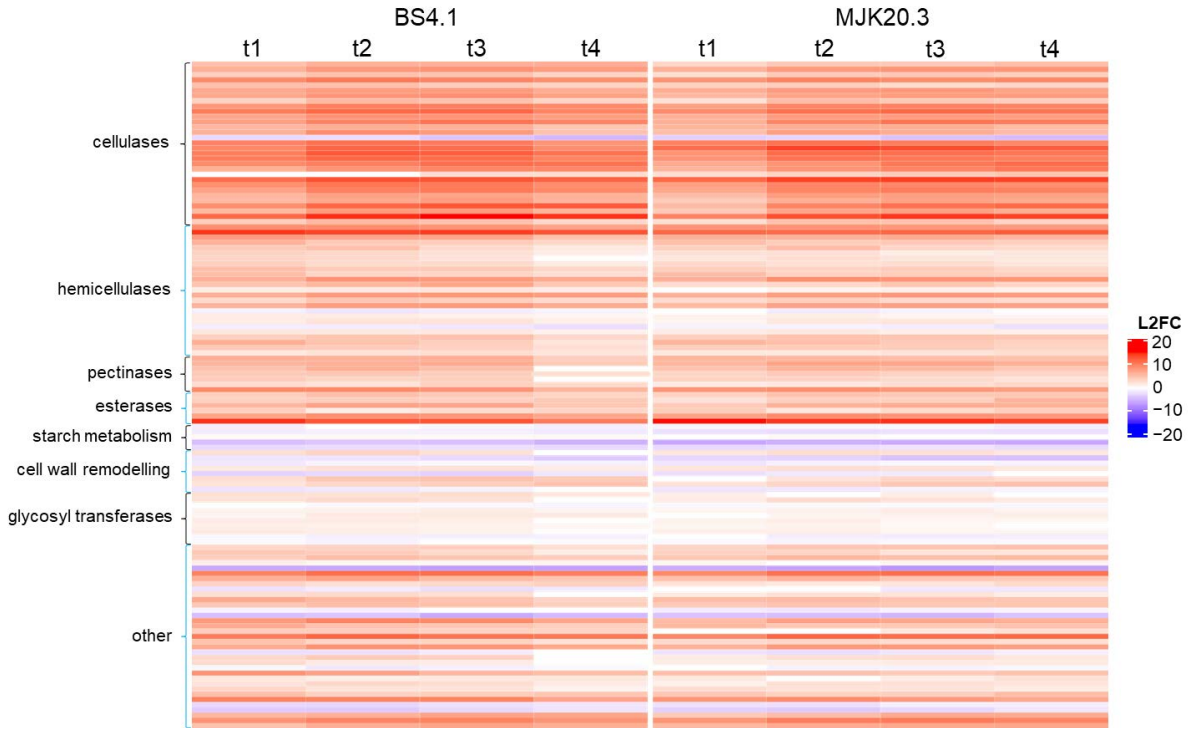

Figure 5.28: Heatmap with expression values of predicted CAZY genes that are differentially expressed at all points in time after the cellulose spike in strains MJK20.3 and BS4.1. Shown are the log2 fold change values (L2FC) of single predicted CAZY genes belonging to different classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

The heatmap of the t1-t4 intersection genes confirms the results of the previous graphs: the expression pattern between both strains is very similar with some predicted CAZY genes having higher expression levels in BS4.1. Generally, it becomes clearer, that the most and strongest differential expression derives from predicted cellulase genes, followed by predicted hemicellulase, pectinase, and esterase genes as well as single genes from the “other” class. Furthermore, it can be observed, that differential

expression decreases toward t4 in BS4.1, whereas in MJK20.3 differential expression levels remain at the approximately same levels.

To investigate and compare gene expression in detail, graphs showing L2FC values as well as graphs showing the mean of the normalized raw count values of the two replicates of the single genes at every point in time after the spike were created as previously (see chapter 5.3.1.2) described (Figures 5.29-5.36).

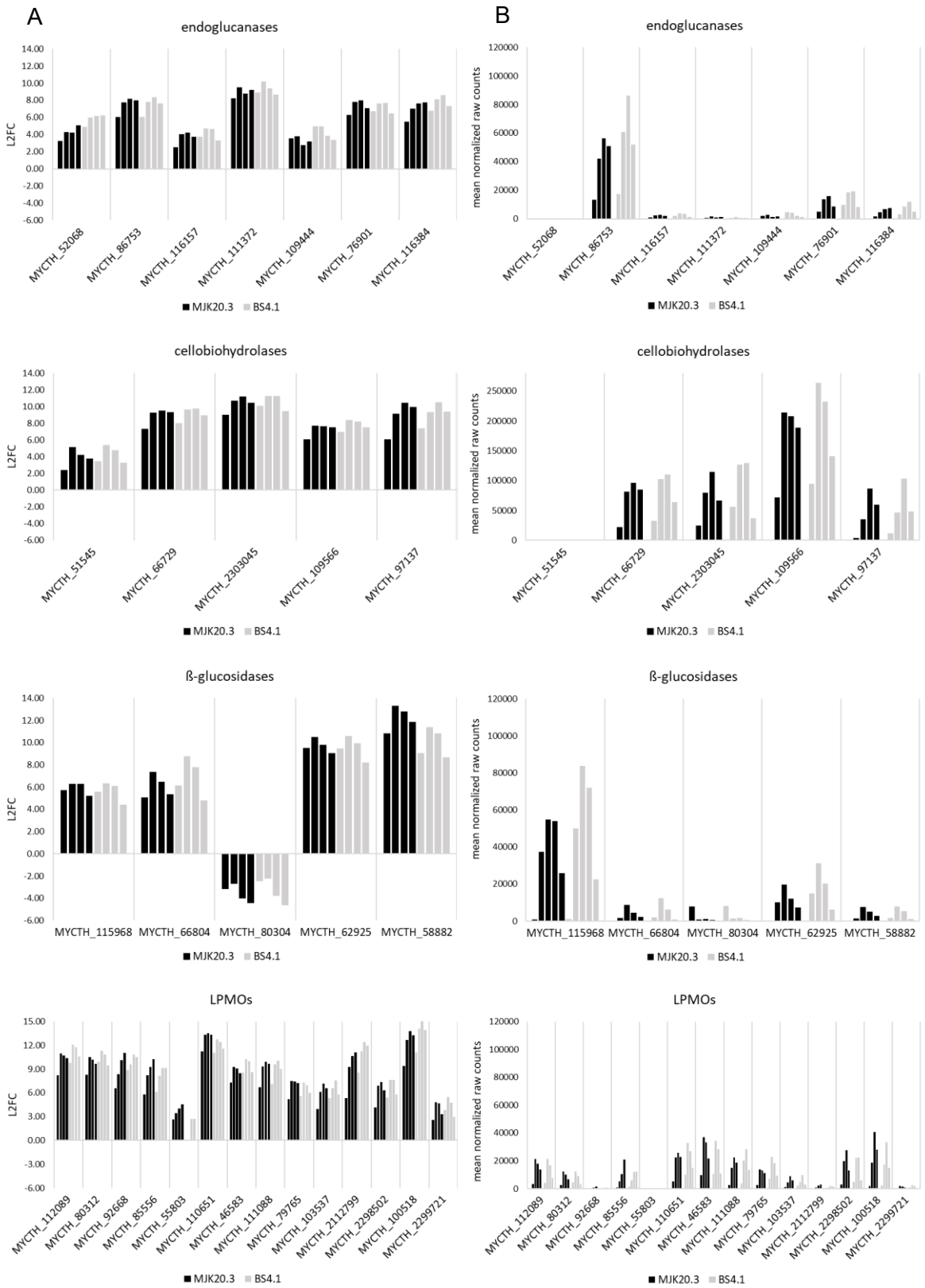

Figure 5.29: Differentially expressed predicted cellulase genes in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

Regarding predicted cellulase gene expression (Figure 5.29), expression levels and differential expression values are very similar for MJK20.3 and BS4.1. Differences are mostly in predicted endoglucanase, cellobiohydrolase and ß-glucosidase gene expression, where differential expression values as well as normalized raw counts are for almost every gene higher in BS4.1 at points in time t1-t3. At t4 those values are lower compared to MJK20.3 for almost every gene. Among the predicted LPMO genes, expression values are for some genes higher in MJK20.3 and for some in BS4.1. Generally, expression levels of the predicted cellulase genes are extremely high (up to 250000 counts after the spike), especially for the predicted LPMO genes, where lots of genes have a very high expression level. The top genes of each category are: MYCTH_86753, MYCTH_76901, and MYCTH_116384 (predicted endoglucanase genes); MYCTH_109566, MYCTH_2303045, MYCTH_66729, and MYCTH_97137 (predicted cellobiohydrolase genes); MYCTH_115968, and MYCTH_62925 (predicted ß-glucosidase genes). Among the predicted LPMO genes, almost all genes have a very high expression level, for which reason top genes are not mentioned separately.

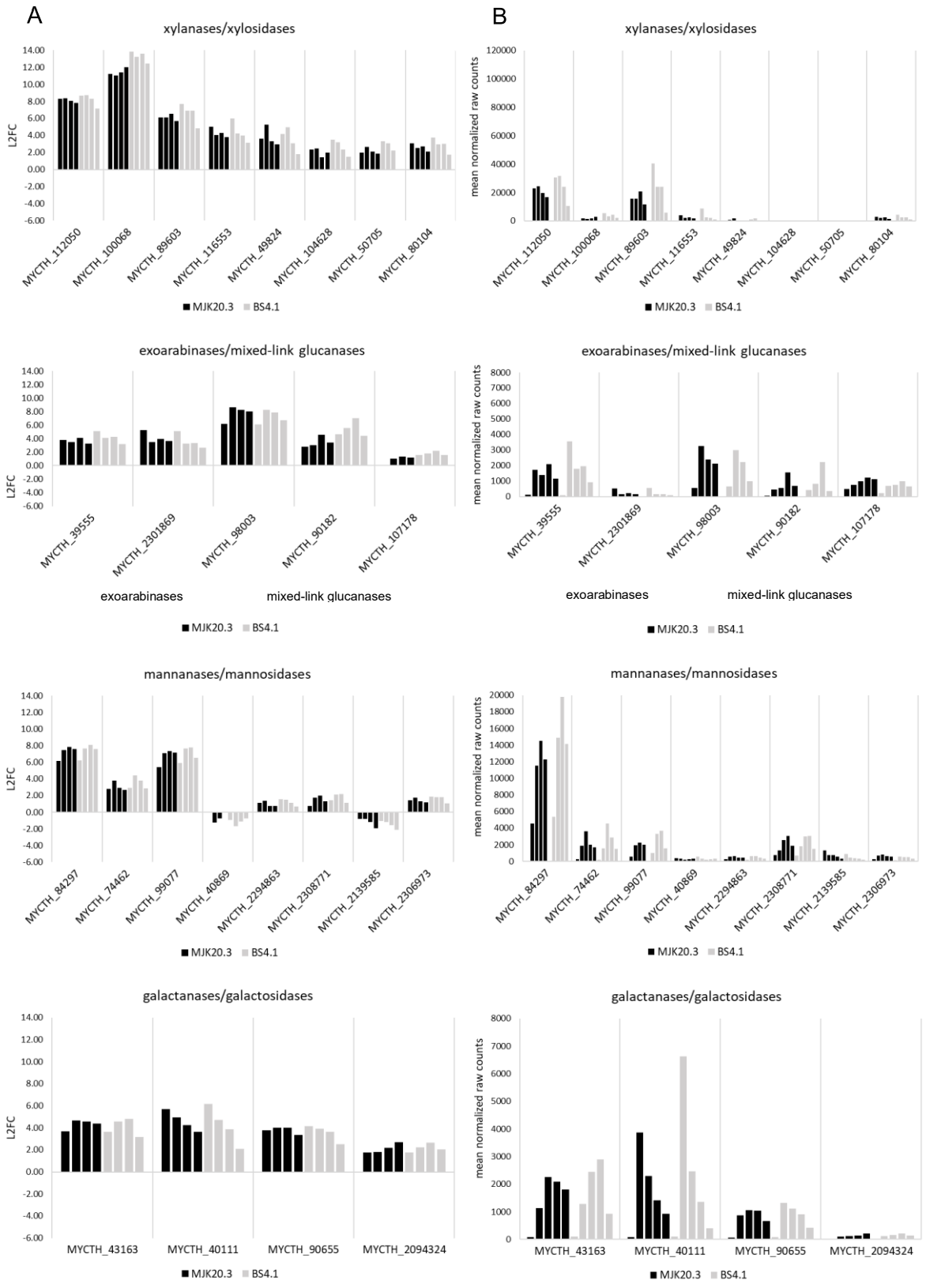

Figure 5.30: Differentially expressed predicted hemicellulase genes in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

The expression of predicted hemicellulase genes (Figure 5.30) is also very similar between MJK20.3 and BS4.1 with most of the genes having higher expression levels in BS4.1 at points in time t1-t3 after the spike as already observed for the predicted cellulase genes . The highest expression levels can be found for the predicted xylanase/xylosidase genes, followed by the predicted mannanase/mannosidase, and the predicted galactanase/galactosidase genes. The top genes of each category are: MYCTH_89603, and MYCTH_112050 (predicted xylanase/xylosidase genes); MYCTH_39555, MYCTH_98003, MYCTH_107178, and MYCTH_98003 (predicted exoarabinase/mixed-link glucanase genes); MYCTH_84297 (predicted mannanase/mannosidase genes); MYCTH_40111, and MYCTH_43163 (predicted galactanase/galactosidase genes).

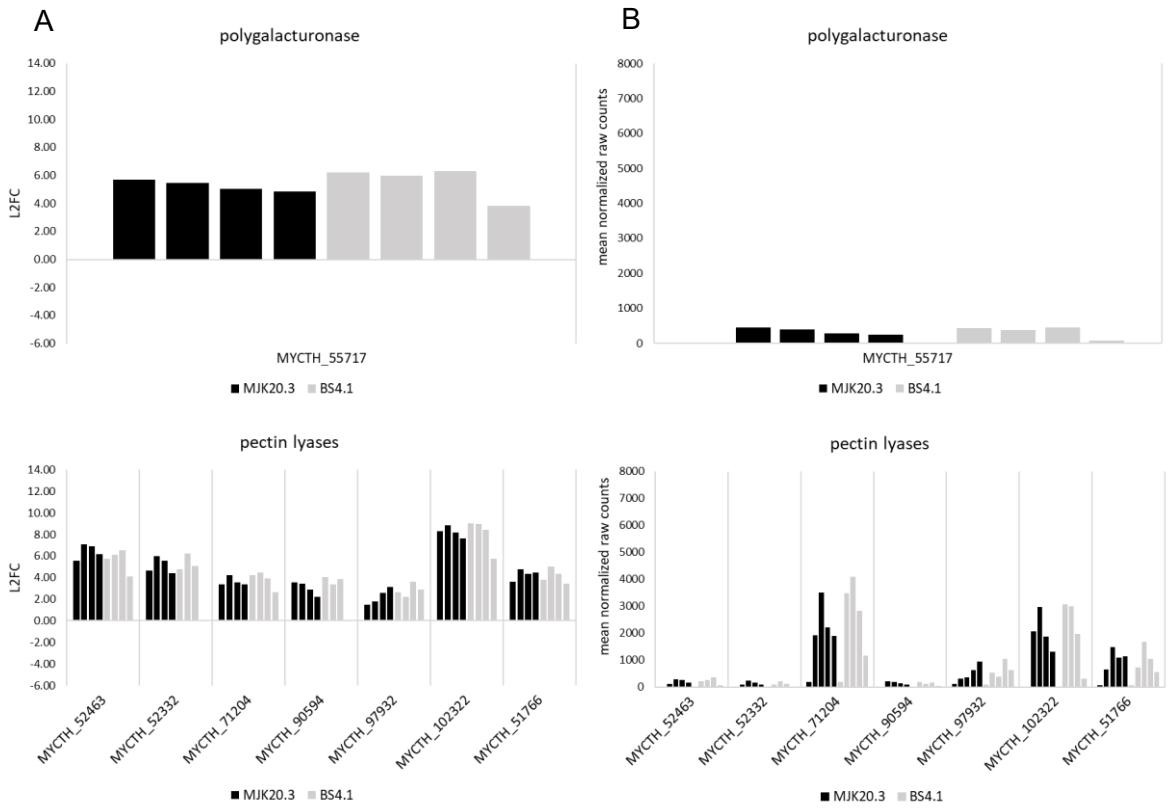

Figure 5.31: Differentially expressed predicted pectinase genes in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

Regarding predicted pectinase genes (Figure 5.31), expression levels of MJK20.3 and BS4.1 are almost identical, with no higher expression levels of BS4.1 for most of the genes except MYCTH_71204 and MYCTH_102322, where expression values in BS4.1 are higher at t1-t3 after the spike. High expression levels can be generally only found for predicted pectin lyase genes, with MYCTH_71204, MYCTH_102322, and MYCTH_51766 as the top genes.

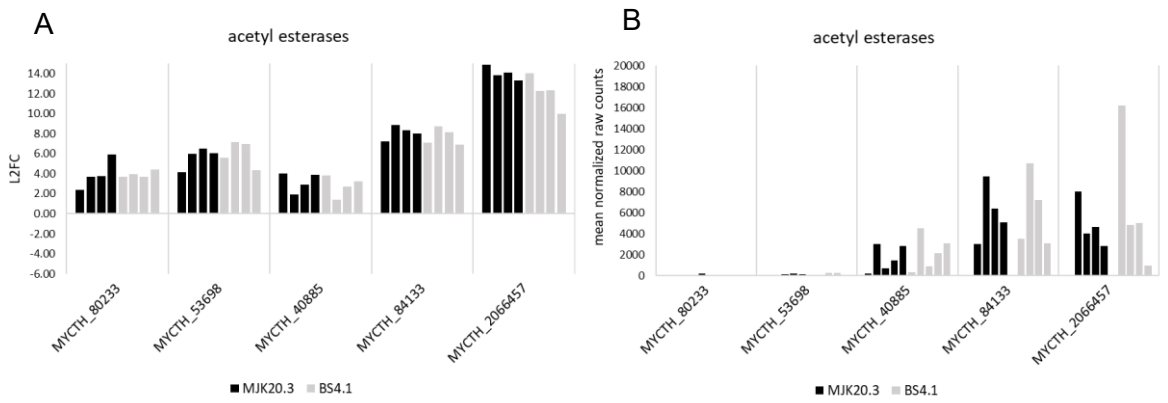

Figure 5.32: Differentially expressed predicted esterase genes in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

Among the predicted esterase genes (Figure 5.32), only acetyl esterase genes are upregulated. Here expression levels of MJK20.3 and BS4.1 are also very high and very similar with a higher expression in BS4.1 but only at early points in time after the spike. Top genes are MYCTH_2066457, MYCTH_84133, and MYCTH_40885.

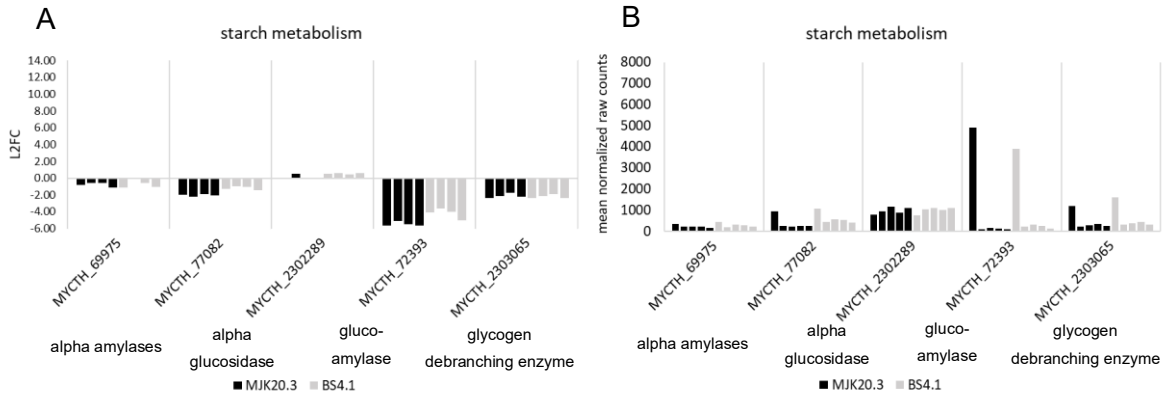

Figure 5.33: Differentially expressed predicted starch metabolism genes in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

Differential expression values for predicted starch metabolism genes (Figure 5.33) are negative for all genes, except for MYCTH_2302289 (predicted alpha glucosidase gene). Generally, no differences between both strains can be detected. The strongest downregulation can be observed for MYCTH_72393 (predicted glucoamylase gene), MYCTH_2303065 (predicted glycogen debranching enzyme gene), and MYCTH_77082 (predicted alpha amylase gene). The general expression levels in this category are low.

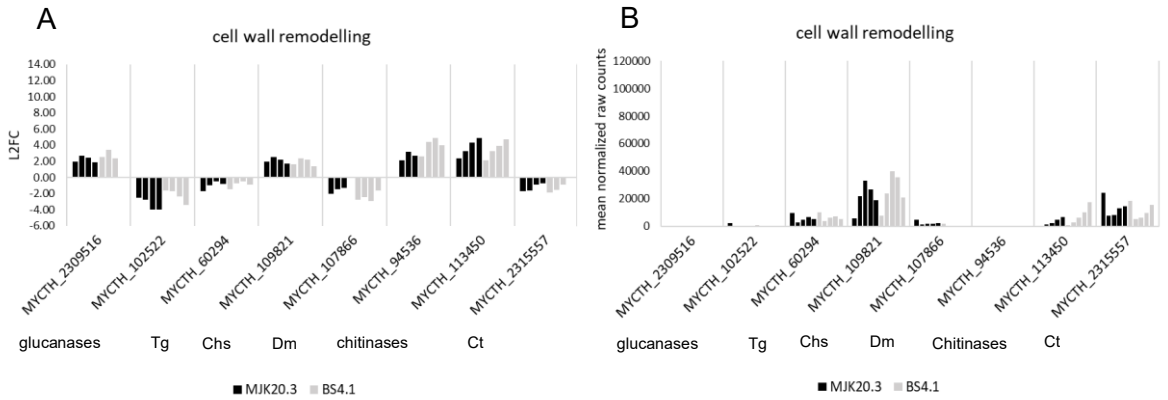

Figure 5.34: Differentially expressed predicted cell wall remodeling genes in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey). Tg= transglucosylase, Chs= chitosanase, Dm= diacetylmuramidase, Ct= crosslinking transglycosidase.

Among the predicted cell wall remodeling genes (Figure 5.34) upregulated as well as downregulated genes can be detected but also no big differences between both strains can be observed, except for MYCTH_113450 (predicted chitinase gene), where expression values are higher at every point in time after the cellulose spike. The top genes that are upregulated and have high expression levels are MYCTH_109821 (predicted chitosanase gene) and MYCTH_113450 (predicted chitinase gene). Genes that are downregulated but nevertheless possess high expression values, are MYCTH_2315557 (predicted crosslinking transglycosidase gene) and MYCTH_60294 (predicted transglucosylase gene). General expression levels in this category are high.

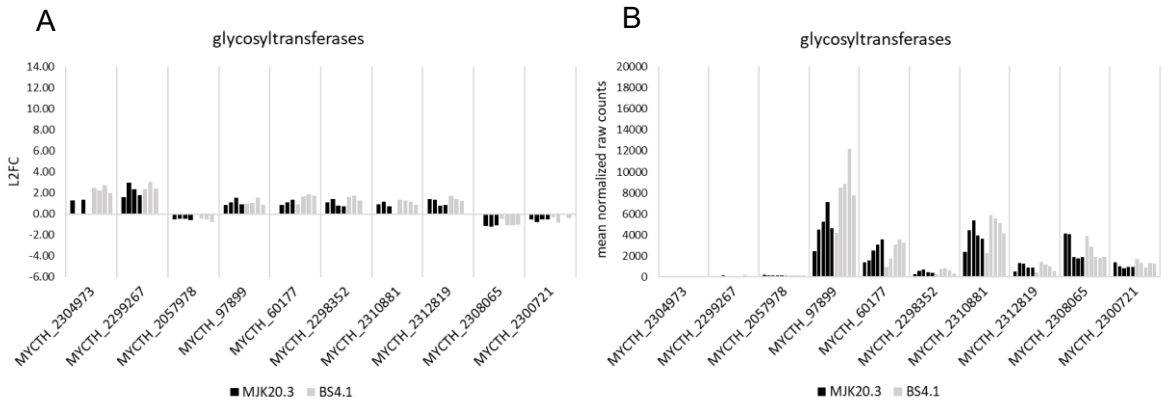

Figure 5.35: Differentially expressed predicted glycosyltransferase genes in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

Expression of predicted glycosyltransferase genes (Figure 5.35) is also not different between both strains except for MYCTH_97899, where expression levels at all points in time after the spike are higher in BS4.1. Most of the genes in this category are upregulated except the genes MYCTH_2057978,

MYCTH_2308065, and MYCTH_2300721. The top genes that are upregulated and have high expression levels are MYCTH_97899, MYCTH_2310881, and MYCTH_60177. Genes that are downregulated but nevertheless possess high expression values, are MYCTH_2308065 and MYCTH_60294.

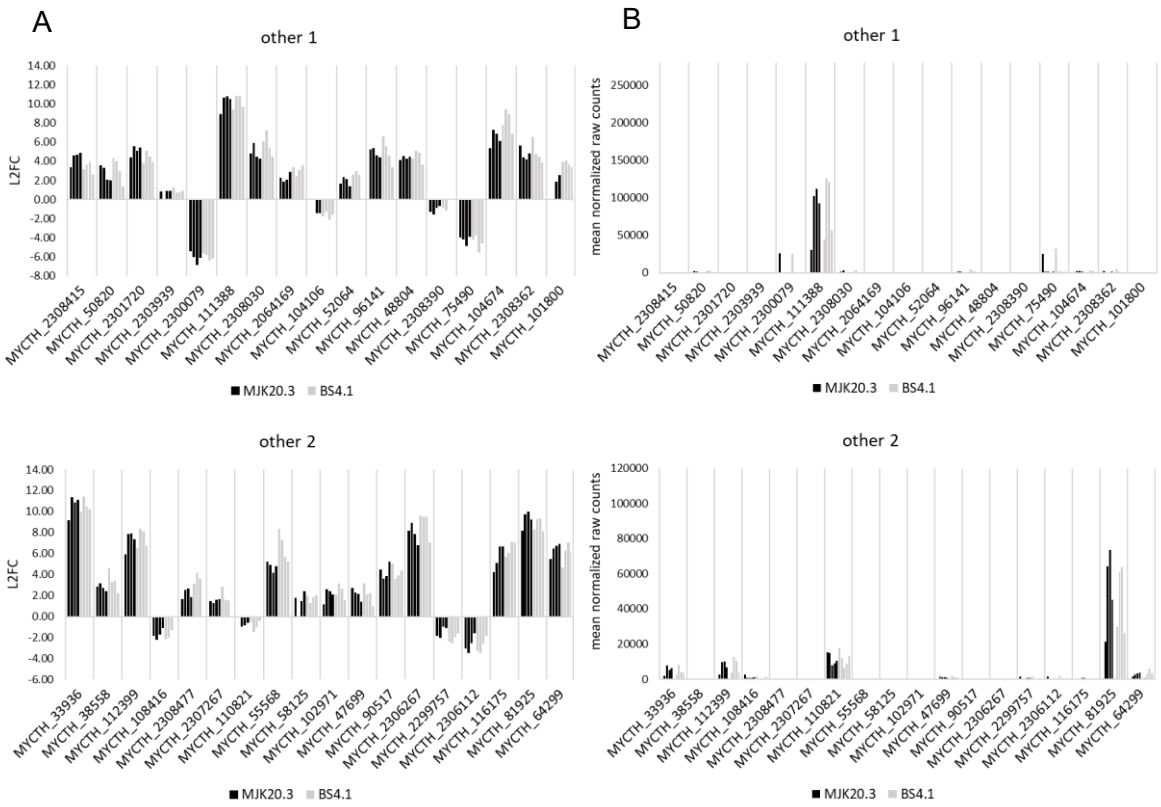

Figure 5.36: Differentially expressed "other" genes in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

In the "other" category (Figure 5.36), some interesting, upregulated genes with an extremely high expression level and some extremely downregulated genes can be identified. Again, a similar expression profile is observed for MJK20.3 and BS4.1. To be able to categorize those genes, that do not have a proper annotation in both literature and databases, information about those genes were retrieved using the publication of Berka et al. 2011 and CAZY database entries (http://www.cazy.org/) to identify the best-known substrate based on the predicted CAZY family the gene belongs to. The findings of this research are summarized in Table 5.10 for all genes that belong to the "other" category. Among the upregulated genes with an extremely high expression level are MYCTH_111388 and MYCTH_81925 (predicted cellobiose dehydrogenase genes) as well as MYCTH_112399 and MYCTH_33936 (predicted CBM1 genes). All these genes do have cellulose as their best-known substrate. Genes that are extremely downregulated are MYCTH_75490 (predicted trehalase gene) and MYCTH_2300079 (predicted GH31 gene). The best-known substrates for those genes are trehalose

and hemicellulose, respectively. One gene that is downregulated but still possesses high expression levels is MYCTH_110821 (hypothetical), the substrate is not known.

Table 5.10: Genes belonging to the “other” class with description of its protein product and best-known substrate. Information about description and best-known substrate was taken from the CAZY database and Berka et al. 2011. GH= glycoside hydrolase, CbD= cellobiose dehydrogenase, CBM= carbohydrate binding module, CE= carbohydrate esterase, GT= glycosyltransferase.

gene	description	best-known substrate
MYCTH_101354	CBM 1 protein	cellulose
MYCTH_101800	GH 30 protein	beta-glycans
MYCTH_102971	oxidase-like protein	hemicellulose
MYCTH_104106	GH 3 protein	beta-glycans
MYCTH_104674	GH 31 protein	hemicellulose
MYCTH_108416	GH 13 - GT 5 protein	alpha-glucan
MYCTH_110061	GH 16 protein	beta-glycans
MYCTH_110821	hypothetical protein	unknown
MYCTH_111388	CbD family protein	cellulose
MYCTH_112399	CBM 1 protein	cellulose
MYCTH_116175	GH 61 protein	cellulose
MYCTH_2064169	GH 43 protein	pectin, hemicellulose
MYCTH_2122625	GH 13 protein	alpha-glucan
MYCTH_2131308	GH 55 protein	beta-1,3-glucan
MYCTH_2294074	GH 79 protein	pectin
MYCTH_2294852	hypothetical protein	unknown
MYCTH_2294895	glyoxal oxidase like protein	unknown
MYCTH_2295307	hypothetical protein	unknown
MYCTH_2299757	hypothetical protein	chitin
MYCTH_2300079	GH 31 protein	hemicellulose
MYCTH_2301720	GH30 protein	beta-glycans
MYCTH_2302517	trehalase	trehalose
MYCTH_2303375	deacetylase	hemicellulose, chitin, peptidoglycan
MYCTH_2303939	GH16 protein	beta-glycans
MYCTH_2304239	CE 4 protein	hemicellulose, chitin, peptidoglycan
MYCTH_2305977	deacetylase	hemicellulose, chitin, peptidoglycan
MYCTH_2306112	hypothetical protein	cellulose
MYCTH_2306267	hypothetical protein	cellulose
MYCTH_2306748	hypothetical protein	unknown
MYCTH_2307267	hypothetical protein	cellulose
MYCTH_2307354	CBM48 protein	glycogen binding
MYCTH_2307864	hypothetical protein	pectin
MYCTH_2308030	GH 94 protein	cellulose
MYCTH_2308362	GH 115 protein	hemicellulose
MYCTH_2308390	GH 114 protein	chitin
MYCTH_2308415	GH79 protein	pectin
MYCTH_2308477	hypothetical protein	unknown
MYCTH_2309011	hypothetical protein	cellulose
MYCTH_2311126	hypothetical protein	unknown
MYCTH_2312626	hypothetical protein	beta-glycans
MYCTH_2312852	hypothetical protein	chitin
MYCTH_2313229	hypothetical protein	starch
MYCTH_2314413	related to plant expansins	cellulose/hemicellulose
MYCTH_33936	CBM 1 protein	cellulose
MYCTH_37110	SUN family-like protein	beta-glycans
MYCTH_37570	GH 16 protein	beta-glycans
MYCTH_37844	CBM 18 protein	chitin
MYCTH_38558	GH 30 protein	beta-glycans
MYCTH_47699	hypothetical protein	pectin
MYCTH_48804	GH 16 protein	beta-glycans
MYCTH_50820	GH43 protein	pectin, hemicellulose
MYCTH_52064	GH 31 protein	hemicellulose

MYCTH_55568	methylesterase	hemicellulose
MYCTH_58125	cbD	cellulose
MYCTH_64299	lipase, GDSL-like protein	unknow
MYCTH_75490	trehalase	trehalose
MYCTH_78090	ferulic acid esterase	hemicellulose
MYCTH_81925	cbD	cellulose
MYCTH_83041	GH 76 protein	hemicellulose
MYCTH_90517	cbD like protein	cellulose
MYCTH_96141	GH 67 protein	hemicellulose
MYCTH_99853	laccase	lignin

In summary, especially predicted cellulase, hemicellulase, pectinase, and esterase genes, as well as single genes belonging to the “others” category that are related to cellulose as substrate are upregulated. Expression levels of those genes are high or even very high. Most of the genes belonging to those categories are upregulated and only a few of them are downregulated. Major differences between MJK20.3 and BS4.1 are higher expression values at especially t1-t3 after the cellulose spike in BS4.1 mainly for predicted cellulase and hemicellulase genes.

5.3.1.4 Differential expression of transcription factors of plant biomass degradation

Knowledge about the expression of inducing or repressing regulators under various industrial conditions, as in the bioreactor experiments of this work, enables direct targeting of genes for deletion or overexpression that are upregulated or downregulated in the respective conditions. The deletion or overexpression of those enzymes, as already mentioned in the introduction, can lead to higher yields of secreted proteins of interest via reducing the expression of, e.g. lignocellulolytic enzymes and therefore, reducing the secretory burden and facilitating downstream processing. On the other hand, it is also possible to examine the conditions under which certain regulators are upregulated/downregulated to improve production conditions of enzymes that are induced or repressed by this regulator for lignocellulosic biorefinery applications. Detailed knowledge about regulation processes also enables the fine tuning of the composition of enzymes cocktails used in plant biomass degradation processes. The annotations of predicted transcription factor genes involved in plant biomass degradation of *T. thermophilus* were retrieved by combining information from the JGI database (https://jgi.doe.gov/), NCBI (https://www.ncbi.nlm.nih.gov/), and the publications mentioned in the respective chapter of the introduction. Orthologs of regulators that are not yet investigated in *T. thermophilus*, were determined via usage of the NCBI protein BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). In Table 5.11 the number and classes of predicted *T. thermophilus* transcription factor genes as they can be found in the JGI database, are shown. Note that several genes were assigned to multiple classes due to the presence of multiple domains.

Table 5.11: Predicted transcription factor genes in *T. thermophilus.*

class	genes
fungal Zn(2)-Cys(6) binuclear cluster domain	101
zinc finger, C2H2 type	53
fungal specific transcription factor domain	61
AT hook motif	18
histone-like transcription factor (CBF/NF-Y)	7
HMG (high mobility group) box	10
GATA zinc finger	6
Myb-like DNA-binding domain	17
homeodomain	5
helix-loop-helix DNA-binding domain	13
bZIP transcription factor	8
basic region leucine zipper	4
MIZ/SP-RING zinc finger	3
NDT80/ PhoG like DNA-binding family	3
copper fist DNA binding domain	2
other	46
total	357

The total number of predicted transcription factor genes that were differentially expressed in the two strains MJK20.3 and BS4.1 at the different points in time after the cellulose spike, is shown in Figure 5.37. Here, a higher number of upregulated genes in BS4.1 for all points in time can be detected. The number of downregulated genes is almost identical between both strains. The strongest differential expression can be detected at t2 in both strains. At t4 a very low number of genes is differentially

expressed. In total approximately 50 % of all predicted transcription factor genes are differentially expressed in BS4.1 and 40 % in MJK20.3.

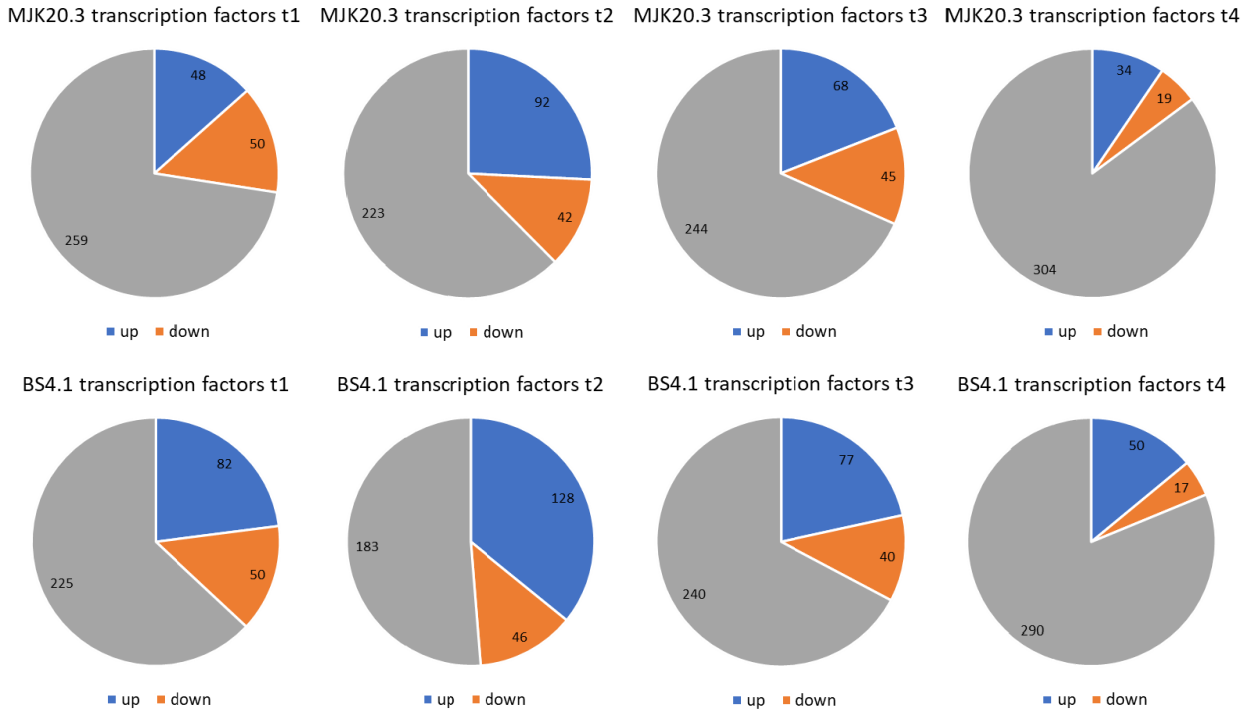

Figure 5.37: Number of differentially expressed predicted transcription factor genes of strains MJK20.3 and BS4.1 at different points in time after the cellulose spike. Shown are the numbers of up- (blue) and downregulated (orange) genes as well as genes with no differential expression (grey) at: 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after cellulose spike (t4) compared to the respective steady state condition.

A detailed overview of the expression profile of the predicted transcription factor genes of the two strains is shown in a heatmap (Figure 5.38). The expression profile of both strains is very similar but a stronger up- and downregulation of single genes (especially upregulation) is observable for BS4.1 as well as a more consistent up- and downregulation over all points in time. A higher total number of differentially expressed genes encoding for predicted transcription factors at the respective points in time as already seen in Figure 5.37 is also detectable. At t4 a lower number of predicted transcription factor genes is differentially expressed in both strains compared to the earlier points in time. Here, almost no differences between both strains are detectable. The classes with the genes that show the strongest upregulation are: fungal Zn(2)-Cys(6) binuclear cluster domain, zinc finger, C2H2 type, fungal specific transcription factor, and bZIP transcription factor. The genes that are showing the strongest downregulation can also be found in those classes but generally, a much higher number of transcription factors are upregulated compared to the downregulated ones.

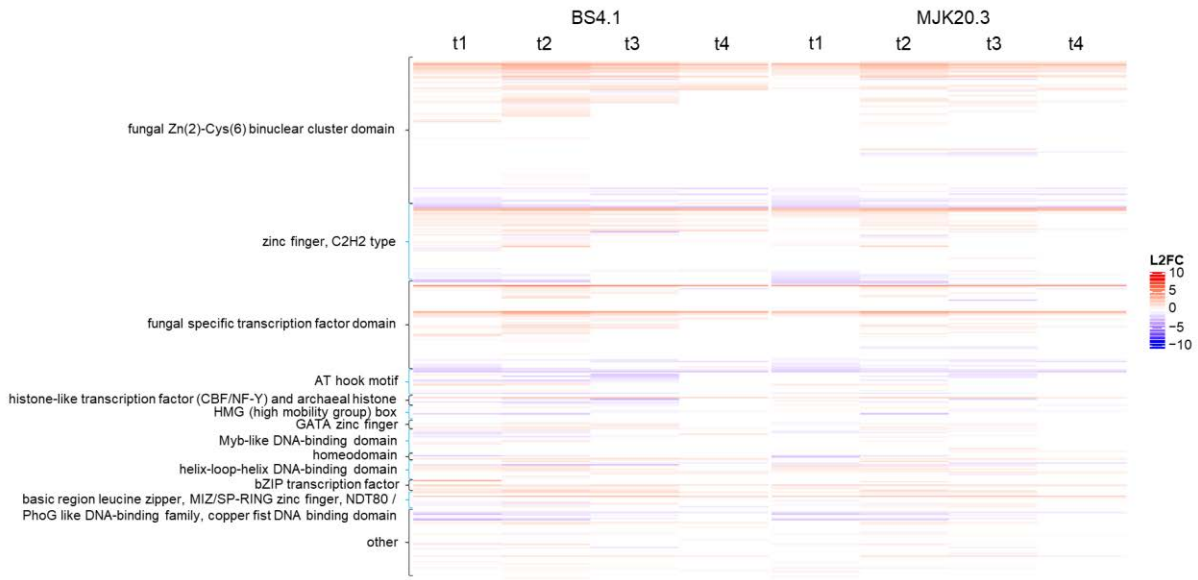

Figure 5.38: Heatmap with differential expression values of all predicted transcription factor genes in strains MJK20.3 and BS4.1. Shown are the log2 fold change values (L2FC) of single predicted transcription factor genes belonging to the different classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

To narrow down the number of differentially expressed genes and filter out the most important differentially expressed genes encoding for predicted transcription factors, the t1-t4 intersection was analyzed next. A gene was included, when differential expression over all points in time after spike in at least one of the two strains was detected. In Table 5.12, the numbers of genes that are differentially expressed at all points in time after the spike are shown for the two strains. A further heatmap (Figure 5.39) visualizes the differential expression levels of the respective genes that belong to the t1-t4 intersection.

Table 5.12: Number of genes belonging to different transcription factor classes that are differentially expressed at all points in time after the spike in strains MJK20.3 and BS4.1.

		MJK20.3		**BS4.1**	
class	genes	t1-t4 up	t1-4 down	t1-t4 up	t1-4 down
fungal Zn(2)-Cys(6) binuclear cluster domain	101	8	4	11	4
zinc finger, C2H2 type	53	3	1	10	0
fungal specific transcription factor domain	61	4	3	4	2
AT hook motif	18	1	0	1	0
histone-like transcription factor (CBF/NF-Y) and archaeal histone	7	0	0	0	0
HMG (high mobility group) box	10	0	0	0	0
GATA zinc finger	6	0	0	1	0
Myb-like DNA-binding domain	17	0	0	0	0
homeodomain	5	0	0	1	0
helix-loop-helix DNA-binding domain	13	1	0	0	0
bZIP transcription factor	8	2	0	2	0
basic region leucine zipper	4	1	0	1	0
MIZ/SP-RING zinc finger	3	0	0	0	0
NDT80 / PhoG like DNA-binding family	3	0	0	0	0
copper fist DNA binding domain	2	0	1	0	1
other	46	0	0	2	0
total	357	20	9	33	7

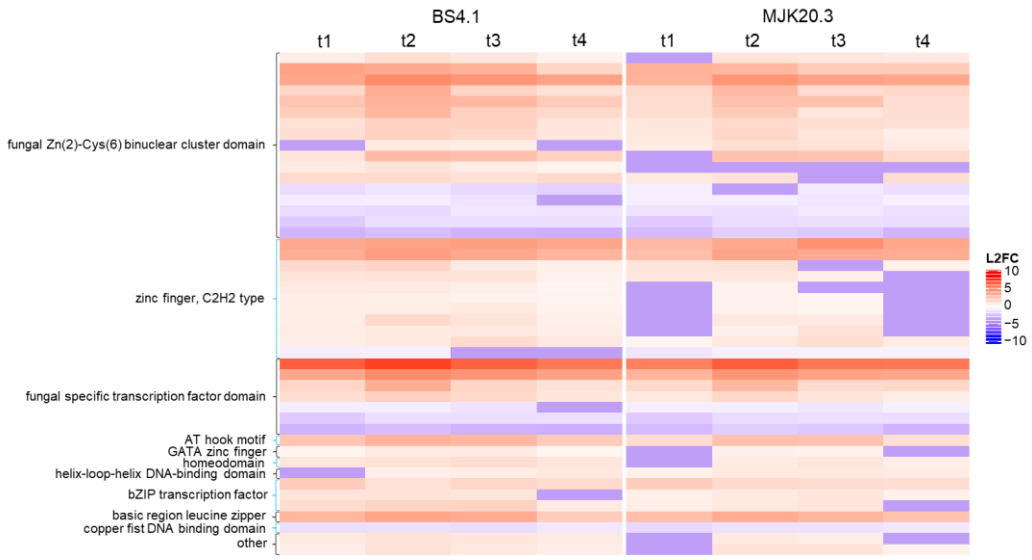

Figure 5.39: Heatmap with expression values of predicted transcription factor genes that are differentially expressed at all points in time after the cellulose spike in strains MJK20.3 and BS4.1. Shown are the log2 fold change values (L2FC) of single predicted transcription factor genes belonging to different classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

The heatmap of the genes that belong to the t1-t4 intersection confirms the results of the previous graphs: the expression pattern between both strains is very similar with some predicted transcription factor genes having a stronger up- and downregulation in BS4.1. Generally, it becomes clearer, that the most and strongest differential expression derives from genes belonging to the classes: fungal Zn(2)-Cys(6) binuclear cluster domain, zinc finger, C2H2 type, fungal specific transcription factor, and bZIP transcription factor. Furthermore, it can be observed, that some predicted transcription factor genes that are upregulated at t1 and t4 in BS4.1 are downregulated in MJK20.3.

To get an idea of the possible functions of these transcription factors, the protein sequence of the predicted transcription factors that belong to the t1-t4 intersection was blasted using all entries of the NCBI database. The descriptions of the blast hits with the lowest E-value or the names of the respective regulator orthologs as well as the general expression trends of the genes are provided in Table 5.13 and are used in the following to describe the genes.

Table 5.13: Up- and downregulated predicted transcription factor genes that belong to the t1-t4 intersection in strains MJK20.3 and BS4.1 with the description of the best protein blast hits (lowest E-values), performed using the NCBI database.

upregulated	downregulated	description best blast hits
MYCTH_105777		transcription factor III A
MYCTH_113457		AreA/Nir2/Nit2
MYCTH_2121737		sterol regulatory Cys6/Ara1
MYCTH_2144297		Xpp1
MYCTH_2295908		Pro1
MYCTH_2296022		halotolerance protein 9
MYCTH_2296328		Lim/homeobox Lhx3/homeobox Pah3/homeobox Hat1
MYCTH_2297012		all development altered
MYCTH_2298160		fluconazole resistance protein 3
MYCTH_2299531		purine utilization
MYCTH_2300377		Gli1/Crz2
MYCTH_2300984		glucose transport rgt1/nitrate assimilation NirA

MYCTH_2302011		sterol uptake control protein 2/oxidase assembly protein 2
MYCTH_2303382		respiration factor 2/ Adr1
MYCTH_2307033		regulator adr1/putative respiration factor 2
MYCTH_2307450		maltose acetyltransferase/female fertility 7/nodulation protein
MYCTH_2307545		krueppel like
MYCTH_2308084		sterigmatocystin biosynthesis/Bea4
MYCTH_2308823		Moc3/OefC/Pro1A
MYCTH_2309600		Hcr1
MYCTH_2310085		Cre1
MYCTH_2310145		Xyr1
MYCTH_2310586		hypothetical
MYCTH_2310995		HacA/1
MYCTH_2312442		krueppel like factor 4/Sp5
MYCTH_38704		Clr2
MYCTH_46266		GaaR/Pdr2
MYCTH_49142		Vad1
MYCTH_59287		Crz1
MYCTH_63778		Sp3
	MYCTH_115802	SteA/Fst12/Ste12
	MYCTH_2295212	nit4/lipase regulator1/cutinase transcription factor 1
	MYCTH_2300024	ArcA like
	MYCTH_2300537	Acu15
	MYCTH_2301920	AmyR/BglR
	MYCTH_2305551	pyrimidine pathway regulatory 1/Ada6/ CtnR
	MYCTH_2307010	copper resistance protein Crf1/Cup2

To investigate and compare gene expression in detail, graphs showing L2FC values as well as graphs showing the mean of the normalized raw count values of the two replicates of the single genes at every point in time after the spike, were created as previously (see chapter 5.3.1.2) described (Figures 5.40-5.43).

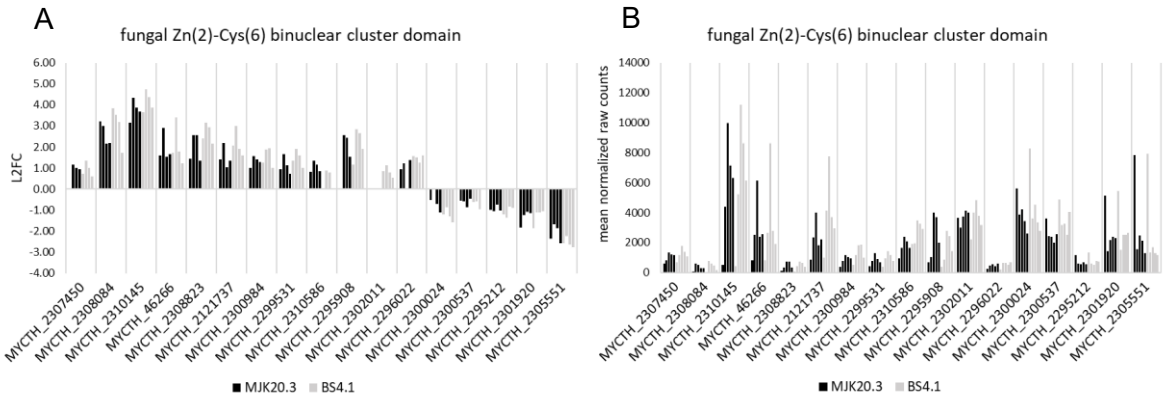

Figure 5.40: Differentially expressed genes encoding for predicted transcription factors with fungal Zn(2)-Cys(6) binuclear cluster domain in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

Regarding expression of genes encoding for predicted fungal Zn(2)-Cys(6) binuclear cluster domain transcription factors (Figure 5.40), expression levels and differential expression values are very similar for MJK20.3 and BS4.1 but higher for almost every single gene at every point in time after the cellulose spike in BS4.1. Generally, expression levels in this category are very high. The top genes that are

upregulated and have high expression levels in this category are: MYCTH_2310145 (Xyr1), MYCTH_46266 (GaaR/Pdr2), MYCTH_2121737 (sterol regulatory Cys6/Ara1), MYCTH_2310586 (hypothetical), and MYCTH_2295908 (Pro1). The genes with the strongest downregulation that still possess high expression levels after the spike, are: MYCTH_2305551 (pyrimidine pathway regulatory 1/Ada6/citrinin CtnR), MYCTH_2301920 (AmyR/BglR), MYCTH_2300024 (ArcA like), and MYCTH_2300537 (Acu15).

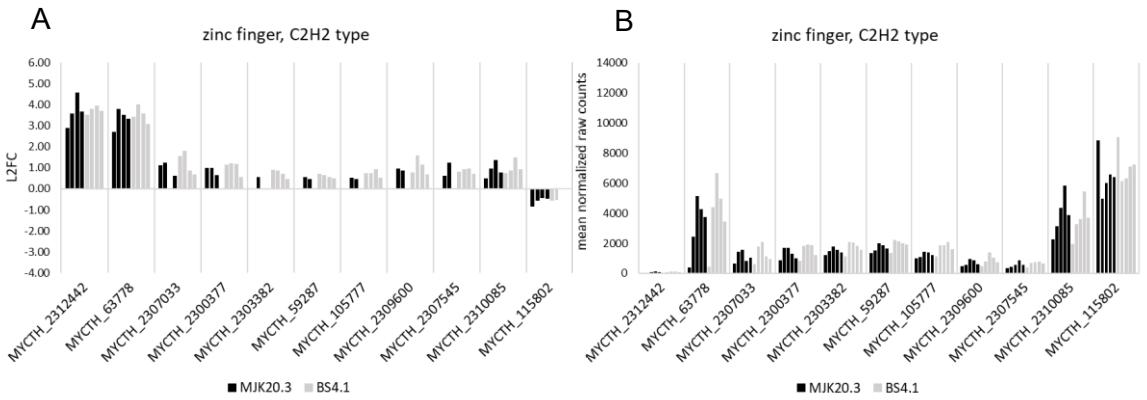

Figure 5.41: Differentially expressed genes encoding for predicted zinc finger C2H2 type class transcription factors in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

The expression of genes encoding for predicted zinc finger C2H2 type transcription factors (Figure 5.41) in BS4.1 is as for the predicted fungal Zn(2)-Cys(6) binuclear cluster domain transcription factor genes very similar for MJK20.3 and BS4.1 and higher for almost every single gene at every point in time after the cellulose spike in BS4.1. General expression levels are very high in this category as well, but not as high as in the previous category. The top genes that are upregulated and have high expression levels in this category are: MYCTH_63778 (Sp3) and MYCTH_2310085 (Cre1). The only downregulated gene that still possess high expression levels after the spike, is MYCTH_115802 (SteA/Fst12/Ste12).

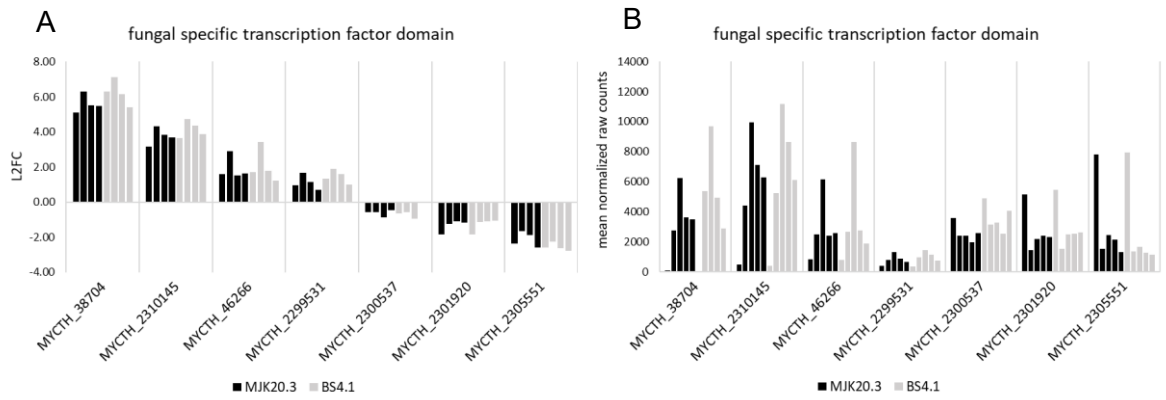

Figure 5.42: Differentially expressed genes encoding for predicted transcription factors with fungal specific transcription factor domain in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey).

Genes encoding for predicted transcription factors with a fungal specific transcription factor domain (Figure 5.40) also have slightly higher differential expression values in BS4.1 for most of the genes, especially at t1-t3 in this category. Expression levels in this category are also very high, with 4 genes upregulated and 3 genes downregulated. The upregulated genes with high expression levels include MYCTH_2310145 (Xyr1), MYCTH_38704 (Clr2), and MYCTH_46266 (GaaR/Pdr2). The downregulated genes include: MYCTH_23005551 (pyrimidine pathway regulatory 1/Ada6/citrinin CtnR), MYCTH_2301920 (AmyR/BglR), and MYCTH_2300537 (Acu15). Of those genes, the only gene that was not already mentioned as a top gene in a previous category is MYCTH_38704 (Clr2).

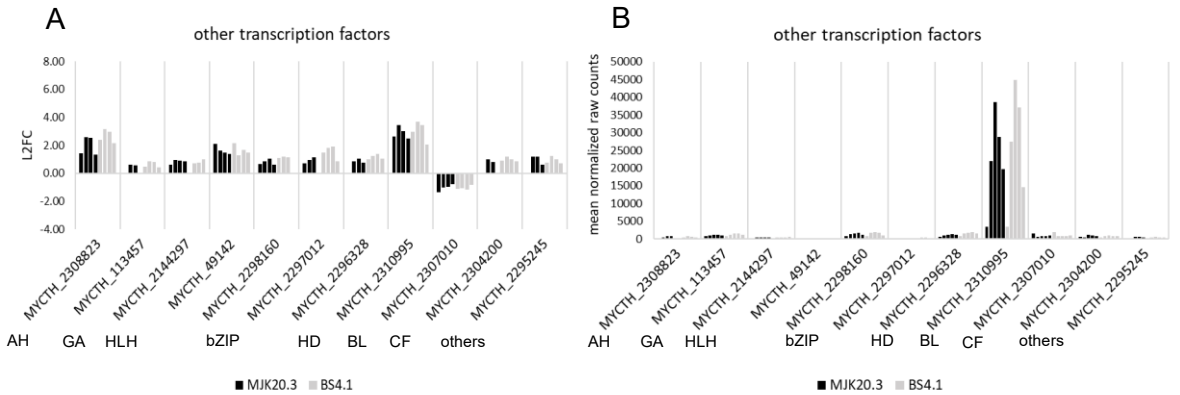

Figure 5.43: Other genes encoding for differentially expressed predicted transcription factors in strains MJK20.3 and BS4.1 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black) and BS4.1 (grey). AH= AT hook motif, GA= GATA zinc finger, HLH= helix loop helix domain, bZIP= bZIP transcription factor, HD= homeodomain, BL= basic region leucine zipper, CF= Copper fist DNA binding domain.

All other predicted transcription factor genes that belong to the t1-t4 intersection are summarized in Figure 5.43. Here, the same trends as in the already discussed categories, regarding the differences between the MJK20.3 and BS4.1 strain, are observable. Expression levels are only for the extremely

upregulated gene MYCTH_2310995 (HacA/1) very high. All other genes of this category are only having low expression levels.

Besides the identification of potentially new regulators according to their expression at all points in time after the cellulose spike, as done above, also the identification of the differential expression of all known regulators of plant biomass degradation, as discussed in the introduction, gives valuable information about regulation processes in the investigated strains. To achieve this, differential expression was investigated for all genes encoding for regulators that possess an ortholog in *T. thermophilus*. Note that several of those regulators have more than one possible ortholog. The L2FC values as well as the mean of the normalized raw counts under the respective conditions are shown in Figure 5.44.

		L2FC								mean normalized raw counts									
		MJK20.3				BS4.1				MJK20.3					BS4.1				
gene	ortholog	t1	t2	t3	t4	t1	t2	t3	t4	SS	t1	t2	t3	t4	SS	t1	t2	t3	t4
MYCTH_2028011	Ace1	0.00	0.00	0.00	0.00	0.89	0.00	0.00	0.00	1572	1910	1889	1831	1446	1574	2925	1832	1671	1306
MYCTH_2308260	Ace2	0.00	0.00	-0.91	0.00	0.96	0.60	-3.17	0.00	378	471	452	201	351	393	762	595	44	269
MYCTH_2301920	AmyR/BglR*	-1.84	-1.23	-1.10	-1.15	-1.85	-1.13	-1.11	-1.05	5142	1435	2186	2405	2310	5460	1516	2496	2535	2629
MYCTH_2307451	Ap3	0.00	0.71	0.00	0.00	0.51	0.56	0.00	0.00	304	386	496	368	334	319	458	473	382	317
MYCTH_113457	AreA/Nir2/Nit2	0.00	0.62	0.58	0.00	0.46	0.83	0.82	0.42	803	958	1227	1201	1068	860	1186	1532	1519	1151
MYCTH_2309867	AreB	-1.08	-0.82	0.00	0.00	-1.25	-1.02	0.00	0.00	527	249	299	519	623	566	240	281	476	727
MYCTH_2063030	ClbR/ClbR2/ClbR3	-0.78	-0.60	-0.68	0.00	-0.76	0.00	0.00	0.00	489	284	322	304	333	511	299	404	419	406
MYCTH_2298863	Clr1*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2270	2072	2425	1929	2031	2597	2790	3014	2402	2138
MYCTH_38704	Clr2*	5.12	6.31	5.52	5.47	6.28	7.13	6.17	5.39	80	2740	6258	3624	3510	69	5371	9693	4954	2895
MYCTH_2306730	ClrC	1.09	0.80	0.00	0.00	1.63	1.35	1.05	1.04	156	332	272	243	238	142	440	361	295	293
MYCTH_2296492	Clr4*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1294	1443	1225	1230	1112	1293	1633	1541	1650	1382
MYCTH_2310085	Cre1*	0.49	0.96	1.37	0.79	0.75	0.88	1.49	0.92	2245	3142	4354	5817	3873	1949	3278	3592	5463	3698
MYCTH_2055311	CreB	0.00	0.47	0.60	0.00	0.57	0.56	0.92	0.00	775	1006	1068	1173	1074	902	1347	1333	1709	1117
MYCTH_2306444	CreC	0.00	0.00	0.00	0.00	0.37	0.51	0.00	0.00	499	457	582	615	491	467	599	665	567	496
MYCTH_2306452	CreD	0.00	-0.53	0.90	0.00	0.00	0.00	1.14	0.00	365	357	252	680	453	369	310	278	810	492
MYCTH_59287	Crz1	0.00	0.57	0.46	0.00	0.71	0.66	0.57	0.49	1361	1504	2018	1874	1633	1353	2205	2142	2003	1902
MYCTH_46266	GaaR/Pdr2	1.60	2.90	1.53	1.65	1.71	3.41	1.77	1.22	826	2506	6143	2391	2582	812	2663	8626	2777	1885
MYCTH_46981	GaaX	1.68	2.07	0.00	0.00	1.85	2.41	0.73	0.00	261	835	1095	344	375	271	976	1437	450	315
MYCTH_2310995	HacA/1	2.64	3.45	3.03	2.49	2.98	3.70	3.42	2.07	3522	21933	38565	28685	19763	3465	27427	44917	37051	14575
MYCTH_2297059	Hap2	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.00	1333	1463	1443	985	1177	1226	1633	1638	1022	1327
MYCTH_41855	Hap3	0.40	0.00	0.00	-0.44	0.00	0.00	-0.49	0.00	1153	1522	1186	977	851	1179	1170	930	841	1163
MYCTH_67051	Hap5	0.00	0.00	-0.31	0.00	0.00	0.00	0.00	0.00	1365	1347	1207	1101	1376	1595	1646	1435	1357	1517
MYCTH_2309600	Hcr1	0.00	0.98	0.87	0.00	0.76	1.59	1.16	0.67	476	557	932	866	620	466	788	1407	1039	742
MYCTH_2295635	Hep1/HP1	-1.68	0.00	0.00	0.00	-1.94	-1.02	0.00	0.00	94	29	55	112	100	79	19	39	98	119
MYCTH_2308921	Ire1	0.00	0.59	0.00	0.00	0.75	0.70	0.00	0.00	1517	1619	2275	1736	1989	1472	2483	2385	1995	1910
MYCTH_2294559	Lae1/A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	213	296	261	199	181	749	1577	618	512	638
MYCTH_2303067	MalR	0.00	0.00	0.84	0.00	0.00	0.55	0.56	0.00	191	166	231	342	200	218	223	318	321	186
MYCTH_2132441	McmA/1	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00	4639	4910	6047	5654	4900	5675	5077	6808	6840	6211
MYCTH_2303918	Mhr1	0.00	0.00	0.00	0.00	0.00	0.57	0.58	0.00	319	338	378	355	326	364	463	538	543	490
MYCTH_2298994	NirA/Nit4	0.00	0.00	0.00	0.00	0.00	0.76	0.00	0.00	386	367	471	453	454	325	428	549	449	418
MYCTH_2302460	NmrA/1	0.00	0.36	0.83	1.00	0.52	1.01	1.09	0.69	598	573	770	1066	1199	618	892	1246	1317	1000
MYCTH_81165	PacC/1	0.00	1.40	0.00	0.00	1.28	1.86	0.66	0.00	1211	1652	3194	1664	1340	1157	2812	4191	1828	1262
MYCTH_2303559	Prk6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	192	202	202	182	182	216	231	269	244	190
MYCTH_2300719	Rca1	0.00	0.00	0.00	0.00	-0.54	0.00	0.00	0.00	1650	1455	1636	1607	1518	1645	1131	1217	1489	1610
MYCTH_2298696	Rce1	0.00	0.00	0.00	0.00	0.00	0.45	0.00	0.00	1025	970	1228	944	781	987	1082	1343	783	898
MYCTH_2302052	Res1	0.00	0.00	0.72	0.00	0.00	0.00	0.00	0.00	916	886	1096	1511	1312	899	1007	969	1188	1088
MYCTH_53224	RhaR/Pdr1	0.00	0.00	0.00	0.00	0.00	0.74	0.00	0.00	225	223	310	243	282	214	261	356	225	202
MYCTH_2306768	Sah2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	730	620	925	827	800	730	844	1011	832	710
MYCTH_2297068	Stk12	0.00	0.66	0.00	0.00	1.00	0.86	1.04	0.00	144	191	227	219	163	141	282	255	290	222
MYCTH_2312657	VeA/Vel1	0.00	0.82	0.63	0.52	0.00	0.82	0.58	0.53	828	924	1458	1280	1187	960	1193	1692	1438	1383
MYCTH_113912	VelB	0.00	0.87	0.75	0.00	0.00	0.83	0.97	0.00	343	440	625	574	449	347	478	616	679	541
MYCTH_46530	Vib1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	350	359	355	365	431	373	477	464	449	331
MYCTH_108157	Vib2	0.00	0.44	0.67	0.00	0.00	0.70	0.00	-0.84	825	707	1119	1313	810	63	51	103	73	35
MYCTH_2309330	Wc1/Blr1	0.00	0.57	0.00	0.00	0.72	0.90	0.70	0.00	588	687	871	751	571	507	833	943	821	515
MYCTH_2294022	Wc2/Blr2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	556	438	485	550	376	518	569	629	627	341
MYCTH_2310145	Xyr1*	3.15	4.33	3.85	3.67	3.64	4.74	4.37	3.87	494	4394	9959	7135	6289	418	5225	11178	8634	6122
MYCTH_2144297	Xpp1	0.63	0.93	0.92	0.83	0.00	0.72	0.76	0.97	263	406	499	497	468	299	388	494	505	589

Figure 5.44: Differential expression of genes encoding for orthologs of known regulators of plant biomass degradation in *T. thermophilus* strains MJK20.3 and BS4.1. Shown are the log2 fold change values (L2FC) at: 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after cellulose spike (t4) compared to the respective steady state condition as well as the mean of the normalized raw counts of the respective condition. Green color indicates an upregulation and red color indicates a downregulation. The dash separates the possible orthologs of this regulator if more than one was found. An asterisk marks the regulators that have already been investigated in *T. thermophilus*.

Generally, expression patterns between BS4.1 and MJK20.3 are also for genes encoding for the orthologs of known regulators very similar. For many of these genes, either the L2FC value or the normalized raw count values are low. The genes encoding for the orthologs to the known regulators that are upregulated in at least two points in time after the cellulose spike and have medium to high

expression levels in both strains are: MYCTH_113457 (AreA/Nir2/Nit2), MYCTH_38704 (Clr2), MYCTH_2310085 (Cre1), MYCTH_2055311 (CreB), MYCTH_59287 (Crz1), MYCTH_46266 (GaaR/Pdr2), MYCTH_46981 (GaaX), MYCTH_2310995 (HacA/1), MYCTH_2308921 (Ire1), MYCTH_2302460 (NmrA/1), MYCTH_81165 (PacC/1), MYCTH_2312657 (VeA/Vel1), MYCTH_108157 (Vib2), MYCTH_2309330 (Wc1/Blr1), MYCTH_2310145 (Xyr1), and MYCTH_2144297 (Xpp1). Out of those the top 5 with the strongest expression are: MYCTH_38704 (Clr2), MYCTH_2310995 (HacA/1), MYCTH_2310145 (Xyr1), MYCTH_46266 (GaaR/Pdr2), and MYCTH_2310085 (Cre1). For all these genes, differential expression values are higher in BS4.1 at t1-t3 and lower for most of the genes at t4 compared to MJK20.3.

Genes encoding for orthologs of known regulators that are downregulated in both strains and have medium to high expression levels only include MYCTH_2301920 (AmyR/BglR). For this gene raw count values at all conditions are constantly slightly higher in BS4.1 and therefore, L2FC values are almost identical.

Genes encoding for orthologs with medium to high expression levels that show no differential expression or only at one point in time after the cellulose spike include: MYCTH_2028011 (Ace1), MYCTH_2298863 (Clr1), MYCTH_2296492 (Clr4), MYCTH_2297059 (Hap2), MYCTH_41855 (Hap3), MYCTH_67051 (Hap5), MYCTH_2132441 (McmA/1), MYCTH_2300719 (Rca1), MYCTH_2298696 (Rce1), MYCTH_2302052 (Res1), and MYCTH_2306768 (Sah2). Out of those the ones with the highest raw count numbers are: MYCTH_2132441 (McmA/1) and MYCTH_2298863 (Clr1). For most of these genes, differential expression values are higher in BS4.1 especially at t1-t3.

In summary, major differences between MJK20.3 and BS4.1 are higher expression values at especially t1-t3 after the cellulose spike in BS4.1 of especially highly expressed predicted transcription factor genes and genes encoding for orthologs of regulators of plant biomass degradation. Most notable differences here are the expression of MYCTH_63778 (Sp3), MYCTH_2121737 (sterol regulatory cys6/ara1), MYCTH_38704 (Clr2), MYCTH_2310995 (HacA/1), MYCTH_46266 (GaaR/Pdr2), and MYCTH_2310145 (Xyr1).

5.3.2 Cellulase regulator deletion mutants BS6.4 (Δ*ku80*, Δ*clr2*), BS7.8 (Δ*ku80*, Δ*clr1*), and JK2.8 (Δku80, Δ*clr4*)

5.3.2.1 General and enrichment analysis

For differential gene expression analysis, the samples taken after the cellulose spike were always compared to the respective steady state condition. The number of the resulting differential expressed genes with padj. ≤ 0.05 is shown in Table 5.14. The number of differentially expressed genes is higher at all points in time after the spike in strains BS6.4, BS7.8, and JK2.8 compared to MJK20.3. Toward t4 the number of up- and downregulated genes is drastically reduced in JK2.8. This was the same in MJK20.3. In the other two strains the number of differentially expressed genes is much higher compared to JK2.8 and does not decrease toward t4. The highest numbers of differentially expressed genes can be observed for BS7.8 at all points in time after the spike. The highest number of differentially expressed genes can be observed 1 h after the cellulose spike (t2) for all strains.

Table 5.14: Number of differentially expressed genes with padj. ≤ 0.05 of the strains BS6.4, BS7.8, and JK2.8 at different points in time after the cellulose spike in relation to the respective steady state condition.

strain	condition	upregulated genes	downregulated genes
BS6.4	t1	1981	1936
	t2	2398	2232
	t3	2228	2057
	t4	2012	1939
BS7.8	t1	2311	2063
	t2	2467	2174
	t3	2436	2251
	t4	2285	2238
JK2.8	t1	1734	1673
	t2	2102	2052
	t3	1517	1535
	t4	733	601

With those numbers, Venn diagrams were generated, as previously described (see chapter 5.3.1.1). The diagrams are shown in Figure 5.45. The Venn diagram of JK2.8 is very similar compared to those of MJK20.3 (Figure 5.18). Here, especially at the early points in time more differentially expressed genes can be found in JK2.8 as well as a higher total number of uniquely up- and downregulated genes. The opposite can be observed for BS6.4 and BS7.8: the number of up- and downregulated genes is lower for the earlier points in time and much higher at the later points in time compared to MJK20.3. All in all, in JK2.8 approximately 60 % and in BS6.4 and JK2.8 70 % of all genes (total number of genes: 9292) are differentially expressed compared to roughly 54 % in MJK20.3.

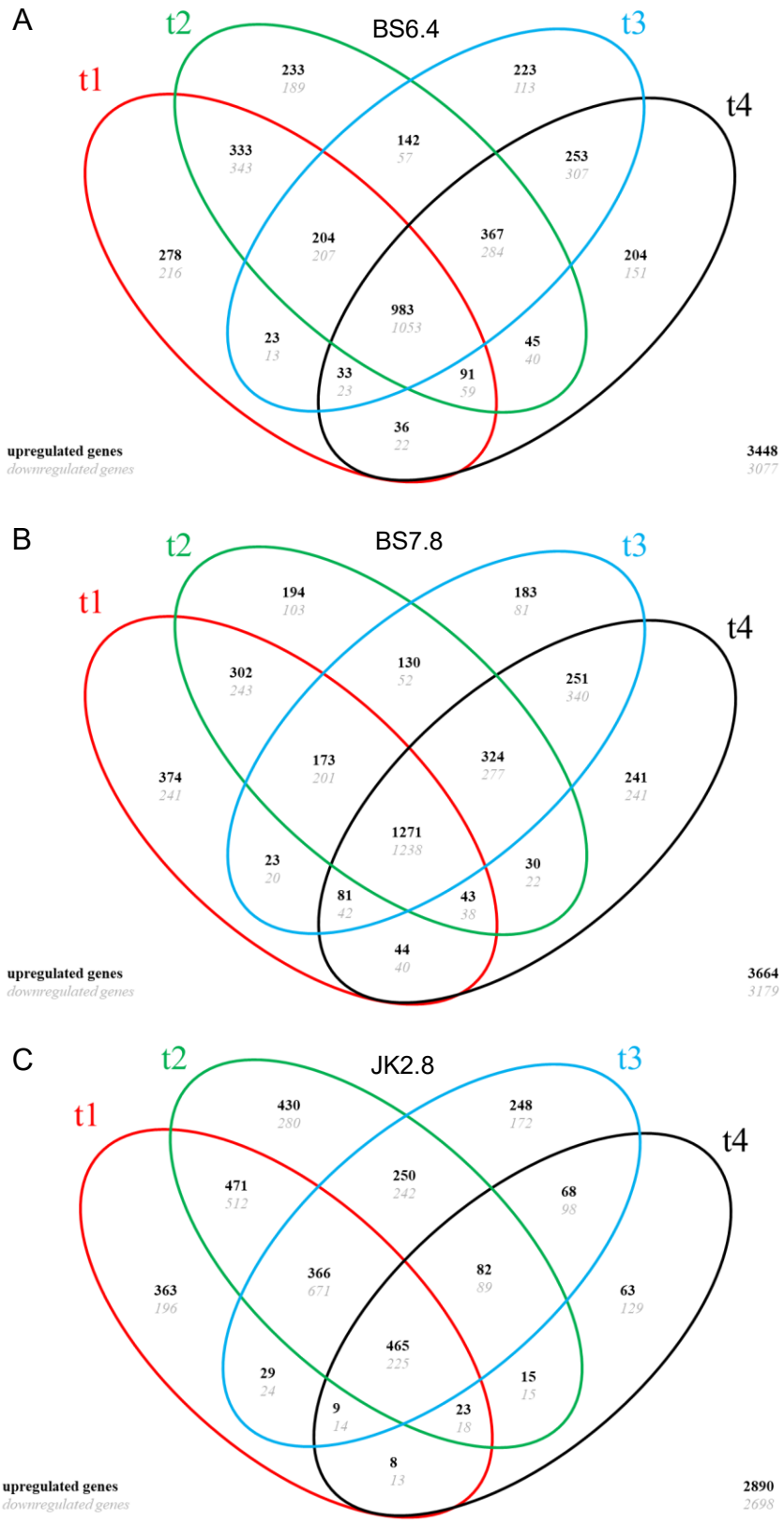

Figure 5.45: Venn diagrams of differentially expressed genes after cellulose spike in relation to the respective steady state condition in strains BS6.4, BS7.8, and JK2.8. Black (grey) numbers refer to upregulated (down-) genes in strain BS6.4 (A), BS7.8 (B), and JK2.8 (C). The total number of uniquely upregulated and downregulated genes is shown in the bottom right corner of (A), (B), and (C). Differential expression (padj. ≤ 0 05) was determined via comparison of expression profiles of t1= 0.5 h, t2= 1 h, t3= 2 h, and t4= 4 h after spiking cellulose with the respective steady state condition.

In many of the following analyses it is referred to the t1-t4 intersection, which was already explained in chapter 5.3.1.1. For JK2.8 higher numbers of up- and downregulated genes in the t1-t4 intersection area (see Figure 5.45) can be detected compared to MJK20.3. The number of genes in the t1-t4 intersection area is for BS6.4 and BS7.8 much higher (~3x) compared to MJK20.3 and JK2.8, whereas in BS7.8 this number is the highest. The relation between up- and downregulated genes in the t1-t4 intersection area is for JK2.8 similar compared to MJK20.3 (up: down= ~2:1). In BS6.4 and BS7.8 this relation amounts approximately 1:1.

To get an idea of the overall transcriptomic response of the strains MJK20.3 and BS4.1 to cellulose, a GO enrichment analysis using the DAVID database (Huang et al. 2009a; Huang et al. 2009b) with the respective gene sets belonging to the t1-t4 intersection was performed as previously described (see chapter 5.3.1.1). The results are shown in Tables 5.15-5.20.

Table 5.15: GO term enrichment analysis of the upregulated genes in the t1-t4 intersection gene set of strain BS6.4. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	carbohydrate metabolic process	3	39
	carbohydrate transport	4	5
	cellulose catabolic process	4	5
	glycogen biosynthetic process	11	3
	late nucleophagy	9	4
	lipid metabolic process	3	7
	mitophagy	7	5
	nitrogen compound metabolic process	5	4
	piecemeal microautophagy of nucleus	4	5
CC	Ada2/Gcn5/Ada3 transcription activator complex	11	4
	extracellular region	3	9
	integral component of membrane	1	189
MF	aldehyde-lyase activity	10	3
	carbohydrate binding	4	8
	heme binding	2	16
	hydrolase activity	2	26
	hydrolase activity, hydrolyzing O-glycosyl compounds	2	18
	ion channel activity	5	4
	metal ion binding	1	34
	oxidoreductase activity	2	36

Table 5.16: GO term enrichment analysis of the downregulated genes in the t1-t4 intersection gene set of strain BS6.4. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	acetyl-CoA biosynthetic process from pyruvate	5	4
	aerobic respiration	2	8
	arginine biosynthetic process	4	5
	ATP synthesis coupled proton transport	4	11
	cellular amino acid biosynthetic process	3	7
	DNA replication initiation	3	12
	DNA strand elongation involved in DNA replication	4	6

BP	formation of translation preinitiation complex	5	16
	lysine biosynthetic process via aminoadipic acid	5	4
	mitochondrial electron transport, ubiquinol to cytochrome c	5	7
	mitochondrial translation	2	7
	pre-replicative complex assembly involved in nuclear cell cycle DNA replication	3	8
	protein import into mitochondrial matrix	3	7
	purine nucleobase biosynthetic process	5	4
	regulation of translational initiation	4	13
	ribosomal large subunit assembly	2	9
	ribosomal small subunit biogenesis	5	6
	rRNA processing	2	12
	transcription of nuclear large rRNA transcript from RNA polymerase I promoter	4	5
	translation	4	81
	UV-damage excision repair	5	6
CC	box C/D snoRNP complex	4	4
	Ctf18 RFC-like complex	4	4
	cytosolic large ribosomal subunit	4	7
	cytosolic small ribosomal subunit	4	7
	DNA replication factor C complex	6	4
	DNA replication preinitiation complex	3	6
	Elg1 RFC-like complex	6	5
	eukaryotic 43S preinitiation complex	6	14
	eukaryotic 48S preinitiation complex	5	14
	eukaryotic translation initiation factor 3 complex	5	6
	eukaryotic translation initiation factor 3 complex, eIF3e	6	6
	eukaryotic translation initiation factor 3 complex, eIF3m	6	6
	Ino80 complex	3	7
	large ribosomal subunit	5	10
	MCM complex	6	6
	mitochondrial large ribosomal subunit	4	13
	mitochondrial nucleoid	4	7
	mitochondrial pyruvate dehydrogenase complex	6	4
	mitochondrial respiratory chain complex III	6	7
	mitochondrial small ribosomal subunit	2	8
	mitochondrion	1	49
	multi-eIF complex	5	7
	nuclear chromatin	2	15
	nuclear pre-replicative complex	4	8
	nucleolus	2	41
	nucleosome	4	5
	preribosome, large subunit precursor	3	12
	ribosome	4	53
	small ribosomal subunit	5	5
	small-subunit processome	3	15
MF	4 iron, 4 sulfur cluster binding	2	8
	amino acid binding	4	5
	ATP binding	1	112
	ATPase activity	2	12
	ATP-dependent 3'-5' DNA helicase activity	6	4
	chromatin binding	2	13
	cytochrome-c oxidase activity	5	7
	DNA helicase activity	6	4
	DNA replication origin binding	3	8
	DNA-directed DNA polymerase activity	4	8
	GTP binding	2	31
	GTPase activity	2	21
	magnesium ion binding	2	16
	NAD binding	2	12

	nucleotide binding	2	26
	proton-transporting ATP synthase activity, rotational mechanism	5	7
	proton-transporting ATPase activity, rotational mechanism	3	8
	RNA binding	2	29
MF	rRNA binding	5	11
	single-stranded DNA binding	2	8
	structural constituent of ribosome	5	98
	translation elongation factor activity	5	6
	translation initiation factor activity	4	21
	ubiquinol-cytochrome-c reductase activity	6	5

Table 5.17: GO term enrichment analysis of the upregulated genes in the t1-t4 intersection gene set of strain BS7.8. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
	autophagy	5	5
	glycogen biosynthetic process	8	3
	late nucleophagy	7	4
BP	lipid glycosylation	8	3
	piecemeal microautophagy of nucleus	3	5
	transcription, DNA-templated	2	28
	transmembrane transport	1	27
	fungal-type vacuole membrane	3	20
	Golgi apparatus	2	19
CC	integral component of membrane	1	265
	intracellular	2	18
	late endosome	5	4
	catalase activity	8	3
	heme binding	2	20
	phosphopantetheine binding	3	7
	phosphorelay sensor kinase activity	6	6
MF	pyridoxal phosphate binding	2	14
	RNA polymerase II transcription factor activity, sequence-specific DNA binding	2	23
	sodium:proton antiporter activity	8	3
	zinc ion binding	1	66

Table 5.18: GO term enrichment analysis of the downregulated genes in the t1-t4 intersection gene set of strain BS7.8. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
	acetyl-CoA biosynthetic process from pyruvate	4	4
	arginine biosynthetic process	4	6
	aromatic amino acid family biosynthetic process	4	5
	ATP synthesis coupled proton transport	3	9
	cellular amino acid biosynthetic process	3	8
	chorismate biosynthetic process	4	4
BP	chromatin silencing at silent mating-type cassette	2	10
	de novo' IMP biosynthetic process	3	6
	DNA replication initiation	3	11
	DNA strand elongation involved in DNA replication	4	6
	fatty acid biosynthetic process	3	7
	formation of translation preinitiation complex	4	16
	G-protein coupled receptor signaling pathway	4	4

BP	regulation of translational initiation	3	12
	histidine biosynthetic process	4	6
	lysine biosynthetic process via aminoadipic acid	4	4
	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	2	11
	methionine biosynthetic process	3	7
	mitochondrial electron transport, ubiquinol to cytochrome c	4	7
	mitochondrial translation	3	10
	mitotic DNA replication	4	5
	pre-replicative complex assembly involved in nuclear cell cycle DNA replication	3	8
	protein folding	2	18
	purine nucleobase biosynthetic process	4	4
	ribosomal large subunit assembly	2	10
	ribosomal small subunit biogenesis	4	6
	RNA processing	2	10
	rRNA processing	2	13
	sister chromatid cohesion	4	4
	translation	4	87
	UV-damage excision repair	4	6
CC	90S preribosome	2	9
	chaperonin-containing T-complex	4	7
	cytosolic large ribosomal subunit	4	8
	cytosolic small ribosomal subunit	4	7
	DNA replication factor C complex	5	4
	DNA replication preinitiation complex	3	7
	Elg1 RFC-like complex	5	5
	eukaryotic 43S preinitiation complex	5	14
	eukaryotic 48S preinitiation complex	5	14
	eukaryotic translation initiation factor 3 complex	4	6
	eukaryotic translation initiation factor 3 complex, eIF3e	5	6
	eukaryotic translation initiation factor 3 complex, eIF3m	5	6
	large ribosomal subunit	5	11
	MCM complex	5	6
	mitochondrial large ribosomal subunit	4	14
	mitochondrial respiratory chain complex III	5	7
	mitochondrial small ribosomal subunit	4	15
	mitochondrion	1	56
	multi-eIF complex	4	7
	nuclear chromatin	2	16
	nuclear chromosome, telomeric region	3	6
	nuclear pore	2	8
	nuclear pre-replicative complex	4	8
	nucleolus	2	46
	preribosome, large subunit precursor	3	13
	ribosome	4	55
	small ribosomal subunit	4	5
	small-subunit processome	3	15
MF	4 iron, 4 sulfur cluster binding	2	9
	amino acid binding	4	6
	ATP binding	1	130
	ATPase activity	2	16
	chromatin binding	2	12
	cytochrome-c oxidase activity	4	7
	DNA helicase activity	5	4
	DNA replication origin binding	3	9
	DNA-directed DNA polymerase activity	3	7
	GTP binding	2	34
	GTPase activity	2	20
	NAD binding	2	11

MF	nucleotide binding	2	27
	oxidoreductase activity, acting on NAD(P)H	5	4
	proton-transporting ATP synthase activity, rotational mechanism	4	7
	proton-transporting ATPase activity, rotational mechanism	3	8
	RNA binding	1	31
	rRNA binding	5	11
	single-stranded DNA binding	2	9
	structural constituent of ribosome	4	107
	structural molecule activity	3	14
	translation elongation factor activity	4	6
	translation initiation factor activity	3	20
	ubiquinol-cytochrome-c reductase activity	5	5

Table 5.19: GO term enrichment analysis of the upregulated genes in the t1-t4 intersection gene set of strain JK2.8. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	carbohydrate metabolic process	7	50
	carbohydrate transport	9	6
	cellulose catabolic process	14	9
	ER-associated ubiquitin-dependent protein catabolic process	5	4
	metal ion transport	10	3
	polysaccharide catabolic process	13	8
	protein autophosphorylation	12	3
	xylan catabolic process	9	4
CC	extracellular region	18	27
	integral component of membrane	1	84
MF	beta-glucosidase activity	8	3
	carbohydrate binding	8	9
	cellobiose dehydrogenase (acceptor) activity	16	3
	cellulase activity	16	5
	cellulose binding	16	16
	endo-1,4-beta-xylanase activity	8	5
	hydrolase activity	2	22
	hydrolase activity, hydrolyzing O-glycosyl compounds	7	33
	mannan endo-1,4-beta-mannosidase activity	16	3
	pectate lyase activity	13	5
	protein disulfide isomerase activity	12	3
	RNA polymerase II transcription factor activity, sequence-specific DNA binding	2	14
	scopolin beta-glucosidase activity	8	3
	transmembrane transporter activity	6	4

Table 5.20: GO term enrichment analysis of the downregulated genes in the t1-t4 intersection gene set of strain JK2.8. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	carbohydrate metabolic process	2	10
	one-carbon metabolic process	16	3
CC	integral component of membrane	1	57
MF	substrate-specific transmembrane transporter activity	7	7
	transferase activity	4	7

The results of the GO enrichment analysis for strain JK2.8 (Tables 5.19 and 5.20) are very similar to the results for strains MJK20.3 and BS4.1 (Tables 5.4-5.7). As in MJK20.3 and BS4.1, mainly polysaccharide catabolic processes like, e.g. “cellulose catabolic process” and “xylan catabolic process” as well as lignocellulolytic enzyme activities like “cellulase activity”, “xylanase activity”, and “pectate lyase activity” are enriched. The main difference to MJK20.3 is the mostly higher number of genes that are associated with those categories in JK2.8. Regarding the analysis of the downregulated genes in JK2.8, the results are also very similar to MJK20.3. Here, mainly “one carbon metabolic process” as well as “transferase activity” are enriched.

The results for BS6.4 and BS7.8 (Tables 5.15-5.18) are completely different to those for MJK20.3 and JK2.8. In BS6.4 and BS7.8, mainly processes like “mitophagy”, “late nucleophagy”, “autophagy”, and “glycogen biosynthesis process” are upregulated. In BS6.4 also some carbohydrate catabolic processes like “cellulose catabolic process” and “carbohydrate transport” (much lower number of genes compared to MJK20.3) are upregulated, which are completely absent in BS7.8. On the other hand, only in BS7.8, genes connected to the cellular components Golgi and endosome are upregulated. Among the downregulated genes in BS6.4 and BS7.8, many GO terms describing essential processes in the cell like replication, transcription, translation, and structures that are connected with those processes such as ribosomes with translation are enriched. Furthermore, biosynthesis processes for, e.g. arginine and histidine are downregulated. Finally, also processes and structures that are required for cell respiration are among the enriched genes of the downregulated gene set. Here, mainly processes connected to the mitochondrial structure and the respiration chain are detectable. The main difference between BS7.8 and BS6.4 is the higher number of downregulated genes connected to the above-mentioned essential processes in BS7.8.

5.3.2.2 Differential expression of proteases

The purpose of the investigation of protease expression has already been mentioned previously (see chapter 5.3.1.2). An overview of predicted *T. thermophilus* proteases was provided in this chapter as well.

The total number of predicted protease genes that are differentially expressed in the strains BS6.4, BS7.8, and JK2.8 at different points in time after the cellulose spike, is shown in Figure 5.46. Here, the results for MJK20.3 are included to enable a better comparison. The number of differentially expressed predicted protease genes in JK2.8 is very similar to MJK20.3, but especially a higher number of downregulated genes at the points in time t1-t2 after the spike can be detected. In BS6.4 the number of up- and downregulated genes is at all points in time much higher compared to MJK20.3. Particularly the number of downregulated genes at t2 is the highest of all strains. In BS7.8 the number of up- and downregulated genes is also higher at all points in time compared to MJK20.3. For BS7.8, the number of upregulated predicted protease genes at t1, t2, and t4 and the number of downregulated genes at t3 and t4 is the highest among all strains. For MJK20.3 and JK2.8 the number of differentially expressed genes peaks at t2 and decreases toward t4. This trend can also be observed for BS6.4, but with a much higher number of differentially expressed genes at all points in time. This is not the case for BS7.8, where a constant increase of differentially expressed genes toward t4 can be observed. In total up to

approximately 40 % of the predicted protease genes are differentially expressed in MJK20.3, 65 % in BS6.4, 55 % in BS7.8, and 45 % in JK2.8.

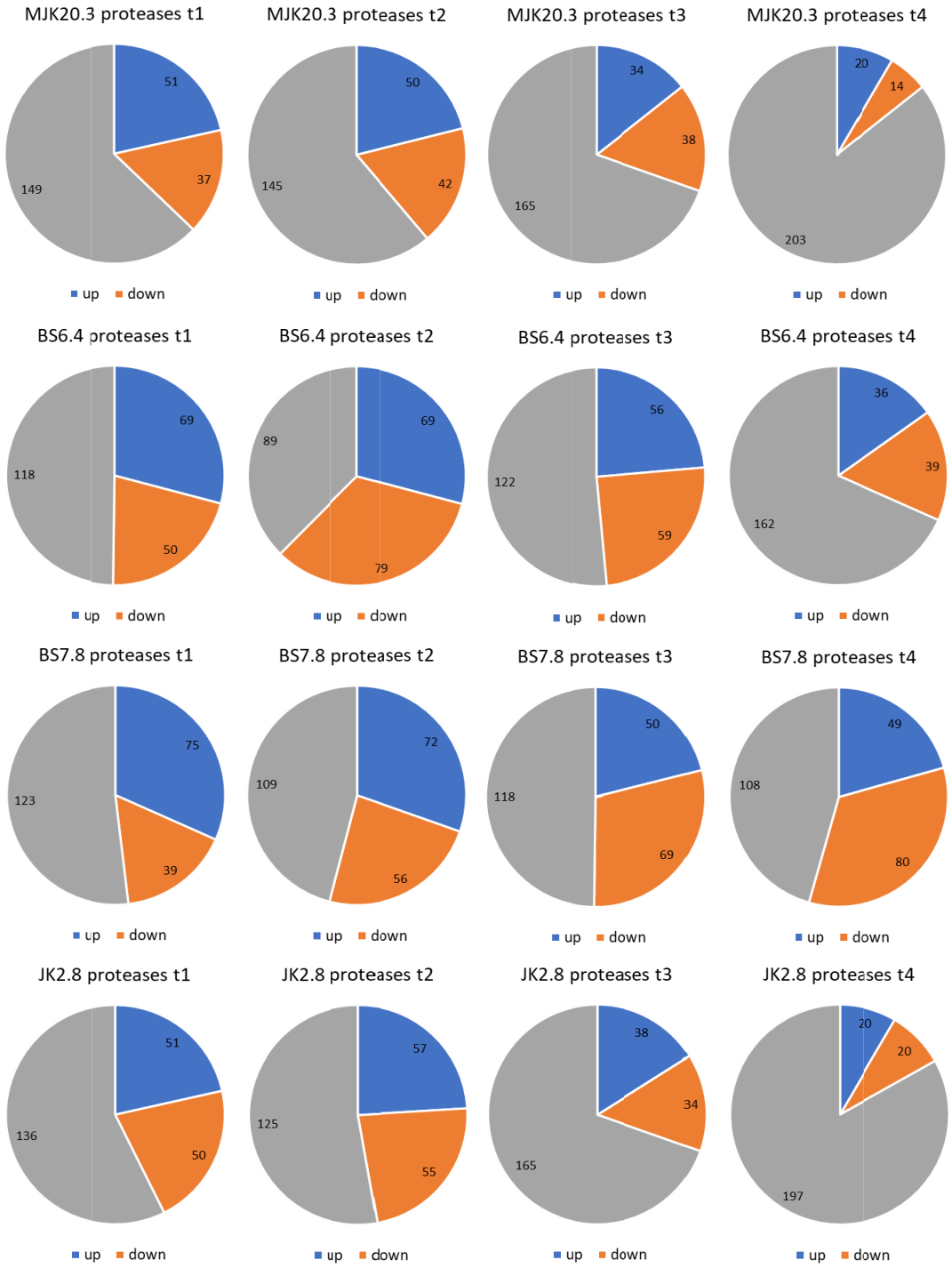

Figure 5.46: Number of differentially expressed predicted protease genes of strains MJK20.3, BS6.4, BS7.8, and JK2.8 at different points in time after the cellulose spike. Shown are the numbers of up- (blue) and downregulated (orange) genes as well as genes with no differential expression (grey) at the points in time: 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after cellulose spike (t4) compared to the respective steady state condition.

To get a detailed overview of the expression profiles of the two strains regarding protease gene expression, a heatmap (Figure 5.47) was created, showing the differential expression levels of genes

belonging to the different classes of proteases among the three strains and the reference strain. The expression profile of strains JK2.8 and MJK20.3 is very similar but a stronger up- and downregulation of the single genes especially for t1-t3 is observable for JK2.8. In addition, there is a higher total number of differentially expressed predicted protease genes at the respective points in time. This has already been seen in Figure 5.46. At t4 a very low number of genes encoding for predicted proteases is differentially expressed in both strains. The expression profile of BS6.4 and BS7.8 is very similar but completely different compared to MJK20.3 and JK2.8. Here, an extremely high number of predicted protease genes is differentially expressed and a stronger up- and downregulation of the single genes similar to the relation between JK2.8 and MJK20.3 can be observed for BS7.8. At t4 still a high number of genes encoding for predicted proteases is differentially expressed in BS6.4 and BS7.8, whereas this number is much higher for BS7.8. Differential expression is evenly distributed among the genes belonging to the different protease classes for all strains.

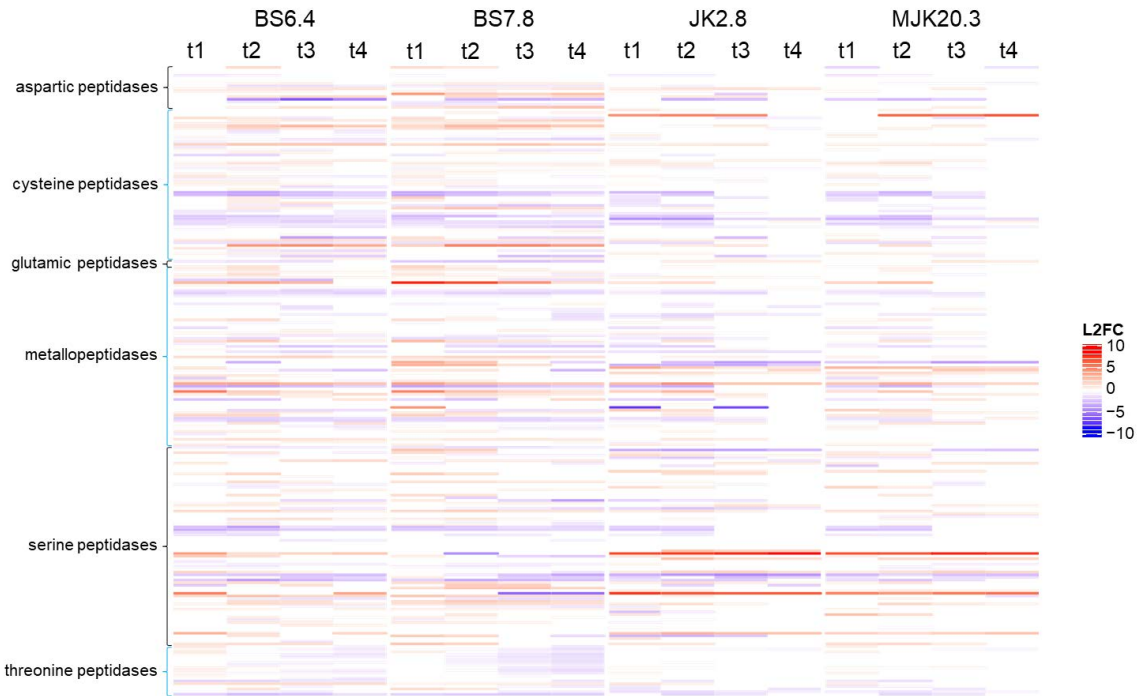

Figure 5.47: Heatmap with differential expression values of all predicted protease genes in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3. Shown are the log2 fold change values (L2FC) of single predicted protease genes belonging to the different protease classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

To narrow down the number of differentially expressed genes and filter out the most important differentially expressed genes, the t1-t4 intersection was analyzed next. A gene was included, when differential expression over all points in time was detected after the cellulose spike in at least one of the three strains or the reference strain. Table 5.21 summarizes these findings. A further heatmap (Figure 5.48) visualizes the differential expression levels of the respective genes that belong to the t1-t4 intersection area.

Table 5.21: Number of genes belonging to different protease classes that are differentially expressed at all points in time after the spike in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3.

class	genes	MJK20.3		BS6.4		BS7.8		JK2.8	
		t1-t4 up	t1-t4 down	t1-t4 up	t1-t4 down	t1-t4 up	t1-t4 down	t1-t4 up	t1-t4 down
aspartic peptidases	17	0	0	1	0	3	2	1	0
cysteine peptidases	56	0	1	5	10	11	9	0	1
glutamic peptidases	2	0	0	0	0	0	1	0	0
metallopeptidases	68	3	1	7	6	9	7	4	3
serine peptidases	75	5	1	5	7	8	4	6	3
threonine peptidases	19	0	1	1	2	1	1	0	1
total	237	8	4	19	25	32	24	11	8

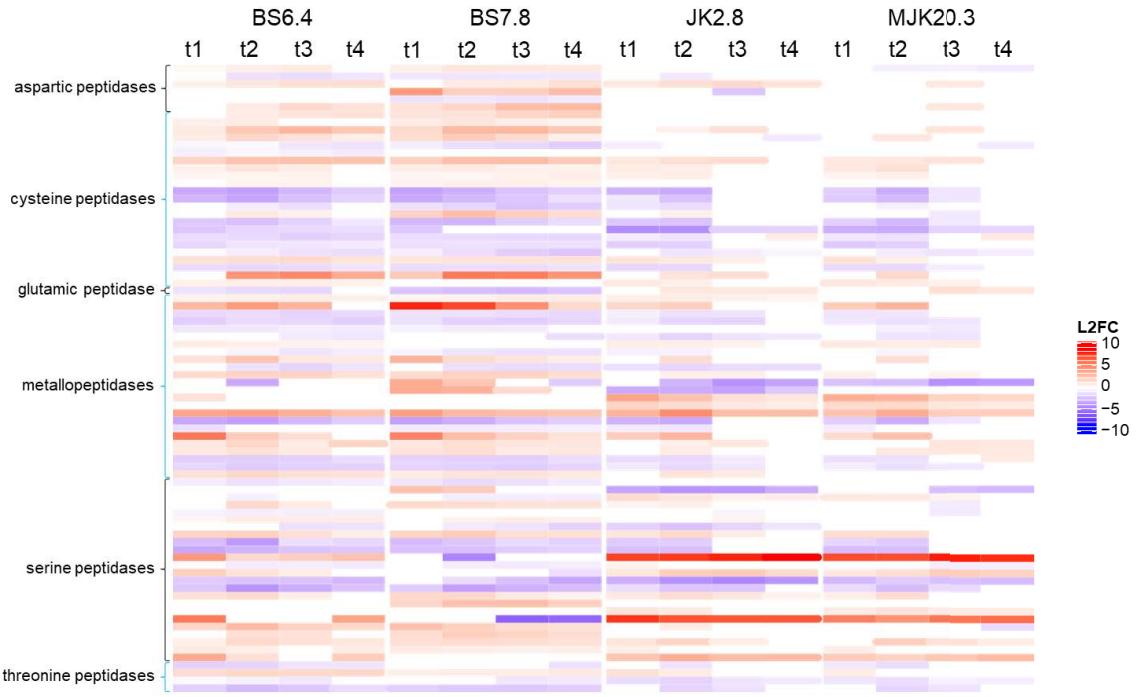

Figure 5.48: Heatmap with expression values of predicted protease genes that are differentially expressed at all points in time after the cellulose spike in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3. Shown are the log2 fold change values (L2FC) of single predicted protease genes belonging to the different protease classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

The heatmap of the t1-t4 intersection genes confirms the results of the previous figures: the expression pattern between strains BS6.4 and BS7.8 on the one hand, and JK2.8 and MJK20.3 on the other hand is very similar with some predicted protease genes having slightly stronger differential expression values in BS7.8 and JK2.8, respectively. Furthermore, the same expression trends, as already described above, can be observed in more detail compared to the previous heatmap.

To investigate and compare gene expression in detail, graphs showing L2FC values as well as graphs showing the mean of the normalized raw count values were created as previously (see chapter 5.3.1.2) described (Figures 5.49-5.52).

As already mentioned, the main purpose for analyzing the expression of proteases in the different strains is to identify highly expressed proteases, especially proteases that are upregulated after the cellulose spike. After the identification of those proteases, deletion experiments with the aim to reduce proteolytic

degradation of secreted enzymes can be conducted in future experiments. With the help of the presented figures, those proteases can be identified for the growth on glucose (SS condition) and cellulose (t1-t4) as well as for the different strains. If for example a *clr2* deletion strain is used as a production host with growth on glucose, the respective highly expressed proteases should be chosen for deletion. This illustrates that choosing proteases for deletion is highly individual according to the condition and strain that is planned to be used. Therefore, in the following only the most interesting expression trends and top up- and downregulated putative proteases of each category will be described briefly.

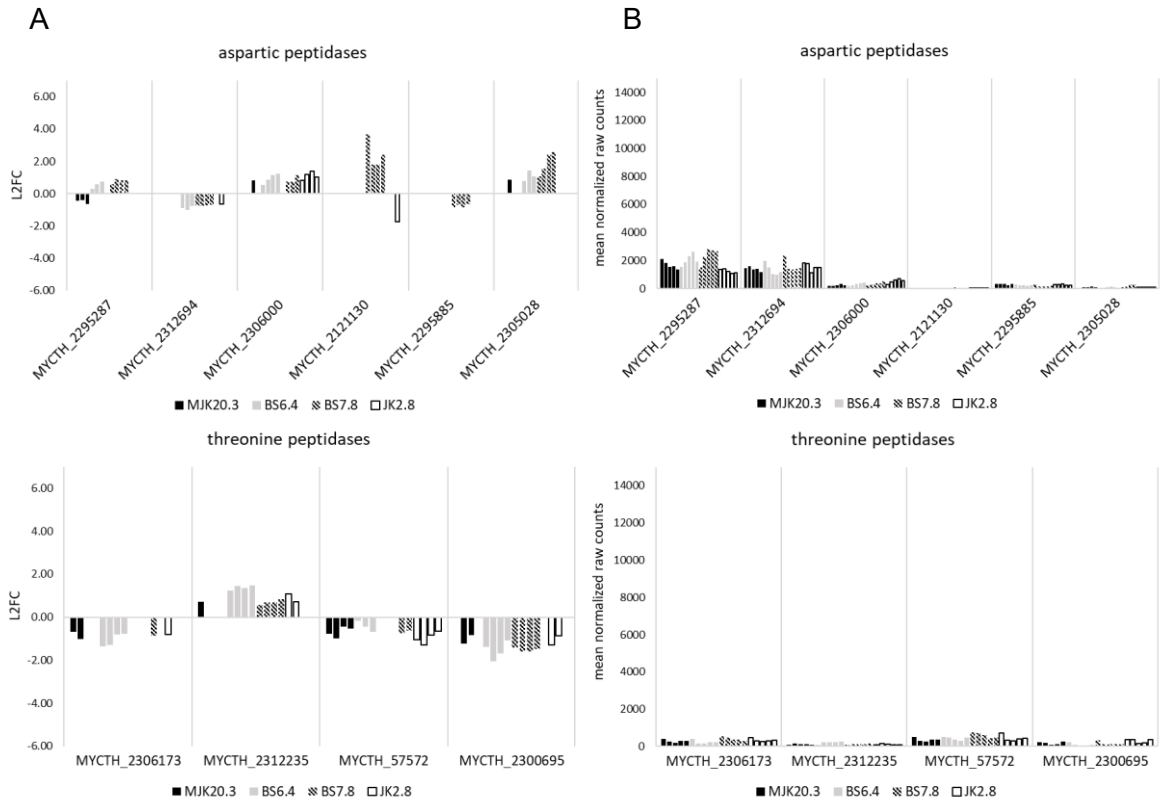

Figure 5.49: Differentially expressed genes encoding for predicted aspartic and threonine peptidases in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Among the genes encoding for predicted aspartic peptidases (Figure 5.49), the highest expressed genes are MYCTH_2295287 which is downregulated in MJK20.3 and upregulated in BS6.4 and BS7.8, and MYCTH_2312694 which has high expression levels in all strains although it is downregulated after the cellulose spike in all strains except MJK20.3.

The threonine peptidase category only consists of predicted protease genes with either low raw count values or low L2FC values.

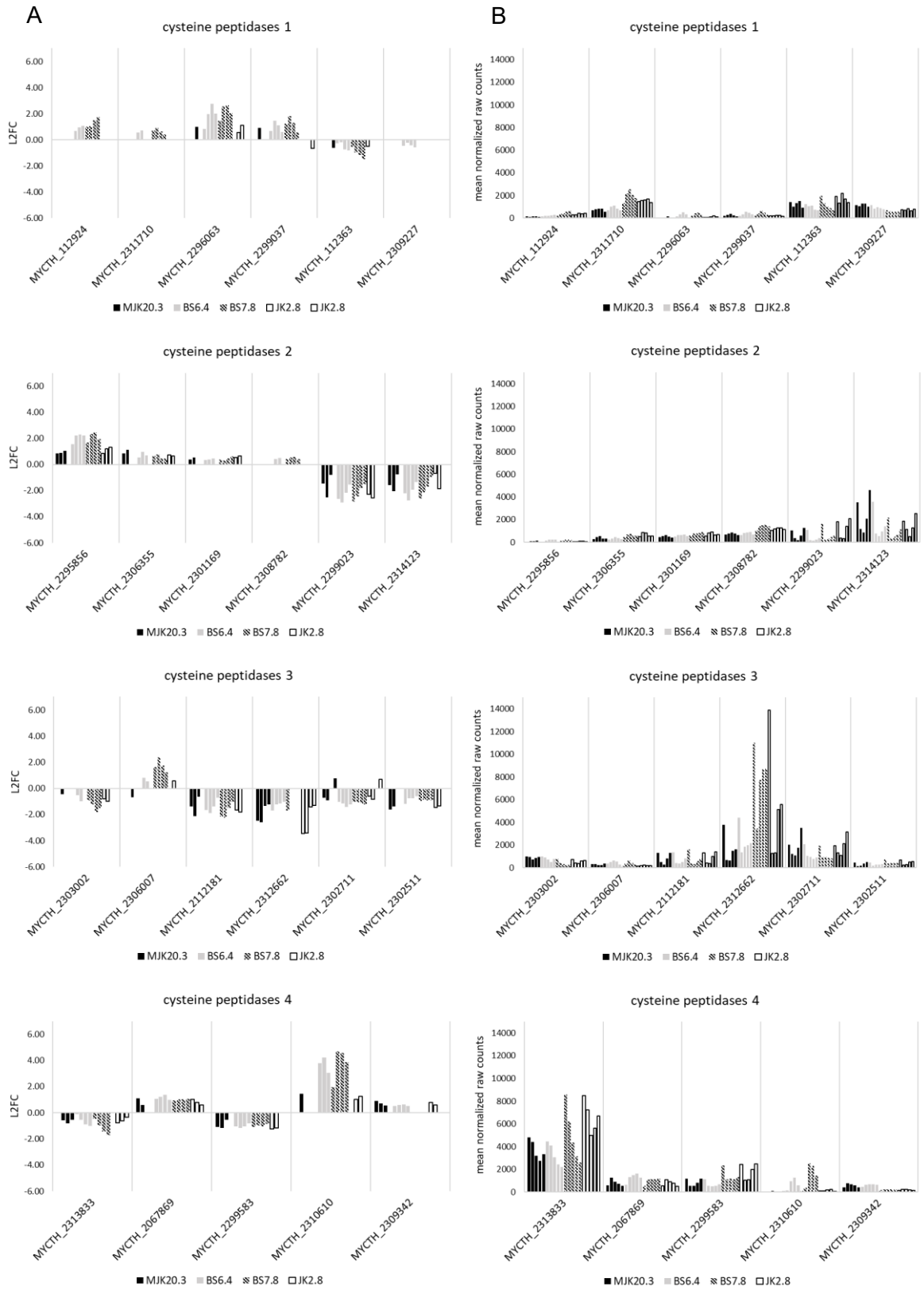

Figure 5.50: Differentially expressed genes encoding for predicted cysteine peptidases in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Among the genes encoding for predicted cysteine peptidases (Figure 5.50), the strongest downregulation and low raw count values after the cellulose spike can be observed for MYCTH_2299023 and MYCTH_2314123. Downregulated genes that still have high raw count values, especially in BS7.8 and JK2.8, are MYCTH_2312662, MYCTH_2313833, and MYCTH_112363. Upregulated genes that have high expression values (highest in BS7.8), are MYCTH_2311710, MYCTH_2308782, and MYCTH_2310710.

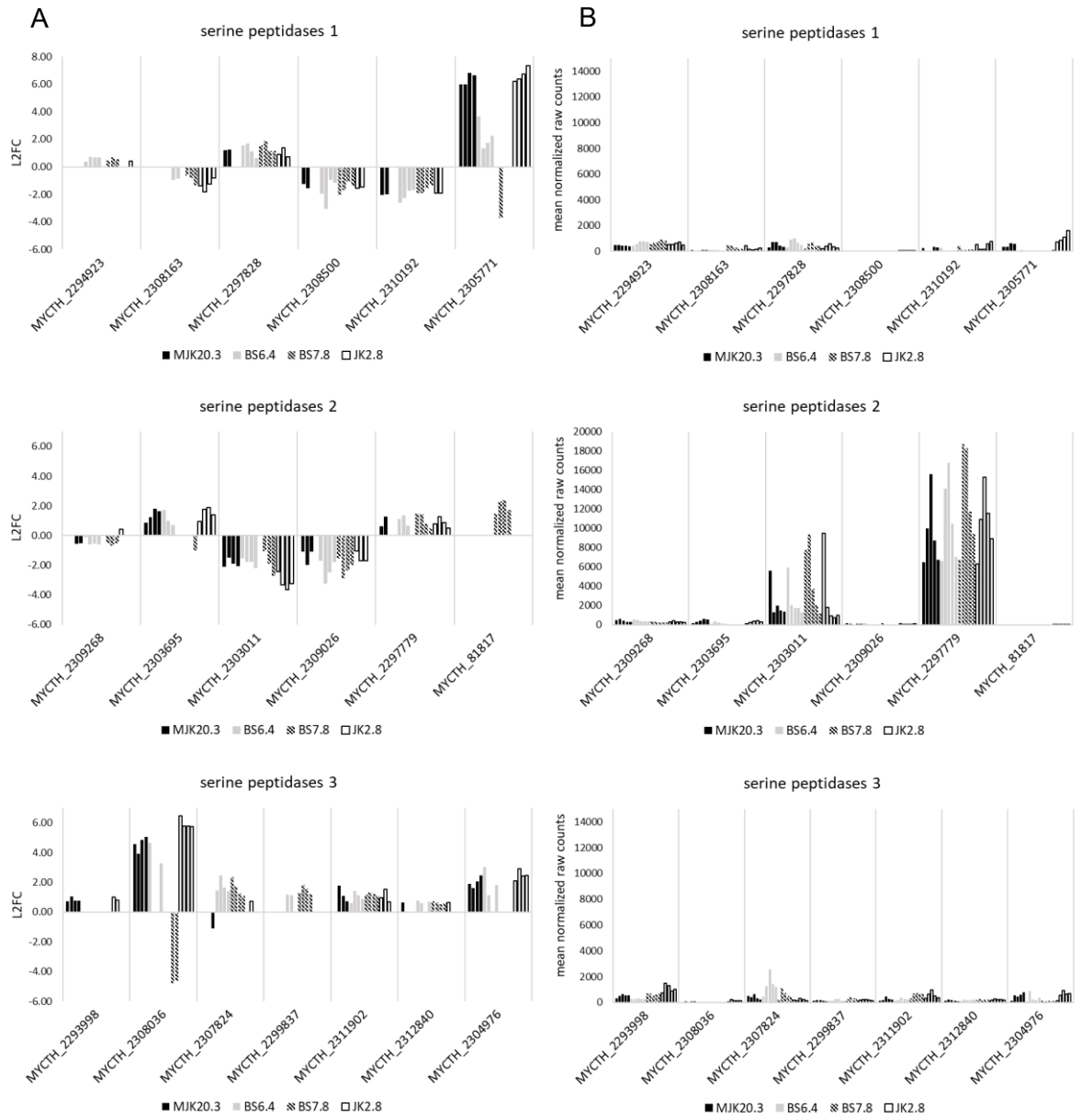

Figure 5.51: Differentially expressed genes encoding for predicted serine peptidases in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Among the genes encoding for predicted serine peptidases (Figure 5.51), many genes with low raw count values can be detected. Among the upregulated genes, MYCTH_2305771 has an extremely high L2FC value for MJK20.3 and JK2.8 only. MYCTH_2297779 has an extremely high raw count value in all strains. A strong downregulation, especially in JK2.8 and BS7.8 can be observed for MYCTH_2303011 together with a low remaining raw count value.

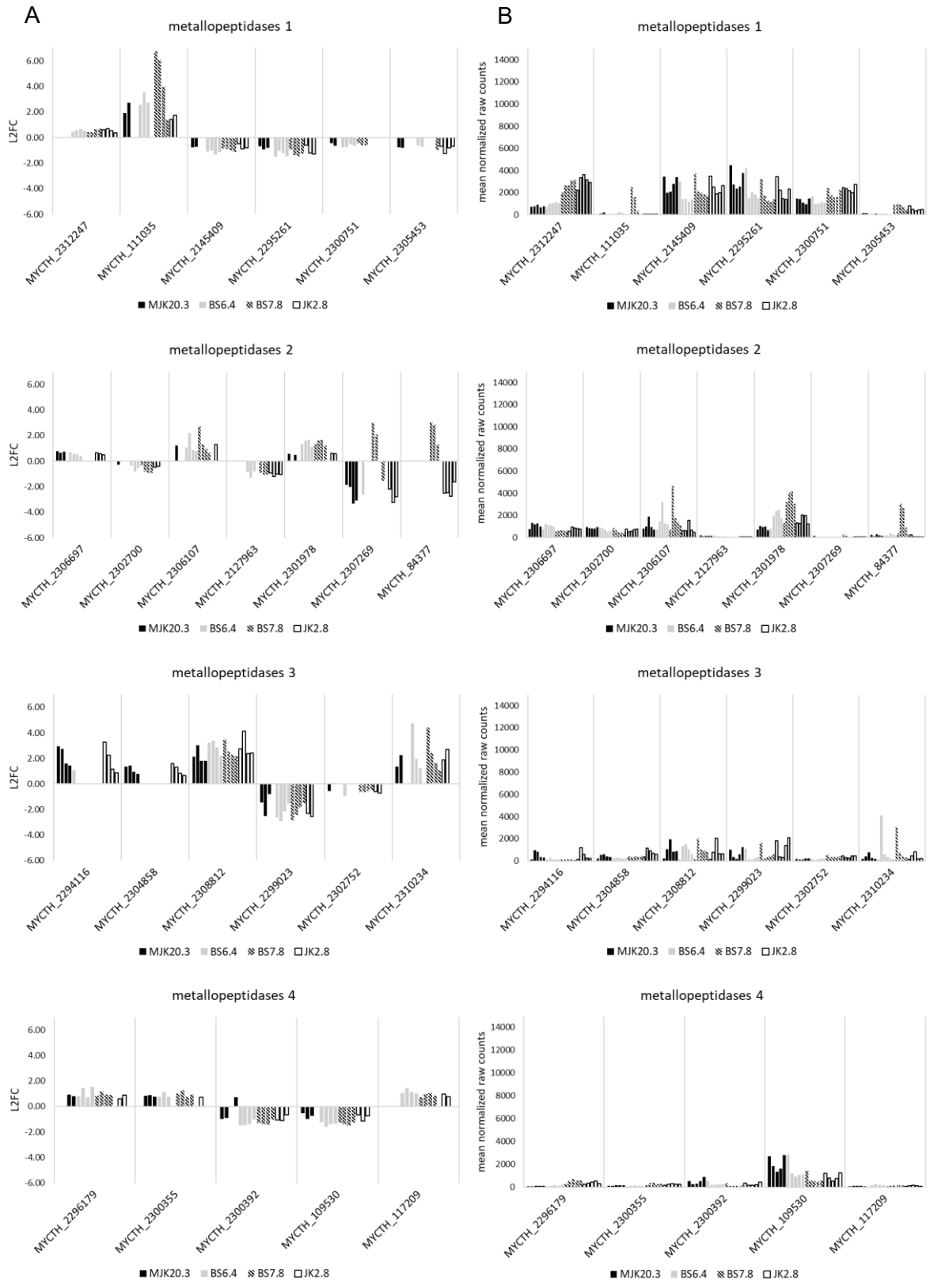

Figure 5.52: Differentially expressed genes encoding for predicted metallopeptidases in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Regarding the expression of genes encoding for predicted metallopeptidases (Figure 5.52), upregulated genes with high expression values at single points in time are MYCTH_111035 (only t1, t2 in BS7.8), MYCTH_2306107 (t1 only, high values in BS6.4 and BS7.8 only), MYCTH_84377 (only t1,t2 in BS7.8), and MYCTH_2310234 (only t1). Upregulated predicted protease genes with high expression levels at all points in time include: MYCTH_2312247 (BS7.8 and JK2.8 only) and MYCTH_2301978 (all strains, except MJK20.3, highest in BS7.8). Among the downregulated genes encoding predicted metallopeptidases that still have a high expression level and are equal for each strain are MYCTH_2145429, MYCTH_2295261, and MYCTH_2300751. The slightly downregulated MYCTH_109530 has high expression levels in MJK20.3 only.

To identify further interesting highly expressed proteases that do not belong to the t1-t4 intersection area and are therefore not mentioned yet, the same graphs, as seen in Figures 5.49-5.52 were created with all genes encoding for predicted proteases, that possess a signal peptide. The results are shown in Figures 5.53-5.54.

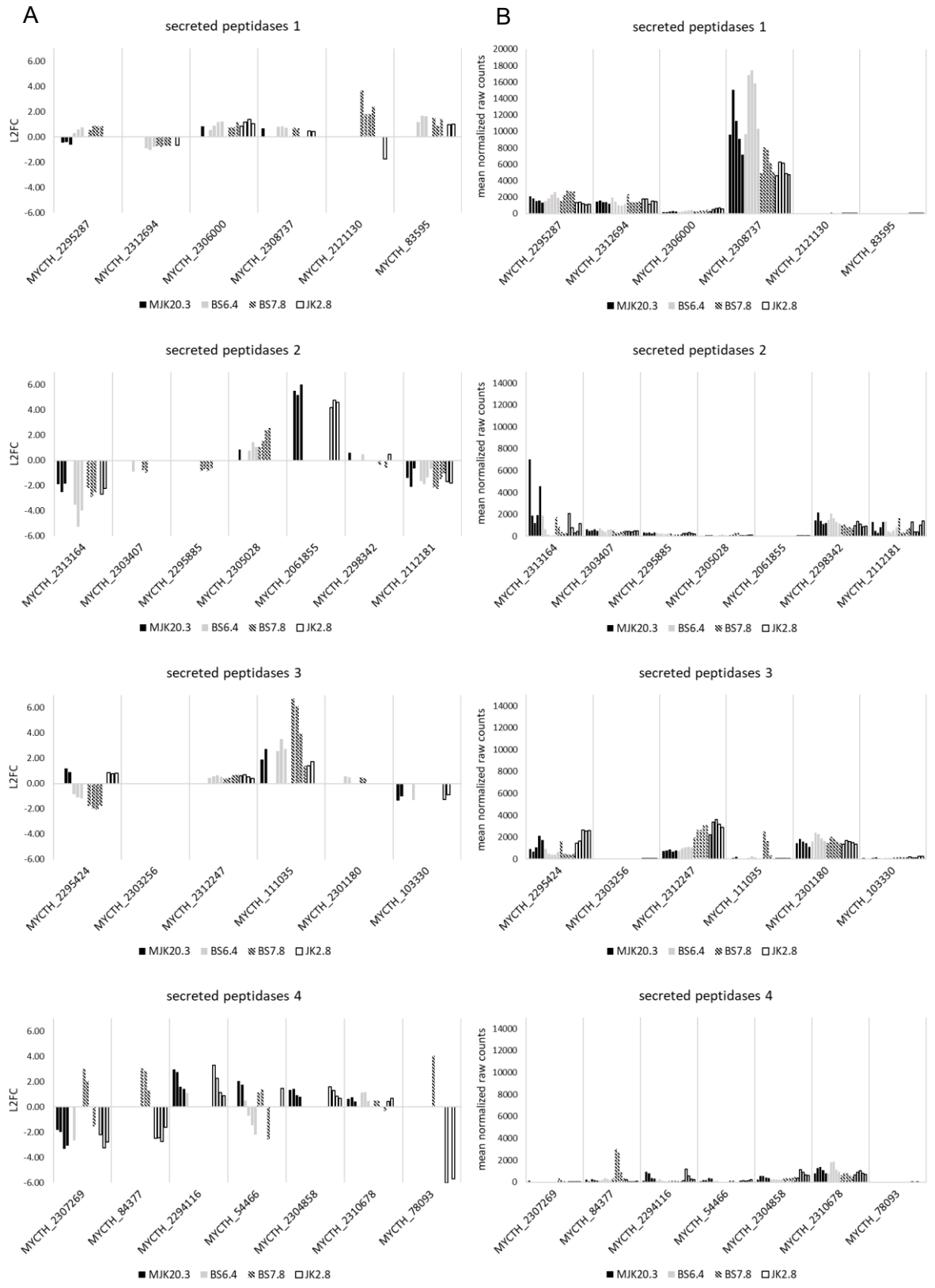

Figure 5.53: Differentially expressed genes encoding for predicted peptidases with a signal peptide in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

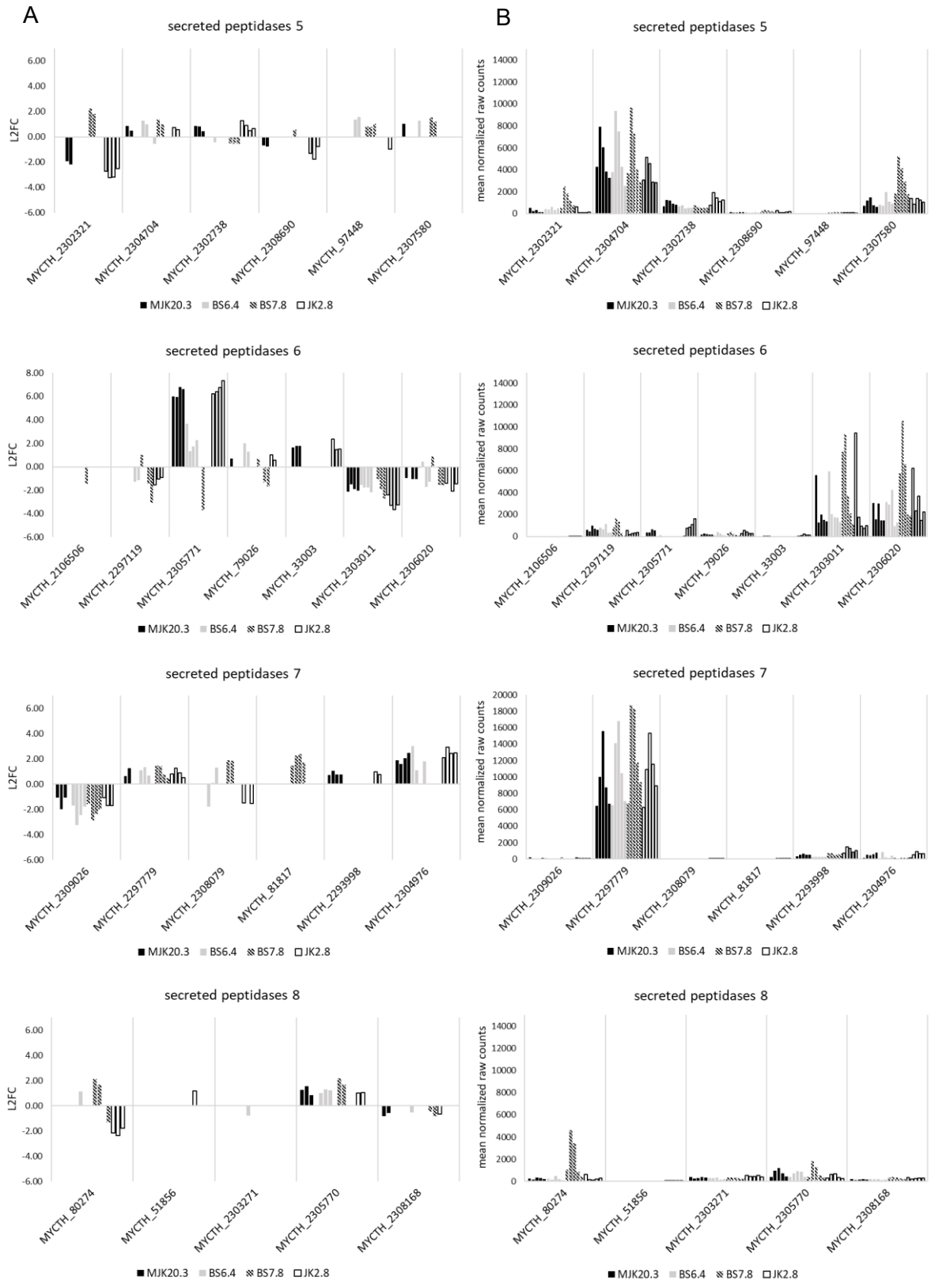

Figure 5.54: Differentially expressed genes encoding for predicted peptidases with a signal peptide in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Genes encoding for predicted secreted peptidases (Figures 5.53-5.54) that are upregulated and have high expression values include MYCTH_2308737 (highest in MJK20.3 and JK2.8), MYCTH_2312247 (high in BS7.8, JK2.8 only), MYCTH_2301180 (equal for all strains), MYCTH_2304704 (only in JK2.8 slightly lower), and MYCTH_2297779 (equal for all strains). Predicted protease genes that only have high expression levels at single points in time are, e.g. MYCTH_2295424 (upregulation in MJK20.3 and JK2.8, downregulation in BS6.4 and BS7.8), MYCTH_111035 (t1 and t2, BS7.8 only), MYCTH_80274 (t1, t2 in BS7.8 only), and MYCTH_84377 (t1, t2 in BS7.8 only). Downregulated genes encoding for predicted secreted proteases include MYCTH_2313164 (high expression values in MJK20.3 only), MYCTH_2303011 (high expression values in SS, low expression values after the spike), and MYCTH_2306020 (high expression values in SS, low expression values after the spike).
In summary, no specific expression pattern is recognizable and protease gene expression itself is very similar among the strains. Only the high expression values in BS7.8 for many predicted protease genes stand out in comparison to the other strains. As already mentioned, the expression of protease genes must be examined individually according to the strain and conditions that will be used in future experiments.

5.3.2.3 Differential expression of carbohydrate active enzymes (CAZYs)

The purpose of the investigation of CAZY expression has already been mentioned previously (see chapter 5.3.1.3). An overview of predicted *T. thermophilus* CAZYs was provided in this chapter as well. The total number of predicted CAZY genes that were differentially expressed in the strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 after the cellulose spike, is shown in Figure 5.55. Here, a higher number of up- and downregulated genes in JK2.8 at all points in time after the spike in comparison to MJK20.3 can be detected. In BS6.4 and BS7.8 a lower number of upregulated genes and a higher number of downregulated genes can be observed at t1-t4 in comparison to MJK20.3. The number of upregulated genes at all points in time is the lowest in BS7.8 and the number of downregulated genes at all points in time the highest in BS7.8. For MJK20.3 and JK2.8 the number of differentially expressed genes peaks at t2 and decreases toward t4. This trend cannot be detected in BS6.4 and BS7.8, where a constant number of differentially expressed genes at all points in time can be observed instead. In total up to approximately 50 % of the predicted CAZY genes are differentially expressed in MJK20.3, 48 % in BS6.4, 48 % in BS7.8, and 55 % in JK2.8. The relation of up- and downregulated genes differs between strains MJK20.3 and JK2.8 (up: down ~3:1) on the one hand and BS6.4 and BS7.8 (up: down ~ 1:1) on the other hand.

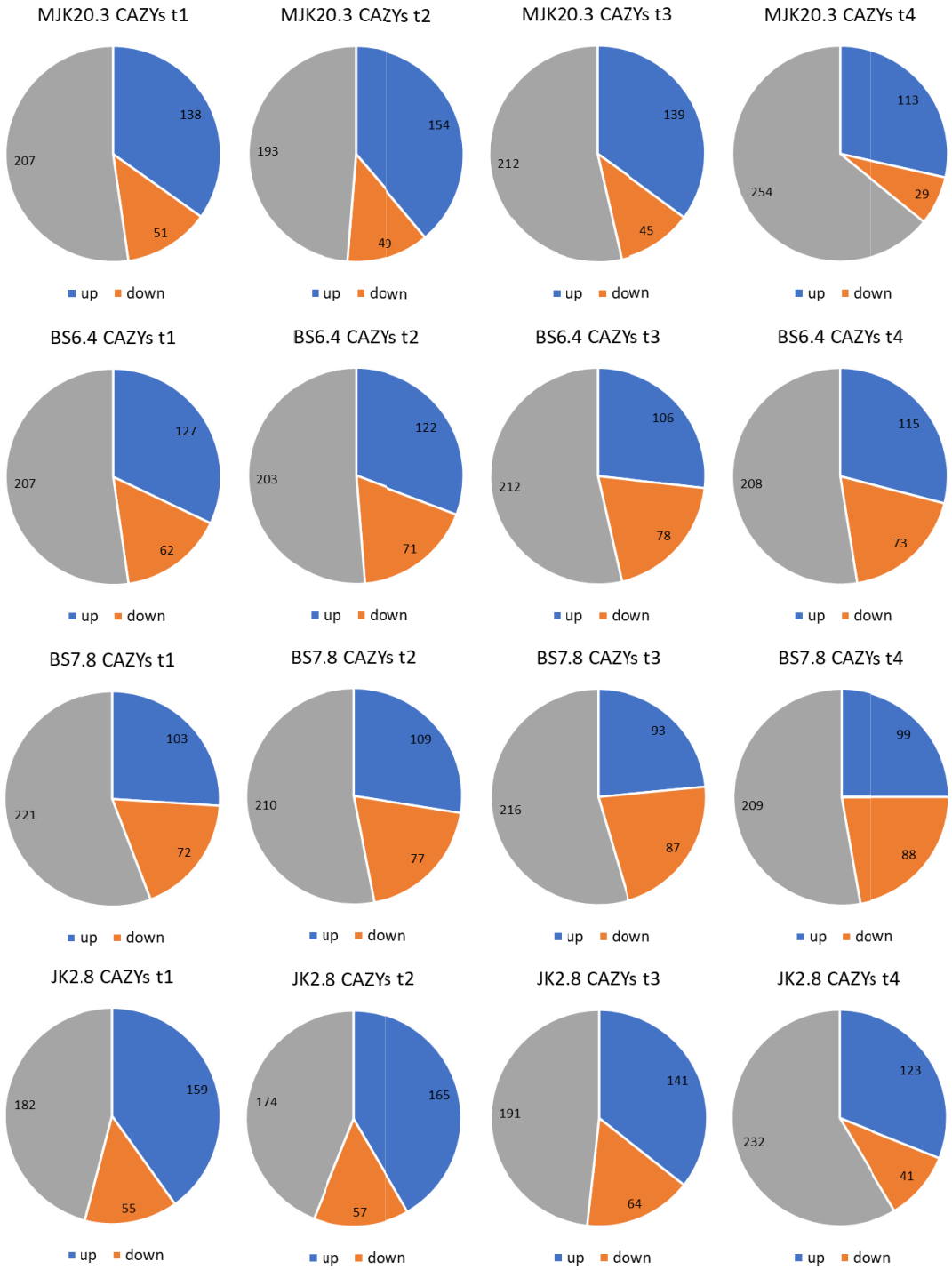

Figure 5.55: Number of differentially expressed predicted CAZY genes of strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 at different points in time after the cellulose spike. Shown are the numbers of up- (blue) and downregulated (orange) genes as well as genes with no differential expression (grey) at the points in time: 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after cellulose spike (t4) compared to the respective steady state condition.

To get a detailed overview of the expression profiles of the two strains regarding predicted CAZY gene expression, a heatmap (Figure 5.56) was created, showing the differential expression levels of genes belonging to the different classes of CAZYs among the two strains. The expression profile of strains JK2.8 and MJK20.3 is almost identical but a slightly stronger upregulation of single genes is observable in JK2.8 as well as a higher total number of differentially expressed predicted CAZY genes at the

respective points in time as already seen in Figure 5.55. At t4 a slightly lower number of genes encoding for predicted CAZYs is differentially expressed in both strains compared to the earlier points in time. The genes that belong to the classes cellulases, hemicellulases, pectinases, and esterases as well as single genes from the “other” class are showing the strongest upregulation and a high expression even at t4 in those strains. The genes that are showing the strongest downregulation belong to the starch metabolism class. The expression profile of strains BS6.4 and BS7.8 is completely different compared to JK2.8 and MJK20.3. In BS6.4, almost all predicted cellulase, hemicellulase, pectinase, and esterase genes as well as single “other” genes that are strongly upregulated in MJK20.3 are only slightly upregulated. In BS7.8 almost all predicted cellulase, hemicellulase, pectinase, and esterase genes as well as single “other” genes, that are only slightly upregulated in BS6.4 do not show any differential expression or are even downregulated. Genes belonging to all other classes do not seem to be severely influenced by the deletion.

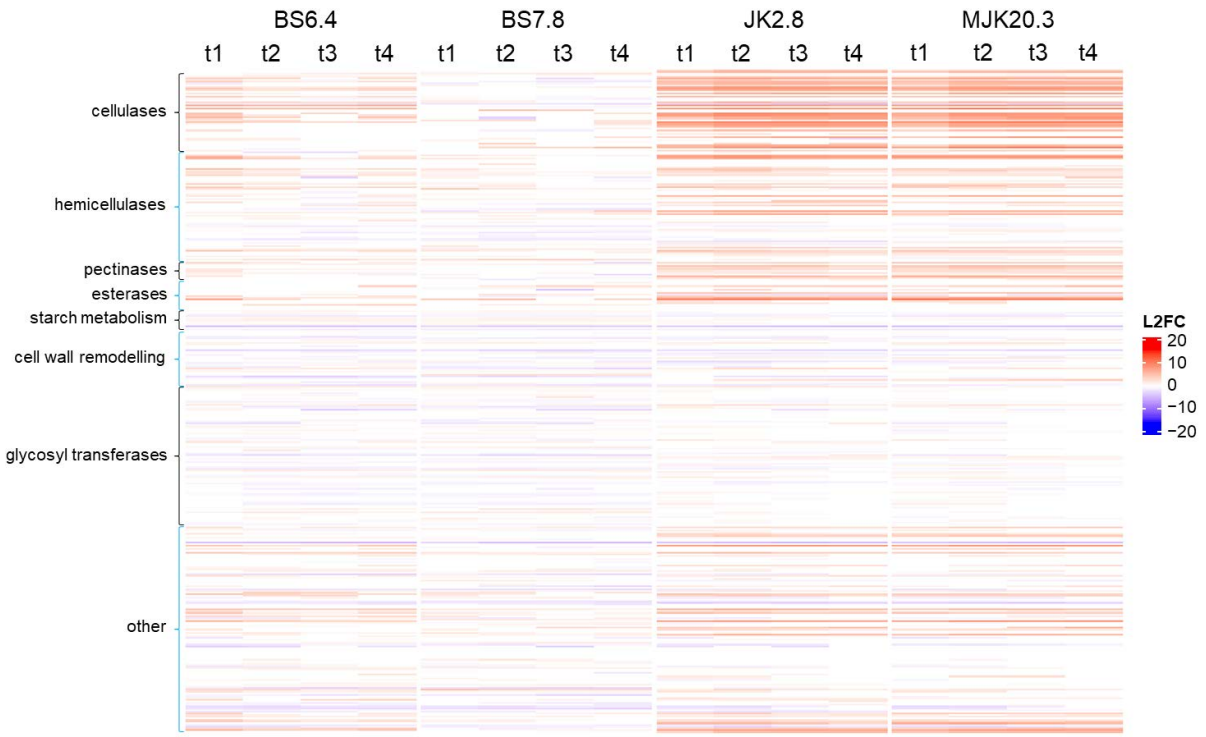

Figure 5.56: Heatmap with differential expression values of all predicted CAZY genes in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3. Shown are the log2 fold change values (L2FC) of single predicted CAZY genes belonging to the different classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

To narrow down the number of differentially expressed genes and filter out the most important differentially expressed genes, the t1-t4 intersection was analyzed next. A gene was included, when differential expression over all points in time after the cellulose spike in at least one of the strains was detected. In Table 5.22, the numbers of genes that are differentially expressed at all points in time after the spike are shown for the strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3. A further heatmap (Figure 5.57) visualizes the differential expression levels of genes that belong to the t1-t4 intersection area.

Table 5.22: Number of genes belonging to different CAZY classes that are differentially expressed at all points in time after the spike in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3.

			MJK20.3		**BS6.4**		**BS7.8**		**JK2.8**	
class	type	genes	t1-t4 up	t1-t4 down	t1-t4 up	t1-t4 down	t1-t4 up	t1-t4 down	t1-t4 up	t1-t4 down
cellulases	endoglucanases	9	7	0	2	0	1	0	8	0
	cellobiohydrolases	7	5	0	3	0	0	0	5	0
	ß-glucosidases	9	4	1	4	1	0	1	4	1
	LPMOs	24	14	0	1	0	0	0	13	0
hemicellulases	xylanases	12	5	0	2	0	1	0	5	0
	xylosidases	4	3	0	2	0	1	0	2	0
	endoarabinases	3	0	0	0	0	0	0	0	0
	exoarabinases/ arabinofuranosidases	11	3	0	2	0	2	0	3	0
	mixed-linked glucanase	5	1	0	0	0	0	1	3	0
	mannanases	10	3	0	1	1	2	1	3	1
	mannosidases	11	3	1	0	1	0	2	3	2
	galactanases	2	1	0	0	0	0	0	1	0
	galactosidases	7	3	0	4	0	2	0	3	0
pectinases	polygalacturonases	2	1	0	0	0	0	0	1	0
	rhamnosidases	1	0	0	1	0	1	0	0	0
	pectin lyases	8	7	0	2	0	0	0	6	0
esterases	feruloyl esterases	4	0	0	0	0	0	0	0	0
	acetyl esterases	9	5	0	1	0	1	0	5	1
	pectin esterases	4	0	0	0	0	0	0	0	0
	glycuronoyl esterases	2	0	0	0	0	0	0	0	0
starch metabolism	alpha amylases	4	0	2	1	1	2	2	0	1
	alpha glucosidases	4	0	0	3	0	3	0	0	0
	glucoamylases	2	0	1	1	1	1	1	0	1
	glycogen debranching enzymes	2	0	1	0	1	0	1	0	1
cell wall remodeling	glucanases	12	1	1	1	3	1	2	0	2
	transglucosylases	4	0	1	0	1	0	1	0	2
	chitosanases	2	1	0	0	0	0	0	1	0
	diacetylmuramidase	1	0	0	0	0	0	0	0	1
	glucosaminidase	3	0	0	0	0	0	1	0	0
	chitinases	8	1	0	2	2	3	2	1	0
	crosslinking transglycosidase	3	0	1	0	1	0	1	0	1
glycosyl-transferases	glycosyltransferases	83	4	2	8	11	10	17	5	4
other	other	124	24	6	13	6	8	6	20	4
	total	396	96	17	9	8	4	7	12	4

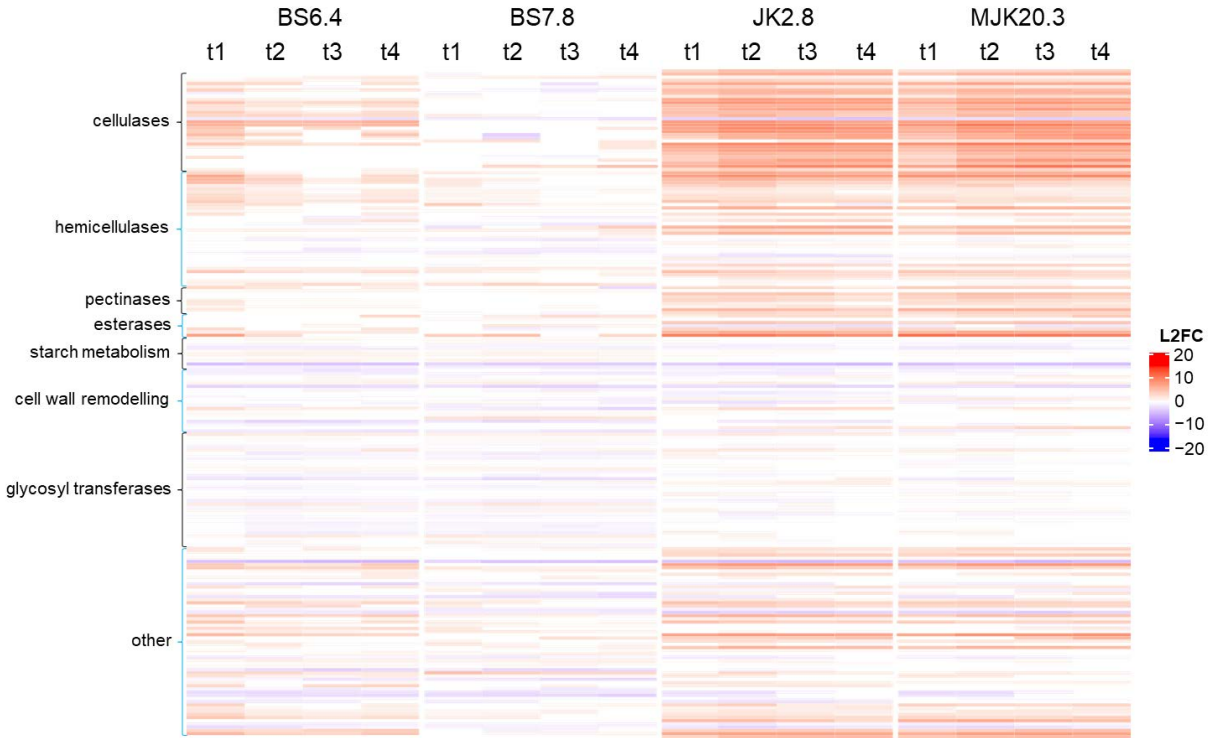

Figure 5.57: Heatmap with expression values of predicted CAZY genes that are differentially expressed at all points in time after the cellulose spike in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3. Shown are the log2 fold change values (L2FC) of single predicted CAZY genes belonging to different classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

The heatmap of the genes that belong to the t1-t4 intersection area confirms the results of the previous graphs: the expression pattern between strains JK2.8 and MJK20.3 is very similar with some predicted CAZY genes having higher expression levels in JK2.8. Furthermore, expression of predicted cellulase, hemicellulase, pectinase, esterase, and single “other” genes is extremely reduced in BS6.4 and even completely missing in BS7.8. All genes belonging to other classes are only showing small expression differences between the strains. Generally, it becomes clearer, that the most and strongest differential expression in JK2.8 and MJK20.3 derives from predicted cellulase genes, followed by predicted hemicellulase, pectinase, and esterase genes as well as single genes belonging to the “other” class.

To investigate and compare gene expression in detail, graphs showing L2FC values as well as graphs showing the mean of the normalized raw count values were created as previously (see chapter 5.3.1.2) described (Figures 5.58-5.70). The expression of predicted CAZY genes in MJK20.3, especially the top genes of each CAZY class have already been discussed in a previous chapter. Therefore, the differences of the strains BS6.4, BS7.8, and JK2.8 to MJK20.3 especially with regards to the highest expressed predicted CAZY genes of each class are described in the following. If not mentioned specifically, the expression of genes is regarded as similar.

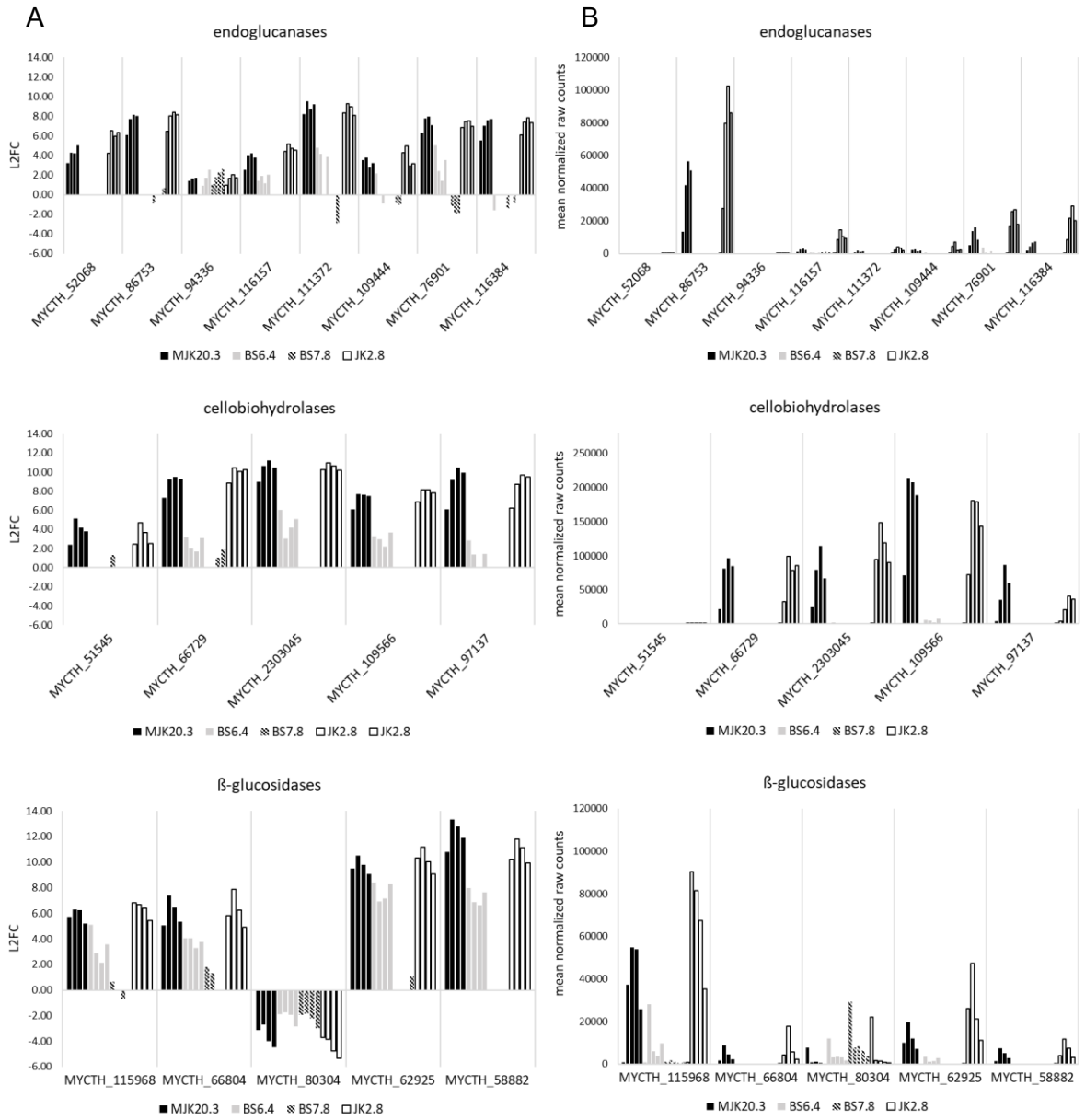

Figure 5.58: Differentially expressed predicted cellulase genes (part 1) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

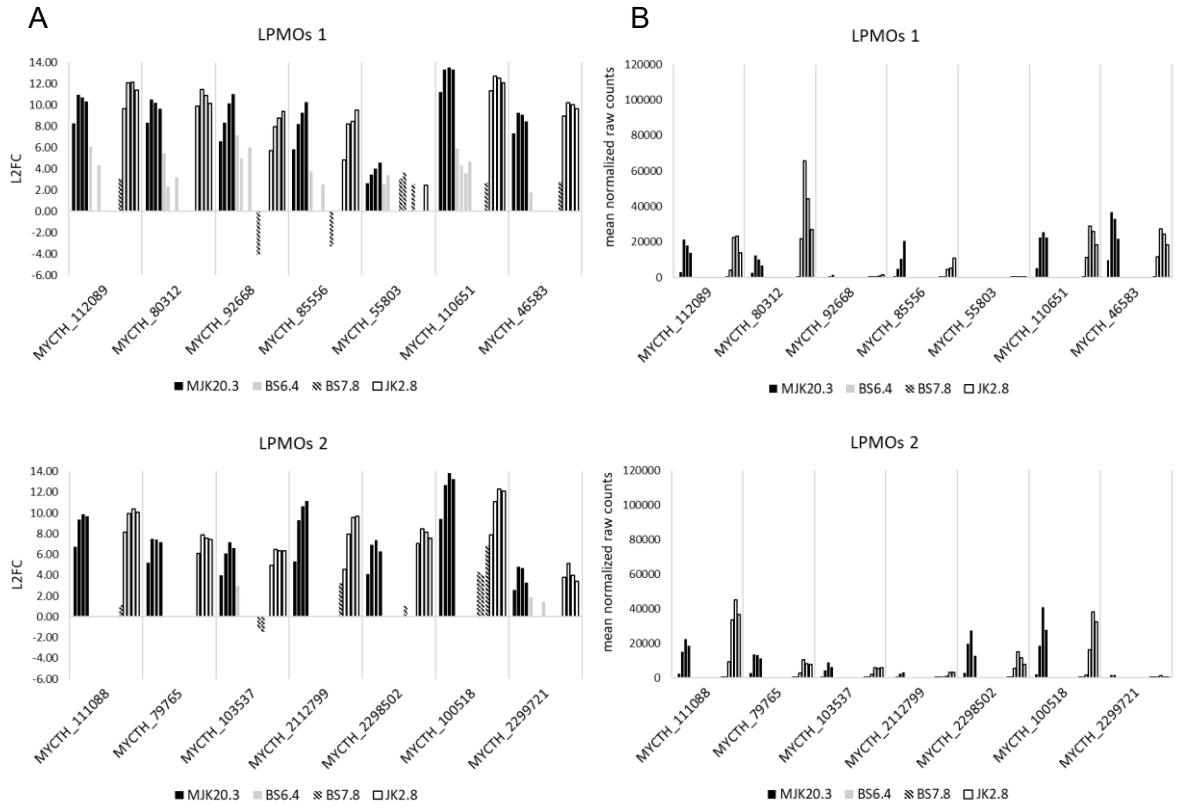

Figure 5.59: Differentially expressed predicted cellulase genes (part 2) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Among the genes encoding for predicted cellulases (Figures 5.58 and 5.59), all major predicted endoglucanase genes are upregulated in JK2.8, similar to MJK20.3. All these predicted endoglucanase genes (MYCTH_86753, MYCTH_76901, MYCTH_116384, and MYCTH_116157) have higher raw count values in JK2.8, especially MYCTH_86753. Due to the higher raw count values in SS, L2FC values for those genes are also very similar for JK2.8 and MJK20.3. These major predicted endoglucanase genes are all showing no expression in BS6.4 and BS7.8, except MYCTH_76901, where at t1 a raw count value, similar to MJK20.3 at this point in time can be observed in BS6.4. Among the genes encoding for predicted cellobiohydrolases, the four major genes are upregulated in JK2.8 as well. Two of those have higher raw count values in MJK20.3 (MYCTH_109566 and MYCTH_97137) but very similar L2FC values. The other ones have equal raw count values with higher L2FC values in JK2.8 (MYCTH_66729) and higher raw count values in JK2.8 (MYCTH_2303045) with equal L2FC values. All major predicted cellobiohydrolase genes are showing no expression in BS6.4 and BS7.8, except MYCTH_109566 (highest raw count value in MJK20.3), where residual expression can be observed regarding the raw count values. In the ß-glucosidase category, all genes have higher raw count values in JK2.8 with equal L2FC change values compared to MJK20.3. The only gene that is downregulated is MYCTH_80304, which has a higher raw count value and lower L2FC value in JK2.8 compared to MJK20.3. Among the upregulated ß-glucosidase genes, no expression can be observed in BS7.8.

Residual expression can be observed in BS6.4 for MYCTH_115968 (highest raw count value in MJK20.3) and MYCTH_62925. Among the genes encoding for predicted LPMOs, higher raw count values (also in SS) can be detected for three genes in JK2.8 compared to MJK20.3. In two of those genes, raw counts values are much higher compared to MJK20.3 (MYCTH_80312: ~3-4 times higher; MYCTH_111088: ~2 times higher). For five predicted LPMO genes, raw count values are higher in MJK20.3 with only two genes with much higher L2FC values (MYCTH_100518 and MYCTH_85556). The strains BS6.4 and BS7.8 show no expression of genes encoding for predicted LPMOs.

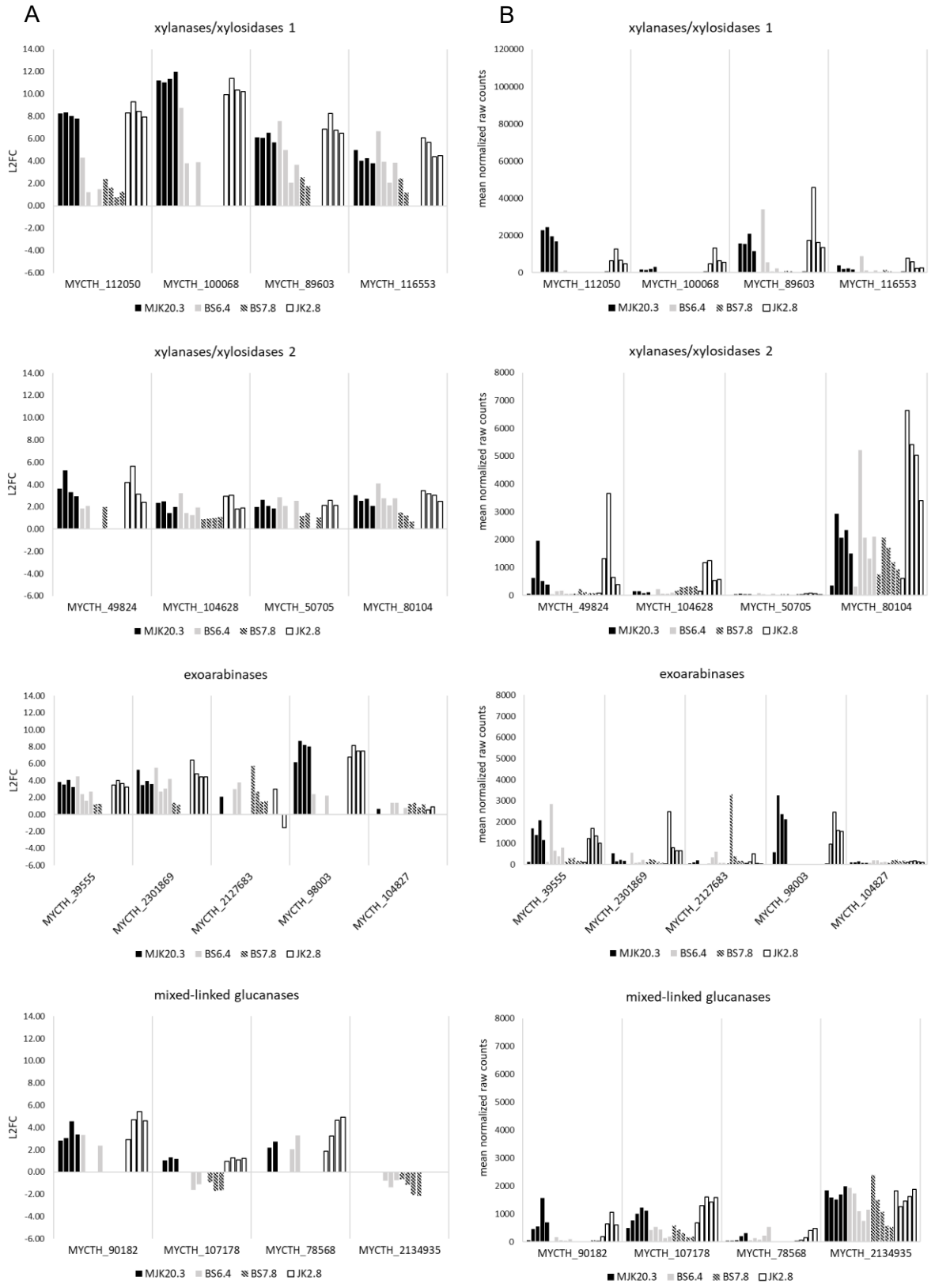

Figure 5.60: Differentially expressed predicted hemicellulase genes (part 1) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

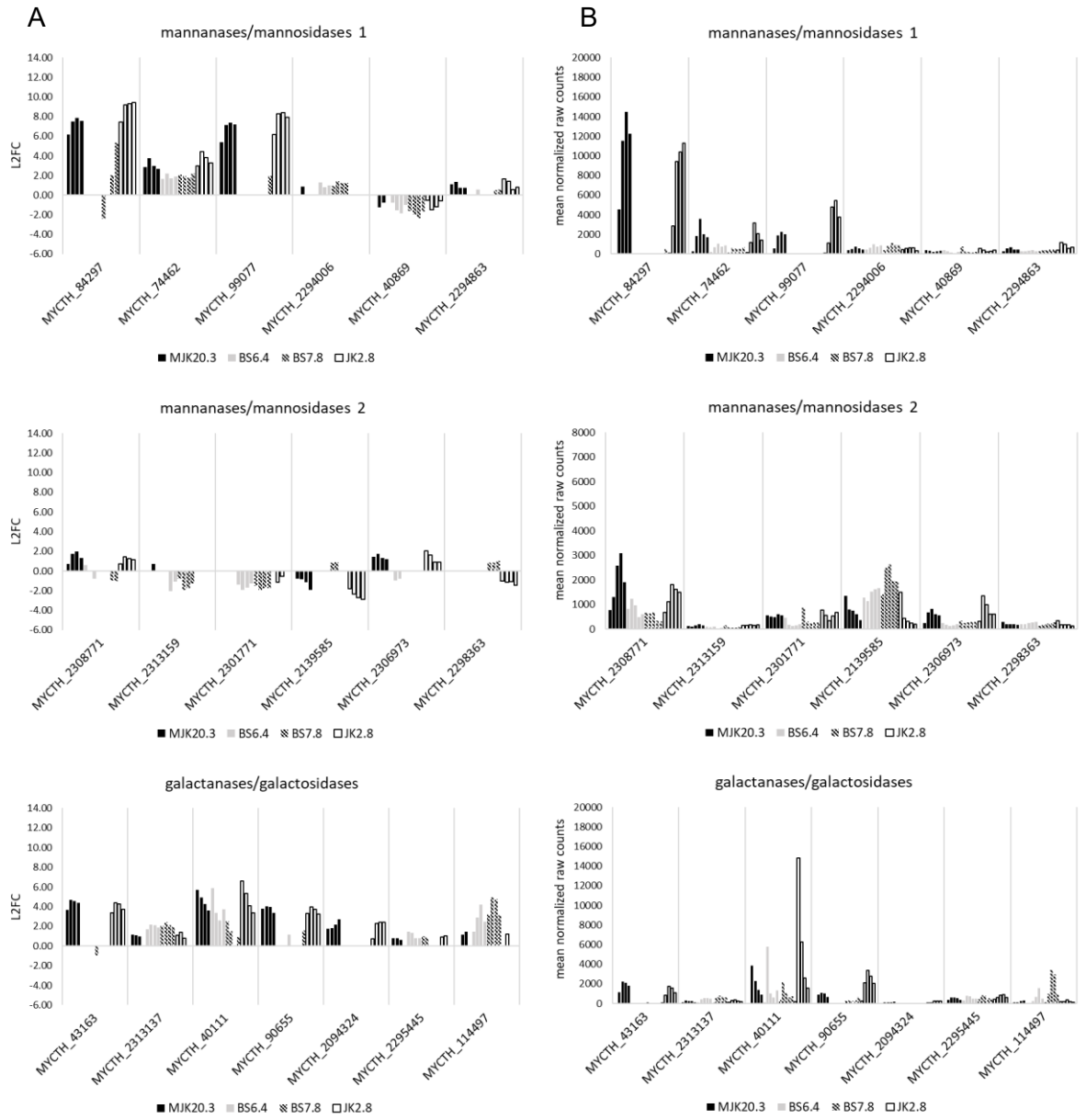

Figure 5.61: Differentially expressed predicted hemicellulase genes (part 2) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Among the genes encoding for predicted hemicellulases (Figures 5.60 and 5.61), for five of the six predicted xylanase/xylosidase genes with the highest raw count values, raw count values are much higher in JK2.8 compared to MJK20.3 (~2 times higher). Also, L2FC values are higher in JK2.8 for most of those genes. Almost no expression can be observed in BS7.8 except for gene MYCTH_80104, where expression levels are approximately similar to those in MJK20.3. In BS6.4 expression can only be observed for MYCTH_89603 (t1 higher, t2-t4 lower compared to MJK20.3), MYCTH_116553 (t1 higher, t2-t4 lower compared to MJK20.3), and MYCTH_80104 (t1 higher, t2-t4 equal compared to MJK20.3). Among the genes encoding for predicted exoarabinases only one gene has higher raw count values in

JK2.8 compared to MJK20.3 (MYCTH_2301869, ~2-3 times higher) as well as higher L2FC values. For two genes raw count values and L2FC values in MJK20.3 are higher. In BS6.4 expression can be observed in all genes similar to MJK20.3 except for MYCTH_39555 (t1 higher, t2-t4 lower compared to MJK20.3) and MYCTH_98003 (no expression). In BS7.8 observations are similar to BS6.4 with lower expression levels for MYCTH_39555. Expression of predicted mixed-linked glucanase genes is for the top genes similar in JK2.8 and MJK20.3 with slightly higher raw count values in JK2.8 for MYCTH_107178. In BS6.4 and BS7.8 expression levels of the top genes are equally lower compared to MJK20.3 with still very high raw count values for MYCTH_2134935 (highest raw count values in MJK20.3). For the top predicted mannanase/mannosidase genes, raw count values and L2FC values are only higher for MYCTH_99077 in JK2.8 compared to MJK20.3. All other top predicted mannanase/mannosidase genes have higher or equal expression levels in MJK20.3. For three of the five top predicted mannanase/mannosidase genes expression in BS6.4 and BS7.8 is detectable. For two of them, expression is lower compared to MJK20.3 but equal among BS6.4 and BS7.8. For MYCTH_2139585 expression is even higher in BS6.4 and BS7.8 (highest in BS7.8) compared to MJK20.3. The highest expressed predicted galactanase/galactosidase genes have higher expression levels in JK2.8 compared to MJK20.3. Expression can be observed for MYCTH_40111 in BS6.4 as well as BS7.8. For MYCTH_114497 expression is even higher in BS6.4 and BS7.8 (highest in BS7.8) compared to MJK20.3. For the other two top predicted galactanase/galactosidase genes, no expression ca be detected in BS6.4 and BS7.8.

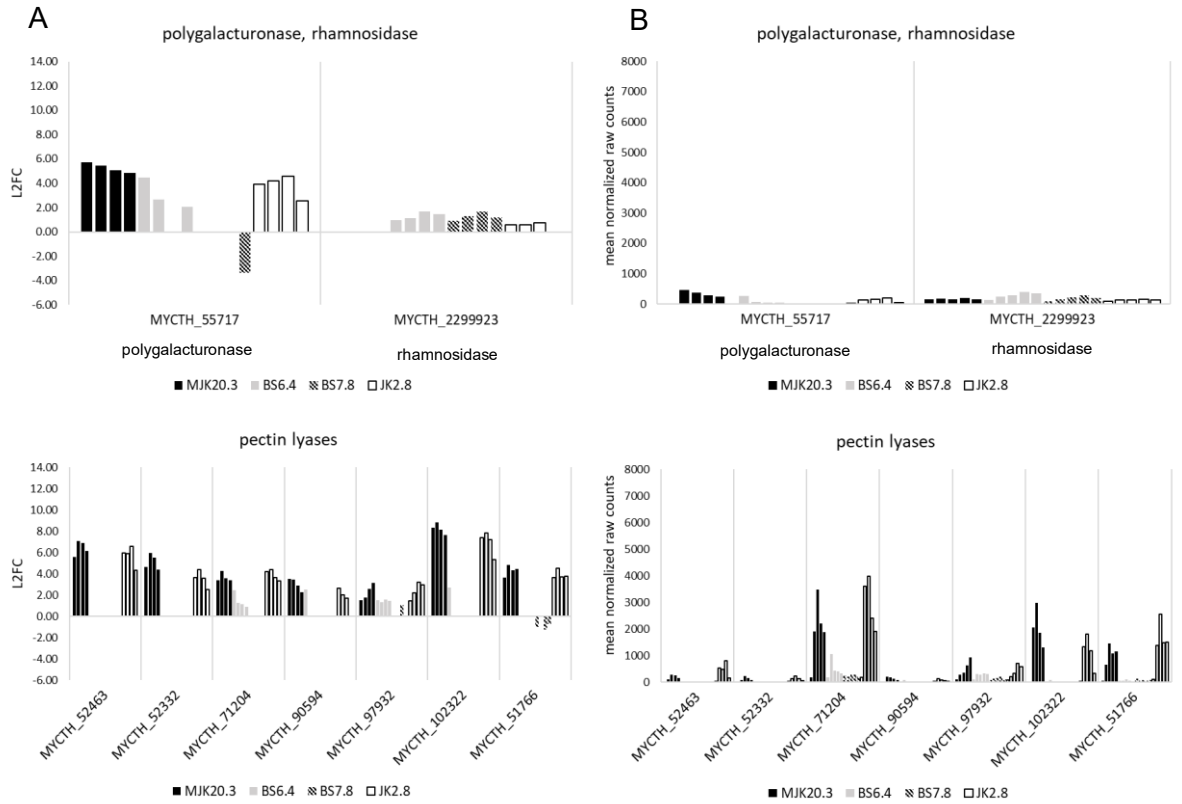

Figure 5.62: Differentially expressed predicted pectinase genes in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Among the genes encoding for predicted pectinases (Figure 5.62), only predicted pectin lyase genes show high expression levels. Here, raw count values for two of the three top predicted pectin lyase genes are higher in JK2.8, but L2FC values are higher or equal in MJK20.3. For MYCTH_102322 raw count values as well as L2FC values are lower in JK2.8 compared to MJK20.3. Only for one of the top predicted pectin lyase genes, MYCTH_71204 (highest raw count values in MJK20.3), expression in BS6.4 as well as BS7.8 can be detected.

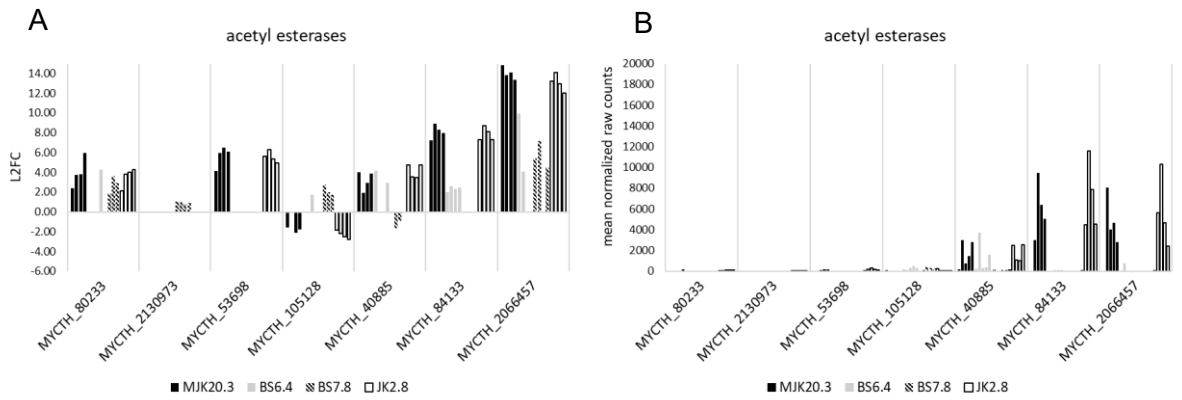

Figure 5.63: Differentially expressed predicted esterase genes in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Among the predicted esterase genes (Figure 5.63), only predicted acetyl esterase genes show high expression levels. For two of the three top genes raw count values are higher in JK2.8 compared to MJK20.3 while L2FC values are equal or lower. Expression in BS6.4 can only be observed for MYCTH_40885 (t1 higher, t2-t4 lower compared to MJK20.3). In BS7.8 no expression can be observed.

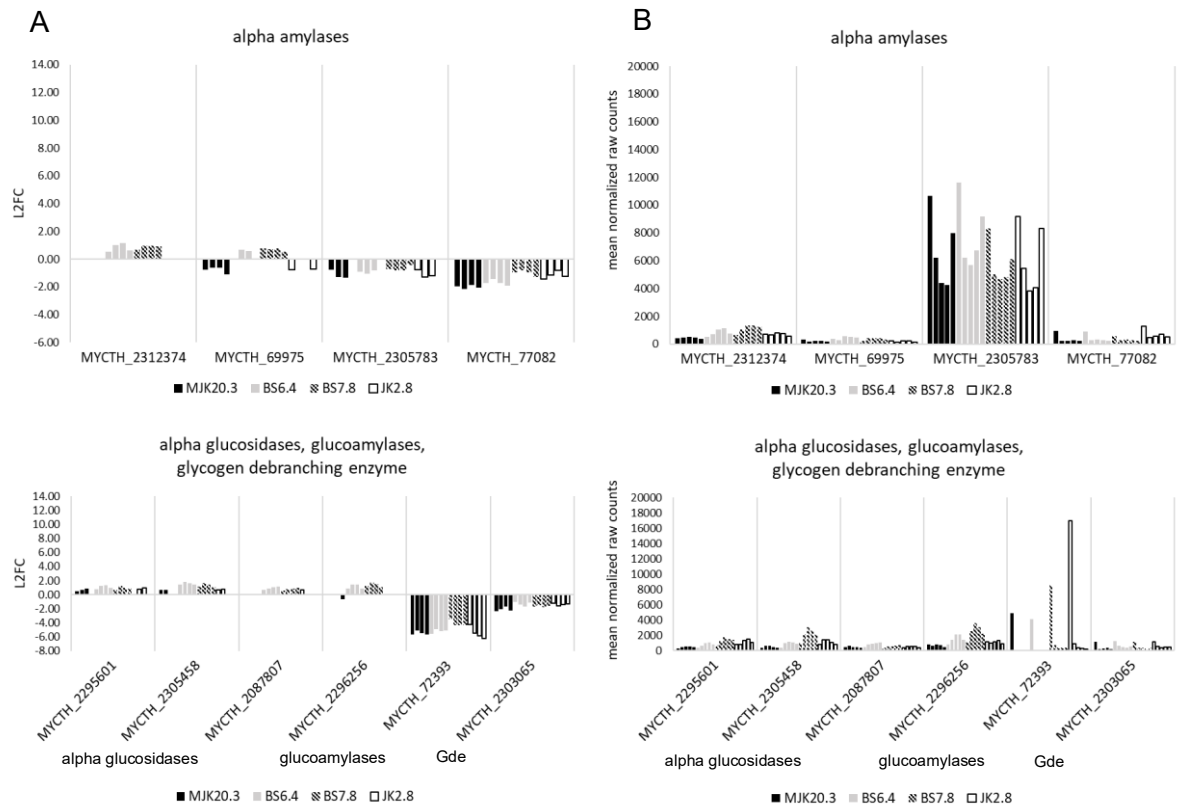

Figure 5.64: Differentially expressed predicted starch metabolism genes in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty). Gde= glycogen debranching enzyme.

Within the starch metabolism category (Figure 5.64), almost all genes are expressed at an equal level with only minor differences. It is worth to mention that in many of the genes, expression levels are highest in BS7.8, but not for MYCTH_2305783 which has the highest raw count value in MJK20.3. One gene that is extremely downregulated in all strains is MYCTH_72393, where very high raw count values in SS can be observed (highest in JK2.8). All other genes have either low expression levels or are equally expressed in all strains.

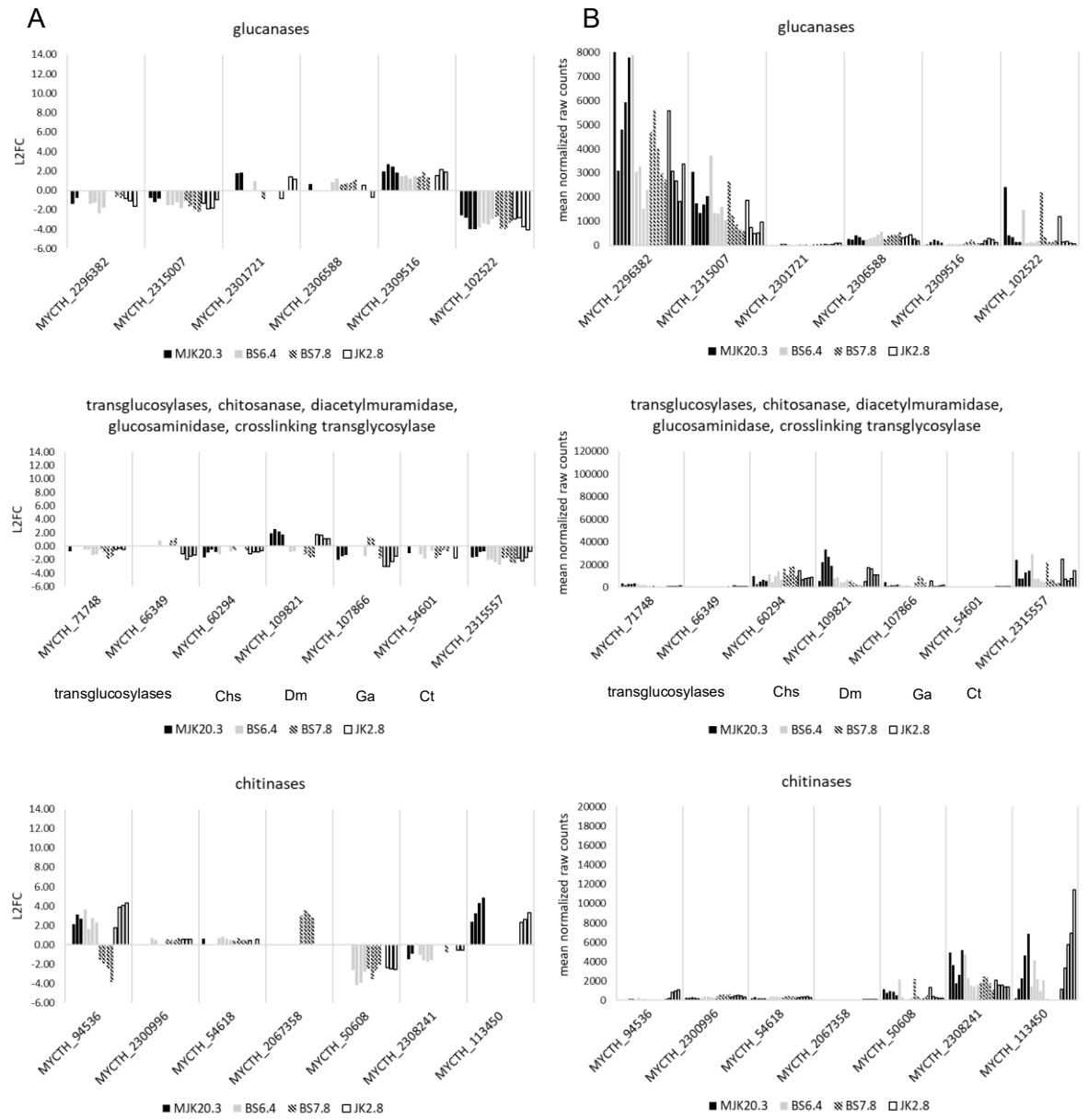

Figure 5.65: Differentially expressed predicted cell wall remodeling genes in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty). Tg= transglucosylase, Chs= chitosanase, Dm= diacetylmuramidase, Ga= glucosaminidase, Ct= crosslinking transglycosidase.

Regarding predicted cell wall remodeling genes (Figure 5.65), almost all genes are equally expressed in all strains. Exceptions are MYCTH_2296382 (predicted glucanase), which has higher expression levels in MJK20.3 especially in SS and at t4, MYCTH_60294 (predicted transglucosylase), which has higher expression levels in BS6.4 and BS7.8 (highest in BS7.8), MYCTH_109821 (predicted chitosanase), which has much higher expression levels in MJK20.3 and JK2.8 (highest in MJK20.3), MYCTH_107866 (predicted glucosaminidase), which has a much higher expression level in BS7.8, MYCTH_2308241 (predicted chitinase), which has higher expression levels in MJK20.3 especially in SS

and at t4 as well as MYCTH_113450 (predicted chitinase), which has higher raw count values in JK2.8 and shows no expression in BS7.8 and a reduced expression in BS6.4.

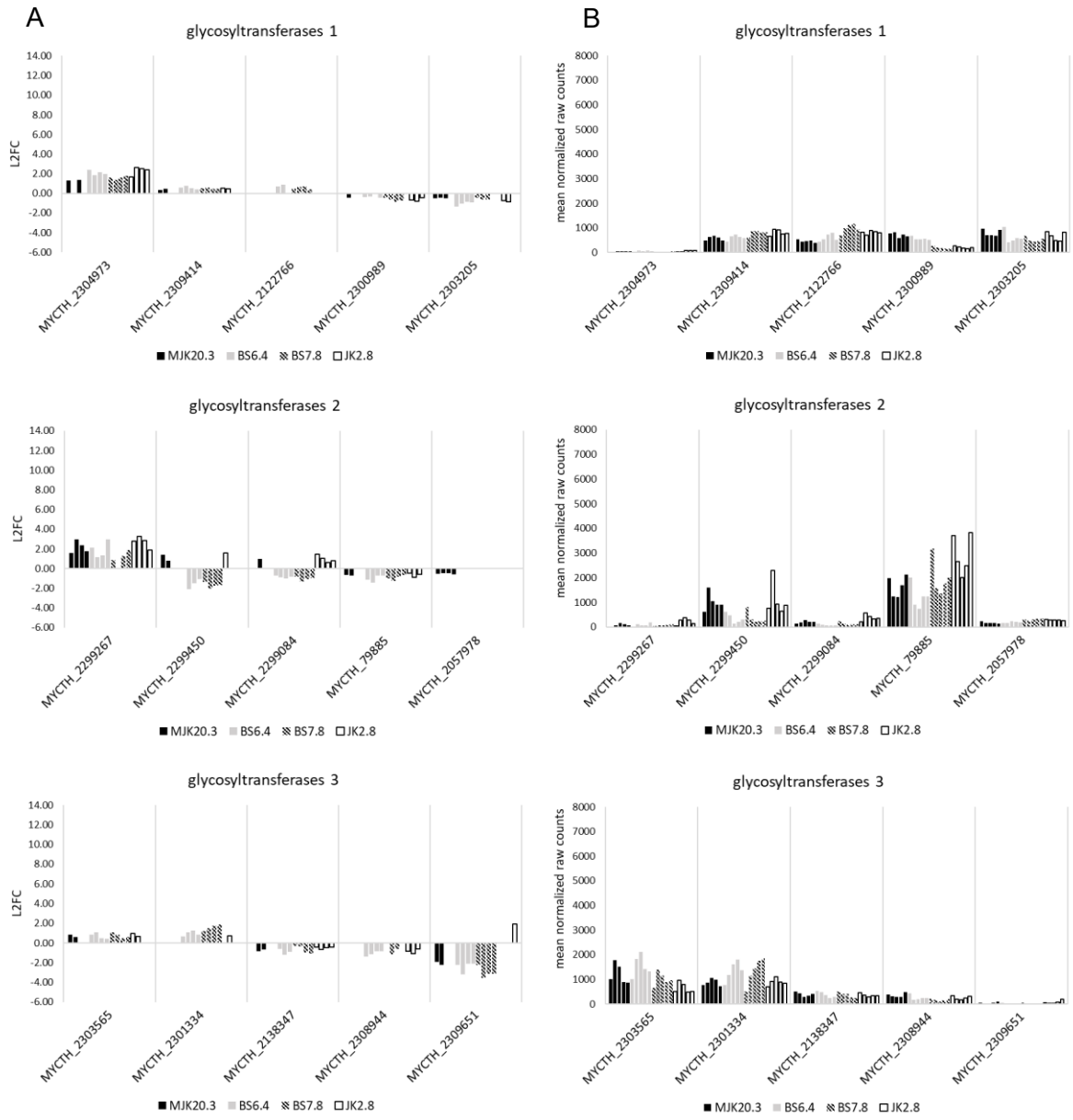

Figure 5.66: Differentially expressed predicted glycosyltransferase genes (part 1) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

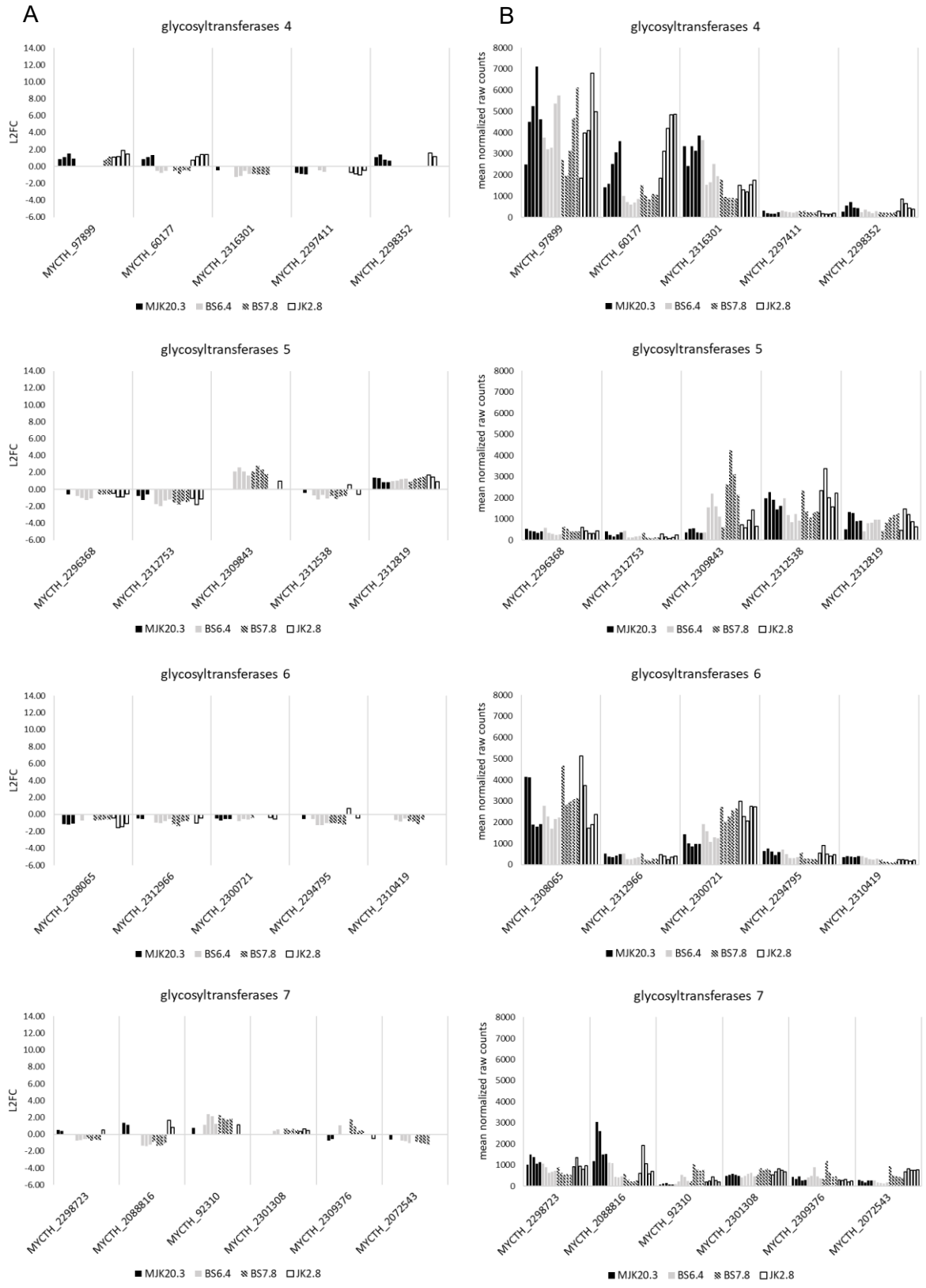

Figure 5.67: Differentially expressed predicted glycosyltransferase genes (part 2) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

For most of the predicted glycosyltransferase genes (Figures 5.66 and 5.67), expression is similar among the different strains. Due to this fact and the high number of differentially expressed predicted glycosyltransferase genes a detailed analysis was not carried out. An observable trend is that expression in MJK20.3 and JK2.8 as well as BS6.4 and BS7.8 behaves very similar, meaning when expression values are higher in MJK20.3 also expression values in JK2.8 are higher compared to the other two strains and the other way around.

In the “other” category (Figures 5.68-5.70), some interesting, upregulated genes with an extremely high expression level and some extremely downregulated genes could already be identified in a previous chapter. Genes belonging to this category have already been categorized in a previous chapter (see Table 5.10 chapter 5.3.1.3) and are described in the following according to this characterization. Due to the high number of differentially expressed genes in the “other” category only the most striking differences regarding single genes between the strains will be discussed.

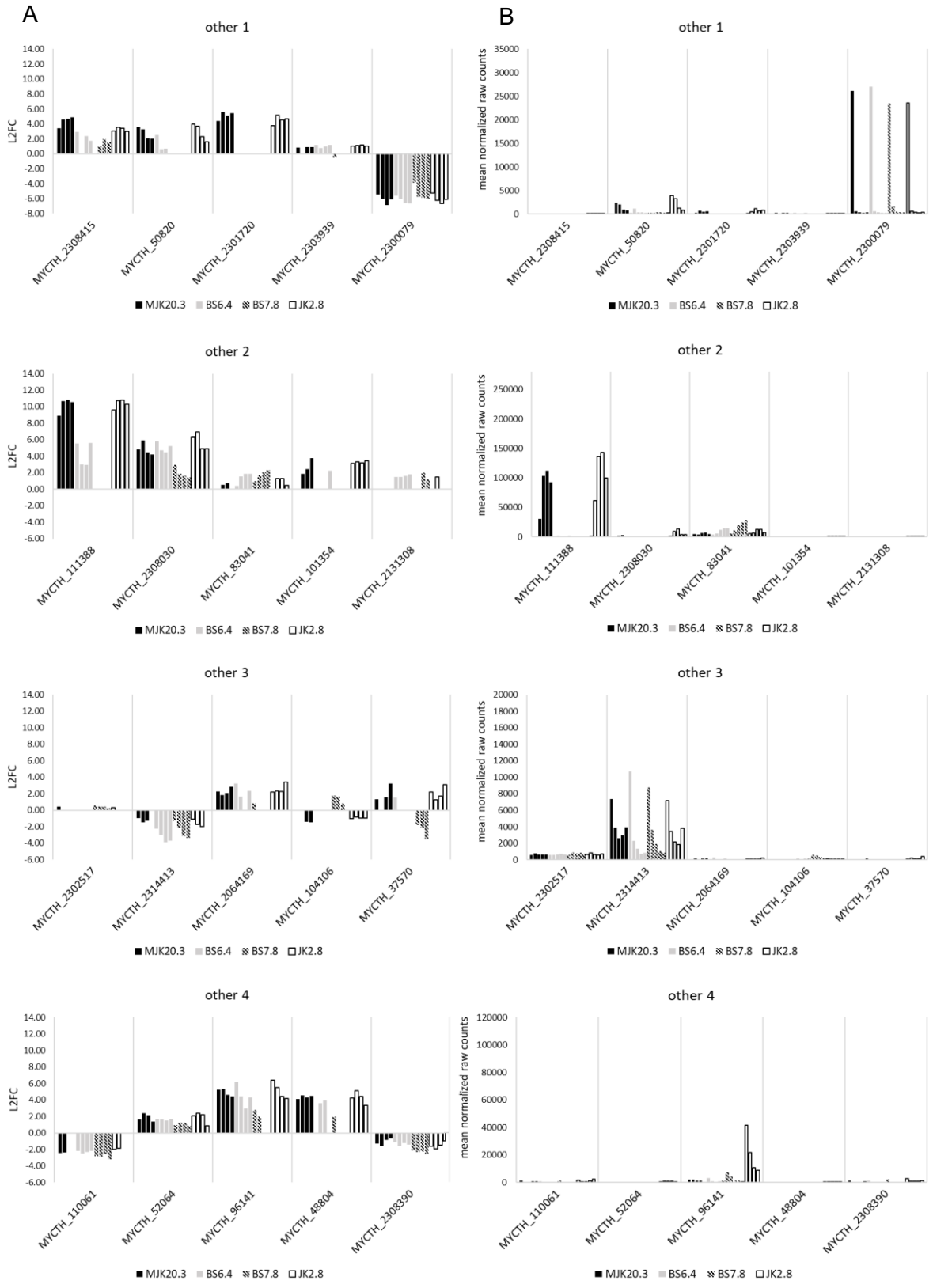

Figure 5.68: Differentially expressed "other" genes (part 1) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

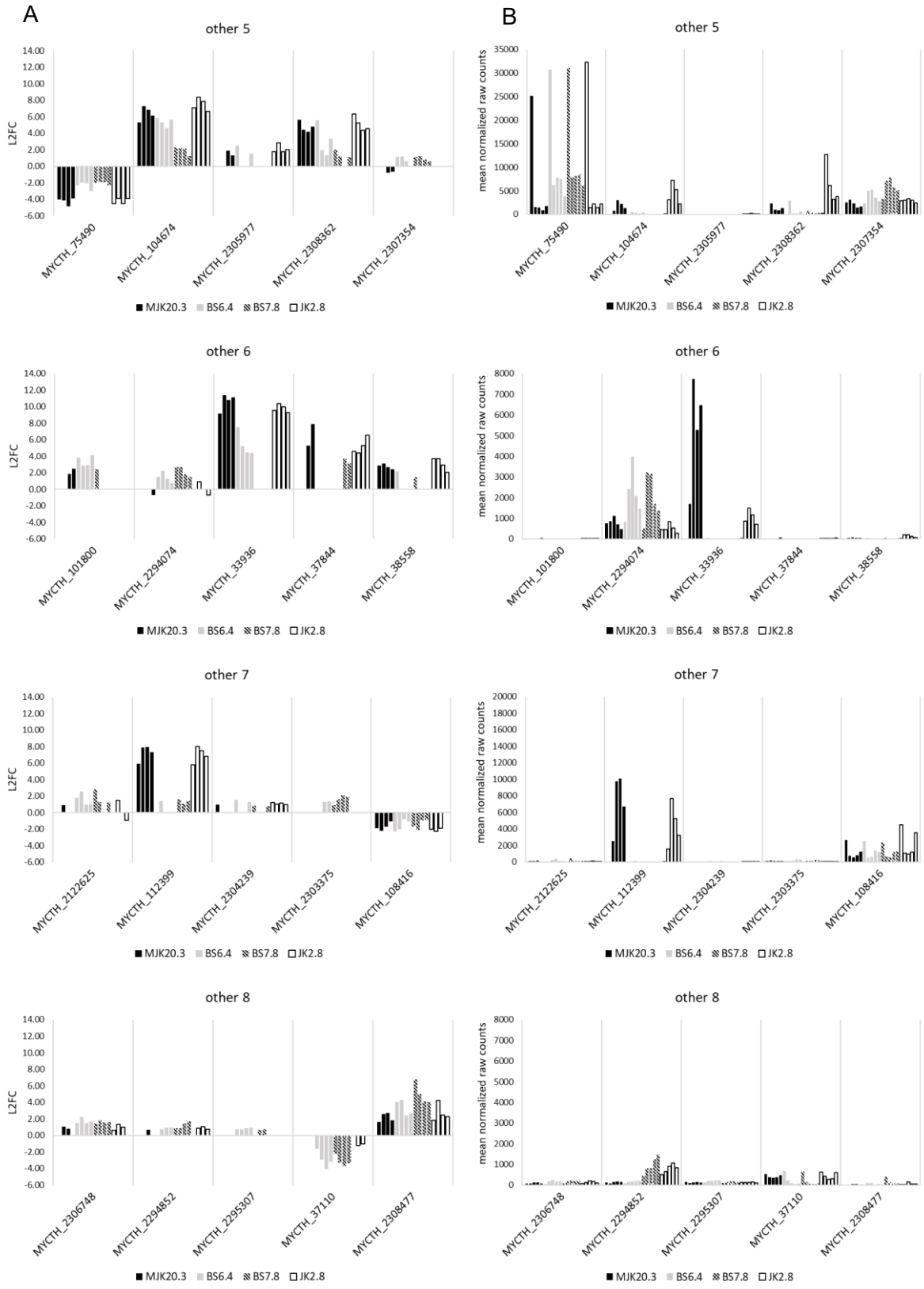

Figure 5.69: Differentially expressed "other" genes (part 2) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

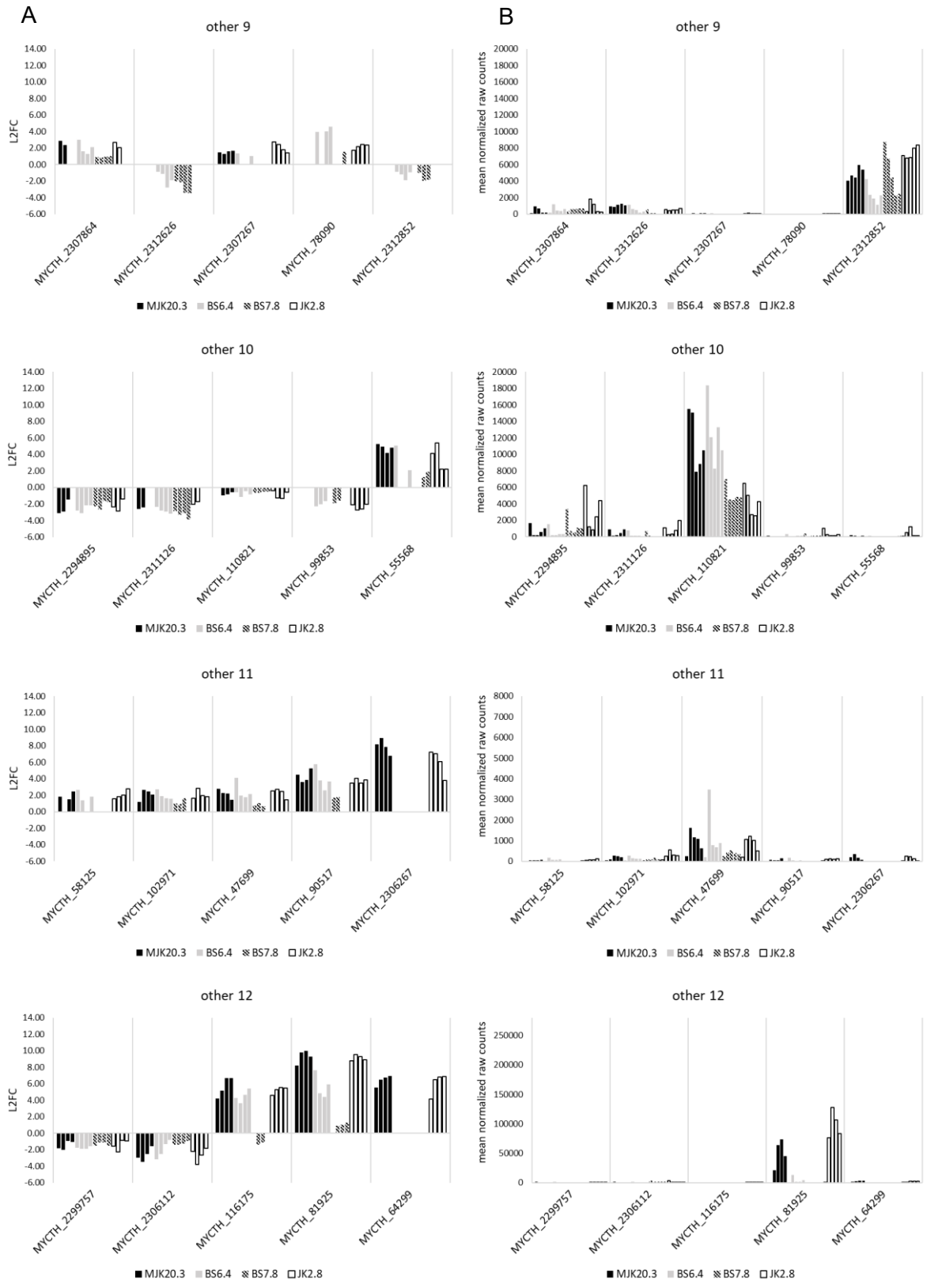

Figure 5.70: Differentially expressed "other" genes (part 3) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

MYCTH_111388 (predicted cellobiose dehydrogenase), has extremely high raw count and L2FC values in MJK20.3 and JK2.8. Raw count values in JK2.8 are higher, and no expression can be detected in BS6.4 and BS7.8. For MYCTH_83041 (predicted hemicellulase), the highest raw count values can be detected in BS7.8. For another interesting gene, MYCTH_96141 (predicted hemicellulase), the expression level in JK2.8 is extremely high compared to all other strains (~7-8 times higher), where much lower expression levels can be observed. The downregulation of MYCTH_75490 (predicted trehalase) is much weaker in the strains BS6.4 and BS7.8, compared to the other strains. For MYCTH_104674 (predicted hemicellulase) expression can only be observed in MJK20.3 and JK2.8 (higher expression level in JK2.8). Regarding MYCTH_2308362 (predicted hemicellulase), the expression level in JK2.8 is extremely high compared to all other strains (~5-7 times higher), where much lower expression levels can be observed. MYCTH_2294074 (predicted pectinase) is expressed higher in BS6.4 and BS7.8 compared to the other strains, whereas MYCTH_33936 and MYCTH_112399 (predicted cellulases) expression cannot be detected in BS6.4 and BS7.8 but MJK20.3 and JK2.8 (much higher in JK2.8). MYCTH_2312852 (predicted chitinase), which has high raw count values in MJK20.3 and JK2.8 (higher raw count values in JK2.8) is only downregulated in BS6.4 and BS7.8. Finally the MYCTH_2294895 (hypothetical) has the highest raw count values in JK2.8, MYCTH_110821 (hypothetical) has high remaining raw count values in all strains although it is slightly downregulated (higher in MJK20.3 and BS6.4) and MYCTH_81925 (predicted cellobiose dehydrogenase) has higher raw count values and equal L2FC values in JK2.8 compared to MJK20.3. A low level of expression for MYCTH_81925 is only detectable in BS6.4 but not in BS7.8. Taken together, predicted cellobiose dehydrogenase genes and some predicted hemicellulase genes were stronger expressed in JK2.8 compared to MJK20.3. For those genes and some other predicted cellulase genes that are stronger expressed in MJK20.3 compared to JK2.8, no or very weak expression was observed in BS6.4 and BS7.8.

In summary, expression of predicted CAZY genes, especially cellulase, hemicellulase, and pectinase genes, is extremely reduced or even not detectable in BS6.4 and BS7.8 in comparison to MJK20.3. This effect is much stronger in BS7.8, whereas in BS6.4 residual expression of some genes, especially predicted ß-glucanase genes can be observed. In JK2.8, in contrary, expression all above-mentioned predicted CAZY genes is mostly higher while the general expression pattern is almost identical compared to MJK20.3.

5.3.2.4 Differential expression of transcription factors of plant biomass degradation

The purpose of the investigation of the expression of transcription factors of plant biomass degradation has already been mentioned previously (see chapter 5.3.1.4). An overview of predicted *T. thermophilus* transcription factors was provided in this chapter as well.

The total number of predicted transcription factor genes that were differentially expressed in the strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 at the different points in time after the cellulose spike, is shown in Figure 5.71.

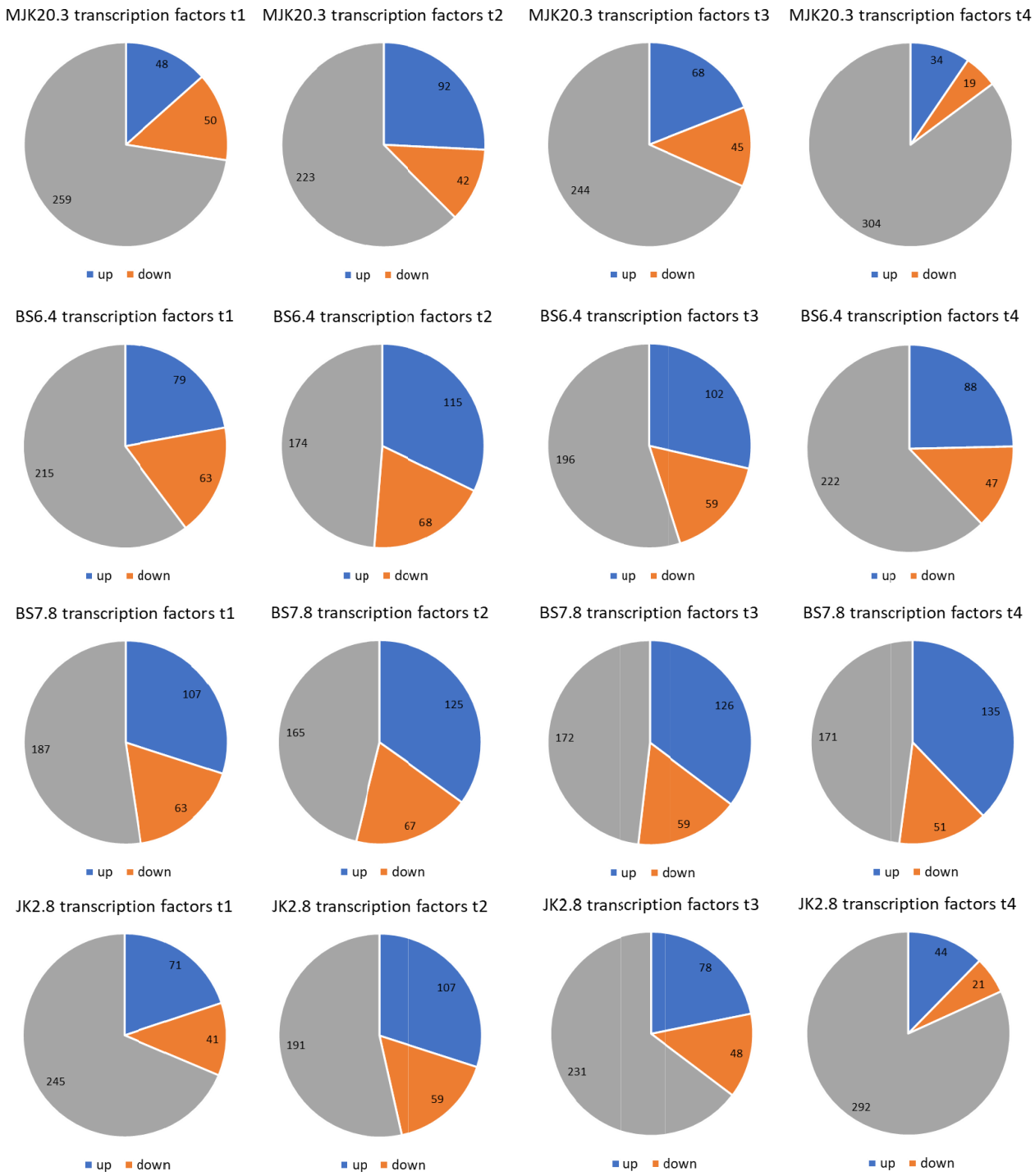

Figure 5.71: Number of differentially expressed predicted transcription factor genes of strains MJK20.3, BS6.4, BS7.8, and JK2.8 at different points in time after the cellulose spike. Shown are the numbers of up- (blue) and downregulated (orange) genes as well as genes with no differential expression (grey) at the points in time: 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after cellulose spike (t4) compared to the respective steady state condition.

Here, a higher number of upregulated genes in JK2.8 at all points in time after the spike in comparison to MJK20.3 can be detected. The number of downregulated genes is at t1 lower and at all other points in time higher in JK2.8 in comparison to MJK20.3. In BS6.4 and BS7.8 a higher number of upregulated genes and a higher number of downregulated genes can be observed at all points in time in comparison to MJK20.3. The number of upregulated genes at all points in time is the highest in BS7.8 and the

number of downregulated genes at all points in time is very similar between BS6.4 and BS7.8. For MJK20.3 and JK2.8 the number of differentially expressed genes peaks at t2 and decreases toward t4. The same trend can be observed for BS6.4 but with much higher numbers of differentially expressed genes at all points in time. In BS7.8 instead, a constant number of differentially expressed genes at all points in time can be observed. In total up to approximately 40 % of the predicted transcription factor genes are differentially expressed in MJK20.3, 50 % in BS6.4, 55 % in BS7.8, and 45 % in JK2.8. The relation between up- and downregulated genes ranges from 1:1 to 2:1 for all strains.

To get a detailed overview of the expression profiles of strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 regarding predicted transcription factor gene expression, a heatmap (Figure 5.72) was created, showing the differential expression levels of the genes belonging to the different classes of putative transcription factors among the four strains. The expression profile of strains JK2.8 and MJK20.3 is almost identical but a slightly stronger upregulation of single genes is observable in JK2.8 as well as a higher total number of differentially expressed predicted transcription factor genes at the respective points in time as already seen in Figure 5.71. At t4 a lower number of predicted transcription factor genes is differentially expressed in both strains compared to the earlier points in time. The expression profile of strains BS6.4 and BS7.8 is also very similar but completely different compared to JK2.8 and MJK20.3. L2FC values are much lower compared to genes that are differentially expressed in MJK20.3 and JK2.8 but much more genes are differentially expressed even at t4. Generally, the genes with the highest/lowest differential expression values are evenly distributed in the different classes.

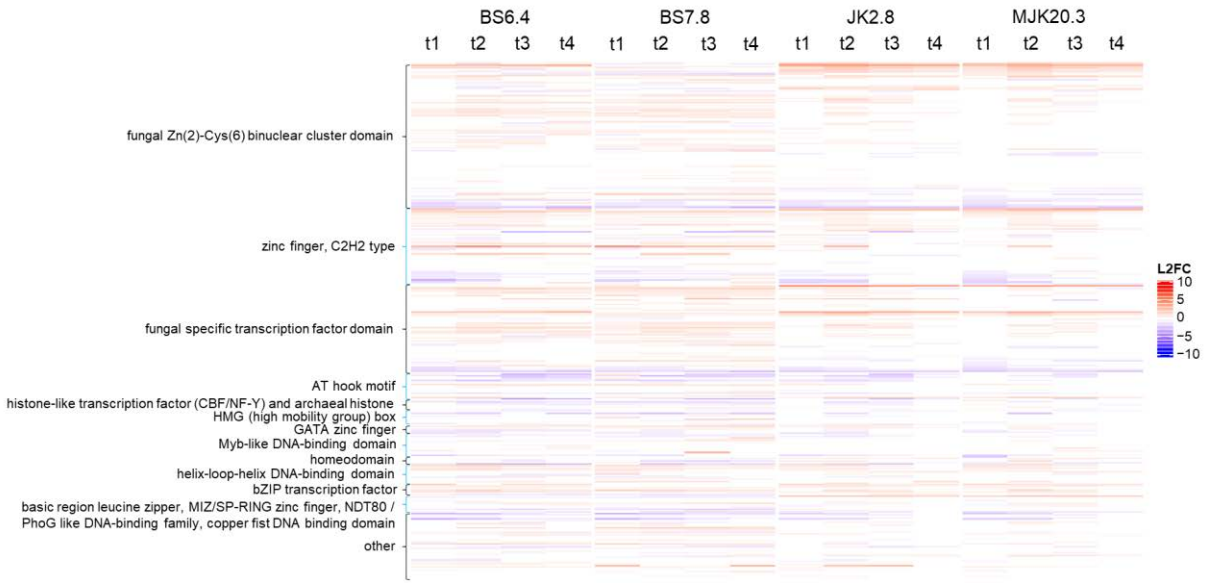

Figure 5.72: Heatmap with differential expression values of all predicted transcription factor genes in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3. Shown are the log2 fold change values (L2FC) of single predicted transcription factor genes belonging to the different classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

To narrow down the number of differentially expressed genes and filter out the most important differentially expressed genes, the t1-t4 intersection was analyzed next. A gene was included, when differential expression over all points in time after the cellulose spike in at least one of the strains was

detected. In Table 5.23, the number of genes that are differentially expressed at all points in time after the spike are shown for the two strains. A further heatmap (Figure 5.73) visualizes the differential expression levels of the respective genes that belong to the t1-t4 intersection area.

Table 5.23: Number of genes belonging to different transcription factor classes that are differentially expressed at all points in time after the spike in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3.

		MJK20.3		**BS6.4**		**BS7.8**		**JK2.8**	
class	genes	t1-t4 up	t1-4 down	t1-t4 up	t1-4 down	t1-t4 up	t1-4 down	t1-t4 up	t1-4 down
fungal Zn(2)-Cys(6) binuclear cluster domain	101	8	4	11	3	23	5	13	4
zinc finger, C2H2 type	53	3	1	8	1	8	2	5	0
fungal specific transcription factor domain	61	4	3	12	2	22	2	6	2
AT hook motif	18	1	0	1	4	2	4	1	0
histone-like transcription factor (CBF/NF-Y) and archaeal histone	7	0	0	0	1	1	3	0	0
HMG (high mobility group) box	10	0	0	0	0	0	1	0	0
GATA zinc finger	6	0	0	2	1	1	0	0	1
Myb-like DNA-binding domain	17	0	0	1	1	1	1	0	0
homeodomain	5	0	0	0	1	0	0	1	0
helix-loop-helix DNA-binding domain	13	1	0	3	0	3	0	0	0
bZIP transcription factor	8	2	0	4	1	2	1	3	0
basic region leucine zipper	4	1	0	0	1	0	1	1	0
MIZ/SP-RING zinc finger	3	0	0	0	0	1	0	0	0
NDT80 / PhoG like DNA-binding family	3	0	0	0	1	0	1	0	0
copper fist DNA binding domain	2	0	1	0	0	1	0	0	0
other	46	0	0	4	4	7	5	0	0
total	357	20	9	46	21	72	26	30	7

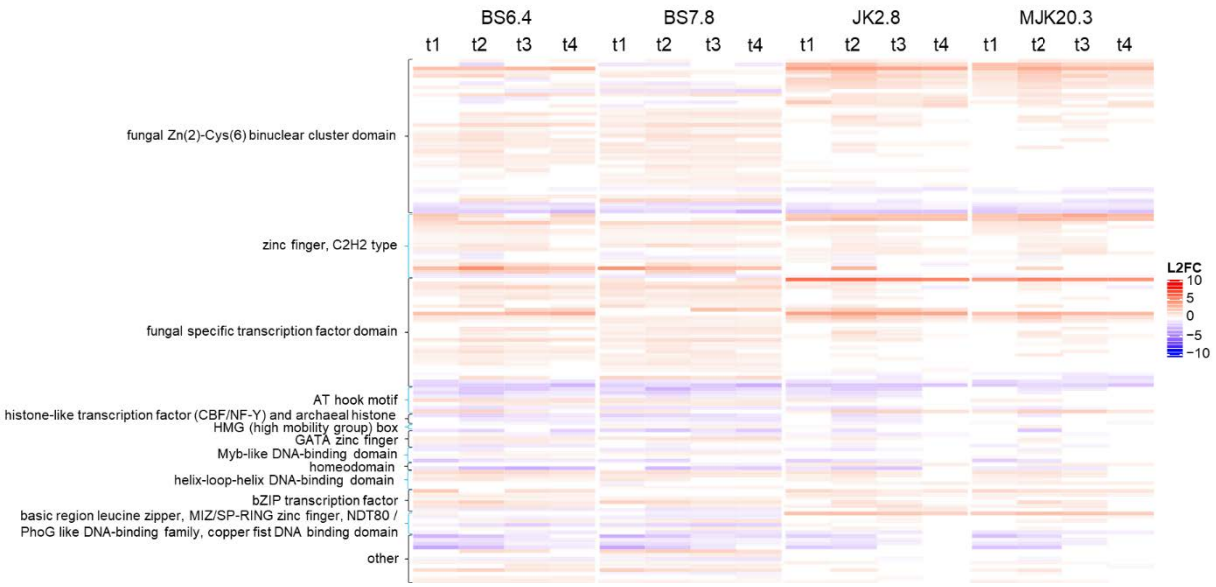

Figure 5.73: Heatmap with expression values of predicted transcription factor genes that are differentially expressed at all points in time after the cellulose spike in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3. Shown are the log2 fold change values (L2FC) of single predicted transcription factor genes belonging to different classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes.

The heatmap of the t1-t4 intersection genes confirms the results of the previous graphs: the expression pattern between strains MJK20.3 and JK2.8 on the one hand and BS6.4 and BS7.8 on the other hand is almost identical with some single genes having a stronger up- or downregulation in one of the strains. Generally, it becomes clear, that the most and strongest differential expression derives from genes belonging to the classes fungal Zn(2)-Cys(6) binuclear cluster domain, zinc finger, C2H2 type, and fungal specific transcription factor, which could not be detected in the previous heatmap. Furthermore, it can be observed, that in BS6.4 as well as BS7.8 a much higher number of especially upregulated genes with low differential expression values can be found mainly in the fungal Zn(2)-Cys(6) binuclear cluster domain, zinc finger, C2H2 type, and fungal specific transcription factor classes.
To get an idea of the possible function of the predicted transcription factors, the protein sequences were blasted as previously described (see chapter 5.3.1.4). Additional interesting genes that first appeared during the analysis of the t1-t4 intersection genes in BS6.4, BS7.8, and JK2.8 are shown in Table 5.24.

Table 5.24: Predicted transcription factor genes that belong to the t1-t4 intersection area in strains BS6.4, BS7.8, and JK2.8 that have not been described yet in Table 5.13. Shown are the descriptions of the best protein blast hits (lowest E-values), performed using the NCBI database and the general expression trend (up- or downregulation at least at one point in time after the spike or no differential expression (no DE)) in the reference strain MJK20.3.

upregulated	downregulated	no DE	description best blast hits
MYCTH_2086212			transcription initiation factor IIA subunit
MYCTH_2132441			McmA/1
MYCTH_2296662			protein Max
MYCTH_2298970			adhesion and hyphal regulator1/Moc3/ sterol uptake protein 2
MYCTH_2305522			LepE/activator of stress/Acu15
MYCTH_2310172			hypothetical
MYCTH_2312477			regulator Rpn4/proteasome 26S regulatory subunit like
MYCTH_2315566			cross pathway control protein 1
MYCTH_57777			regulatory protein Cys3
	MYCTH_2300935		cutinase transcription factor 1 beta
	MYCTH_2301174		multicopy suppressor protein 2
	MYCTH_2314437		multiprotein-bridging factor 1
		MYCTH_111741	endonuclease Nob1
		MYCTH_2296492	Clr4
		MYCTH_2298696	Rce1
		MYCTH_2298994	NirA/Nit4
		MYCTH_2305050	forkhead protein Sep1/2/ meiosis specific transcription factor Mei4
		MYCTH_2313746	regulator Nrg1/conidial separation

To investigate and compare gene expression in detail, graphs showing L2FC values as well as graphs showing the mean of the normalized raw count values were created as previously (see chapter 5.3.1.2) described (Figures 5.74-5.84). The expression of predicted transcription factor genes in MJK20.3, especially the top genes, have already been discussed in a previous chapter. Therefore, focus of the following is on describing the differences of strains BS6.4, BS7.8, and JK2.8 to MJK20.3, especially regarding the top predicted transcription factor genes of each class. If not mentioned specifically, expression of genes is regarded as similar.

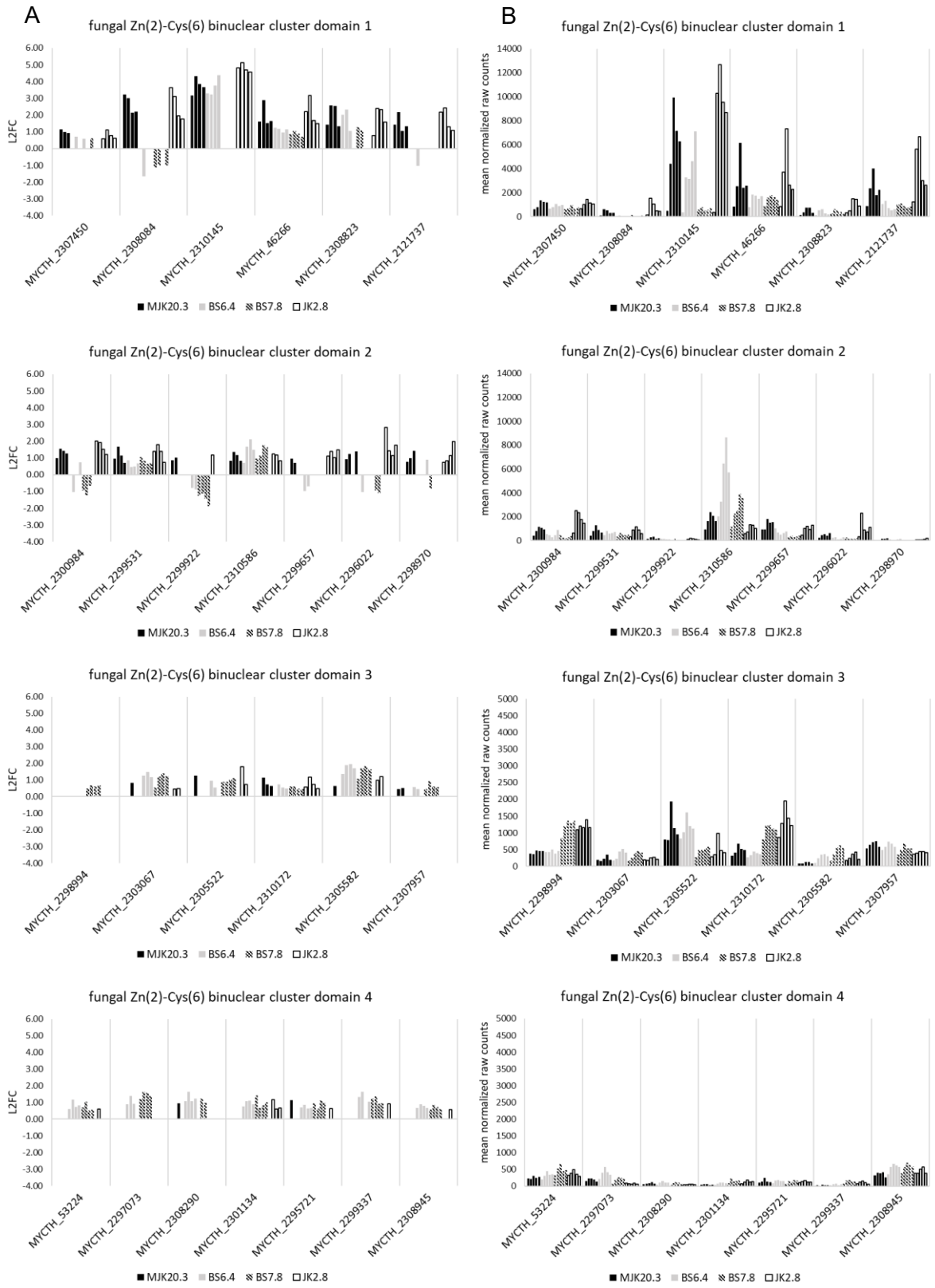

Figure 5.74: Differentially expressed genes encoding for predicted transcription factors with fungal Zn(2)-Cys(6) binuclear cluster domain (part 1) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

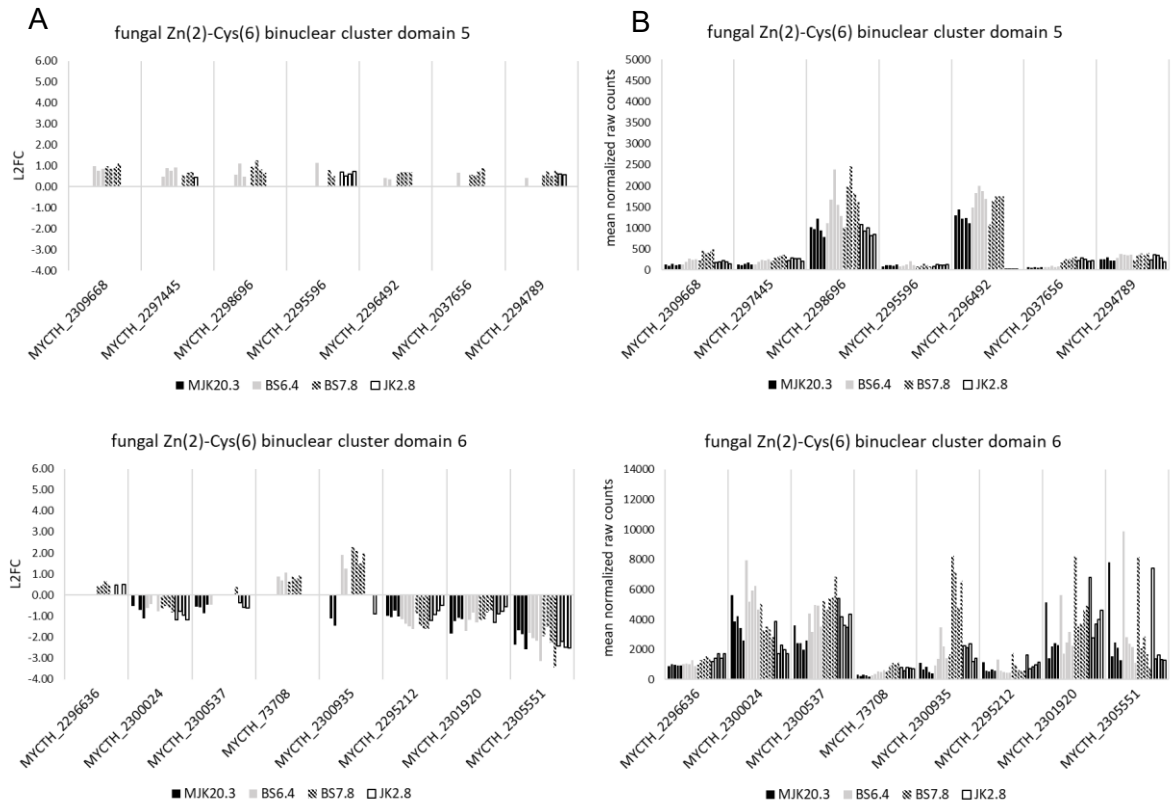

Figure 5.75: Differentially expressed genes encoding for predicted transcription factors with fungal Zn(2)-Cys(6) binuclear cluster domain (part 2) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

The genes encoding predicted transcription factors with a fungal Zn(2)-Cys(6) binuclear cluster domain (Figures 5.74 and 5.75) are mainly upregulated. MYCTH_2310145 (Xyr1) shows almost no expression in BS7.8 whereas in the other strains a strong expression can be observed (highest in JK2.8, lowest in BS6.4). MYCTH_46266 (GaaR/Pdr2) and MYCTH_2121737 (sterol regulatory Cys6/Ara1) are only showing strong expression in MJK20.3 and JK2.8 (highest in JK2.8). For MYCTH_2300984 (glucose transport Rgt1/nitrate assimilation NirA) and MYCTH_2298970 (adhesion and hyphal regulator1/Moc3/sterol uptake protein 2) high expression levels can be detected in JK2.8 only. For MYCTH_2310586 (hypothetical) the strongest upregulation and expression can be observed in BS7.8 and BS6.4 (BS7.8 highest, JK2.8 lowest). The predicted transcription factor genes MYCTH_2298994 (NirA/Nit4) and MYCTH_2310172 (hypothetical) are showing high expression in BS7.8 and JK2.8 only and for MYCTH_2305522 (LepE/activator of stress/Acu15) only high expression levels in MJK20.3 and BS6.4 can be observed. Finally, MYCTH_2298696 (Rce1) and MYCTH_2296492 (Clr4) are stronger expressed in BS6.4 and BS7.8 compared to the other strains. MYCTH_2300935 (cutinase transcription factor 1 beta) is only highly expressed in BS7.8.

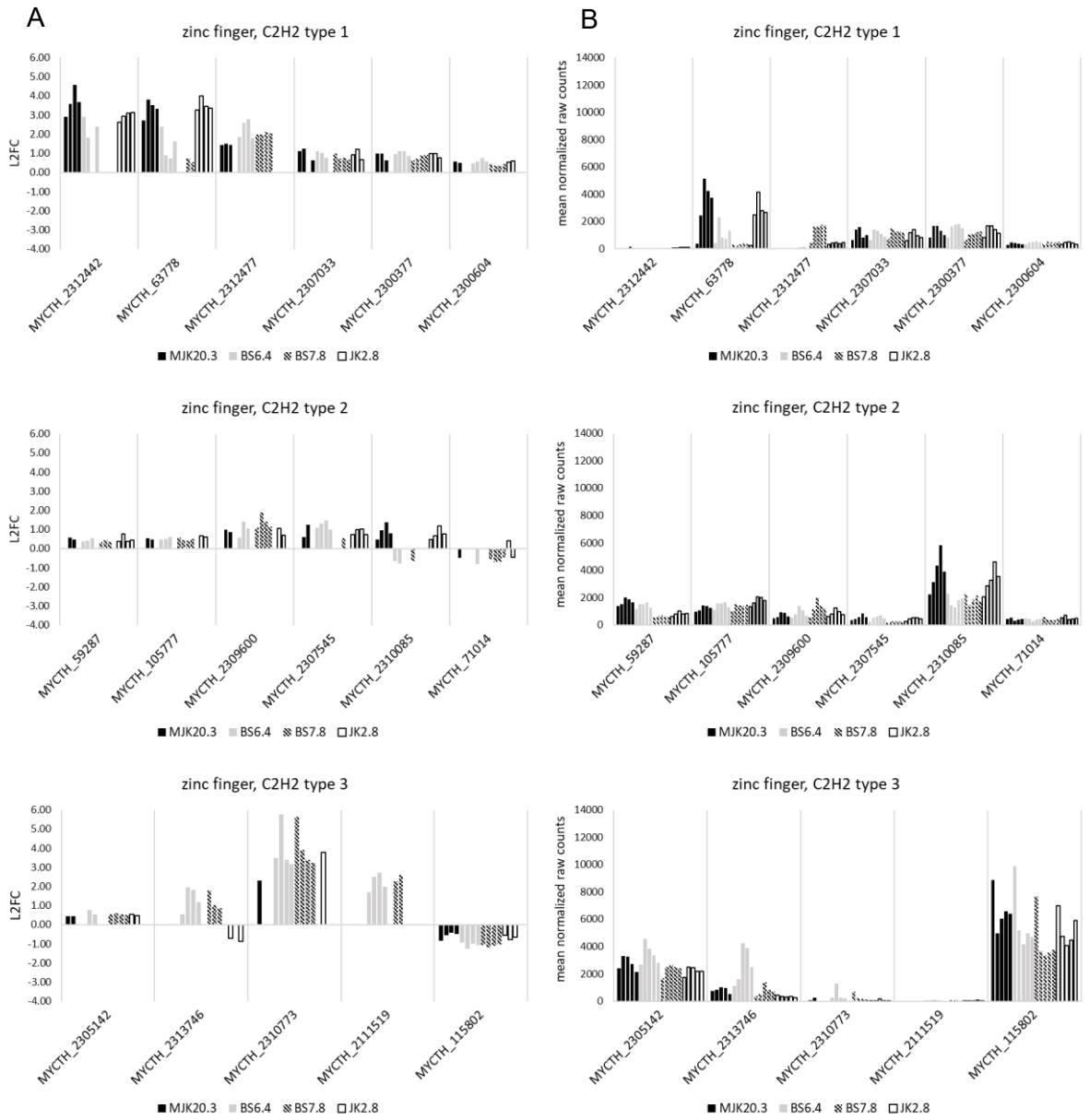

Figure 5.76: Differentially expressed genes encoding for predicted zinc finger, C2H2 type transcription factors in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Among the genes encoding for predicted zinc finger, C2H2 transcription factors (Figure 5.76), the genes MYCTH_63778 (Sp3) is highly expressed in MJK20.3 and JK2.8 (higher in MJK20.3) whereas in BS6.4 only low expression levels and in BS7.8 no expression is observable. For MYCTH_2312477 (regulator Rpn4/proteasome 26S regulatory subunit like) only in BS7.8 high expression levels are detectable. MYCTH_2310085 (Cre1) is strongly expressed and upregulated in MJK20.3 and JK2.8 only (slightly higher in MJK20.3). Finally, MYCTH_2313746 (regulator Nrg1/conidial separation) is upregulated with high expression levels in BS6.4 only.

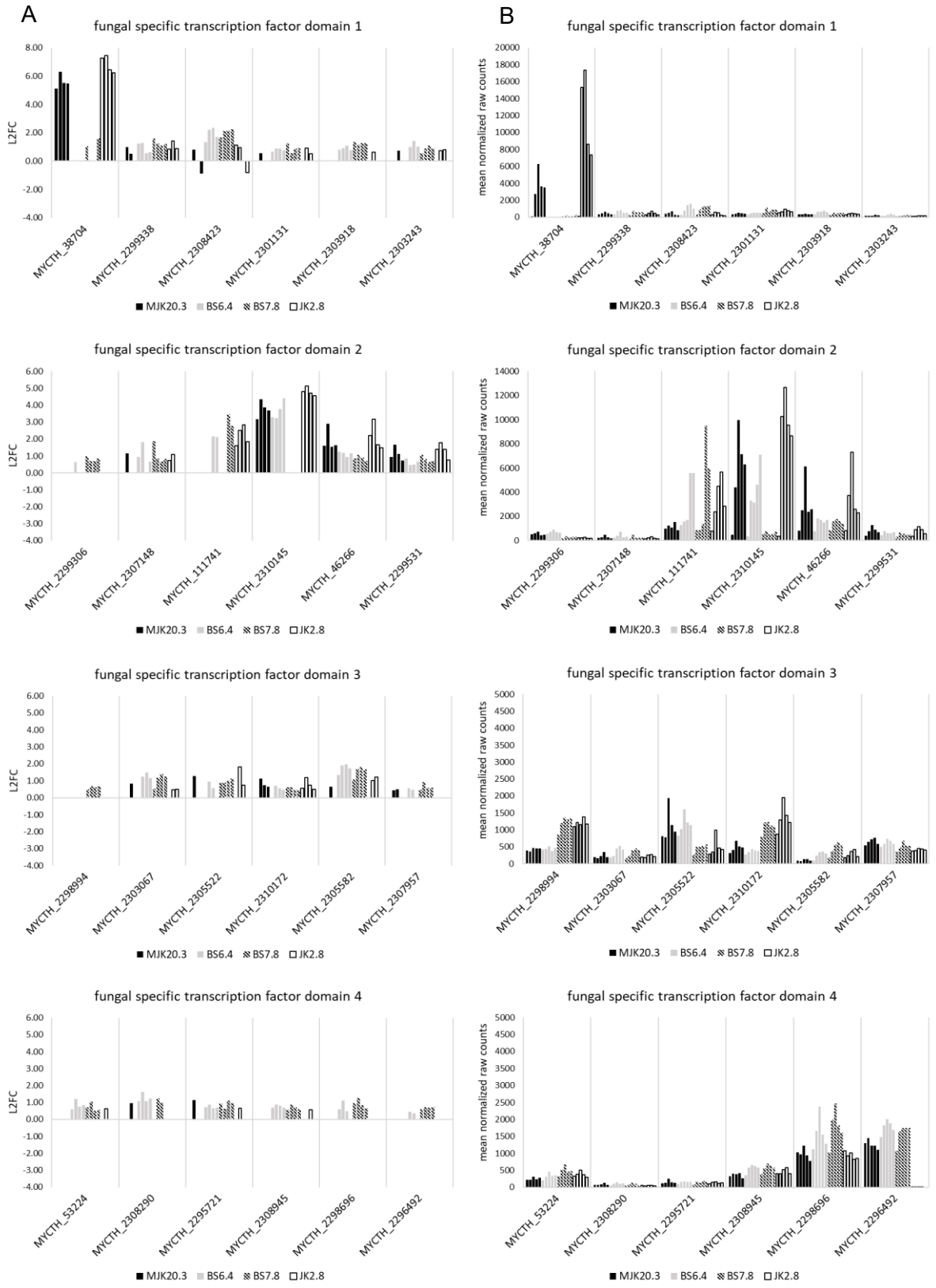

Figure 5.77: Differentially expressed genes encoding for predicted transcription factors with fungal specific transcription factor domain (part 1) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

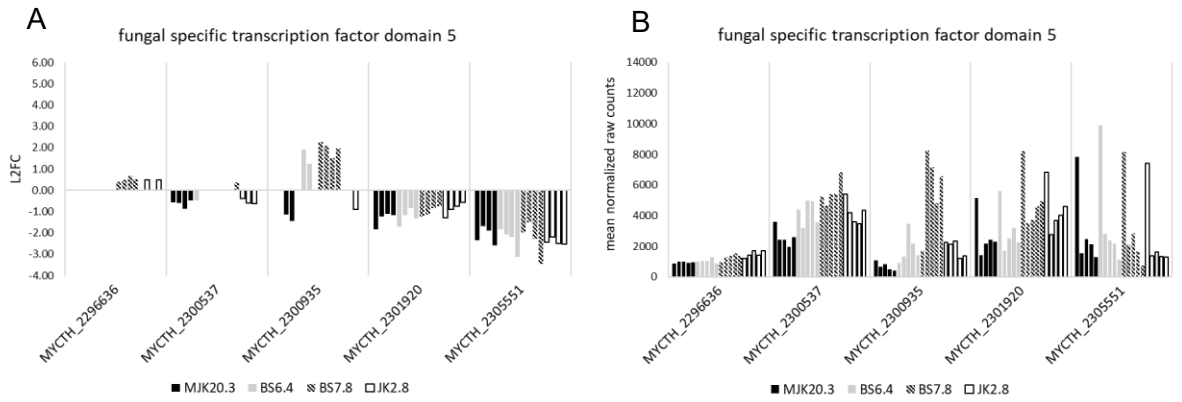

Figure 5.78: Differentially expressed genes encoding for predicted transcription factors with fungal specific transcription factor domain (part 2) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Expression for many of the genes encoding for predicted transcription factors with fungal specific transcription factor domain (Figures 5.77 and 5.78) has already been described for genes belonging to the fungal Zn(2)-Cys(6) binuclear cluster domain class due to the fact that some predicted transcription factors have multiple domains and are therefore, assigned to multiple classes. In the fungal specific transcription factor class, MYCTH_38704 (Clr2) is the upregulated predicted transcription factor gene with the highest differential expression and raw count value. This transcription factor gene is highly expressed in MJK20.3 and JK2.8 only whereas expression levels in JK2.8 are much higher (~2-3 times higher) especially at the first points in time after the spike (t1, t2). MYCTH_111741 (endonuclease Nob1) is highly expressed and upregulated in BS6.4, BS7.8, and JK2.8 only, starting at t2 after the spike.

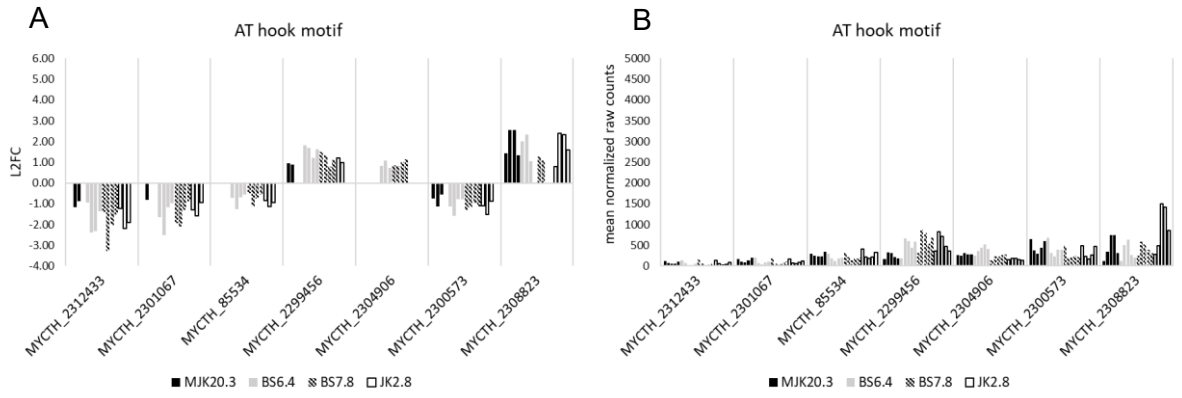

Figure 5.79: Differentially expressed genes encoding for predicted transcription factors with AT hook motif in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Regarding genes encoding for predicted transcription factors with an AT hook motif (Figure 5.79), only MYCTH_2308823 (Moc3/OefC/Pro1/A) is highly expressed and shows differences in expression values. Here, strong expression can be observed for JK2.8 only.

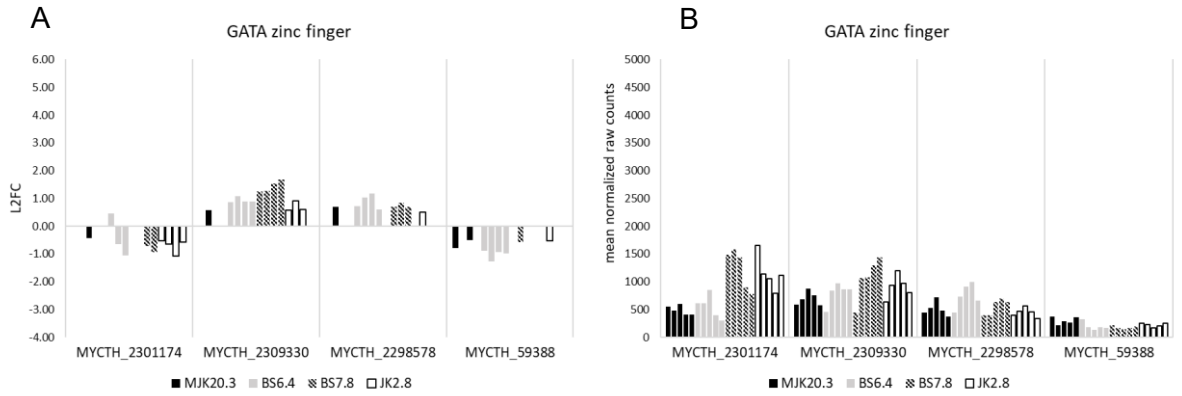

Figure 5.80: Differentially expressed genes encoding for predicted transcription factors with GATA zinc finger in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Among the genes encoding for predicted GATA zinc finger transcription factors (Figure 5.80) only for MYCTH_2301174 (multicopy suppressor protein 2) notable differences between the strains can be detected. Here, expression levels are higher in BS7.8 and JK2.8 compared to the other two strains.

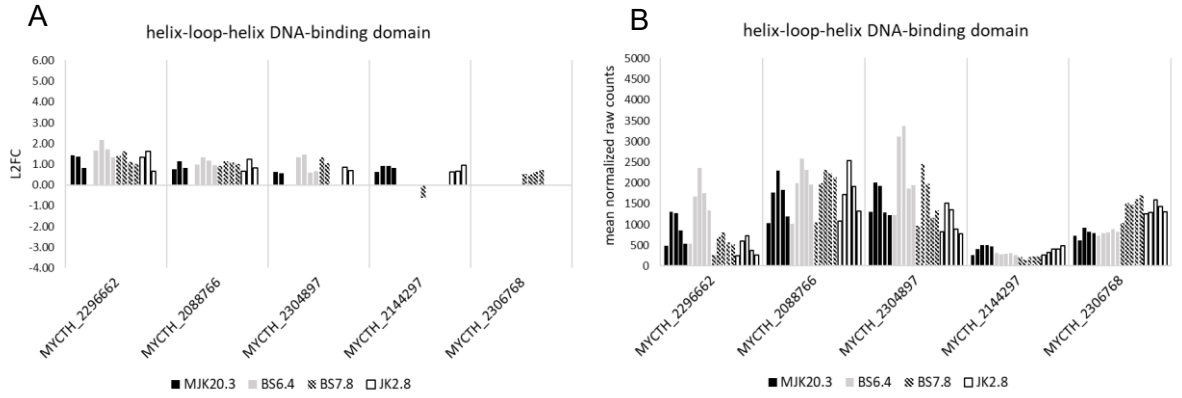

Figure 5.81: Differentially expressed genes encoding for predicted transcription factors with helix loop helix DNA binding domain in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Notable differences in gene expression among genes encoding for predicted transcription factors with helix-loop-helix DNA-binding domain (Figure 5.81) are only observable for MYCTH_2296662 (protein Max), where expression levels are highest in MJK20.3 and JK2.8 (higher in JK2.8).

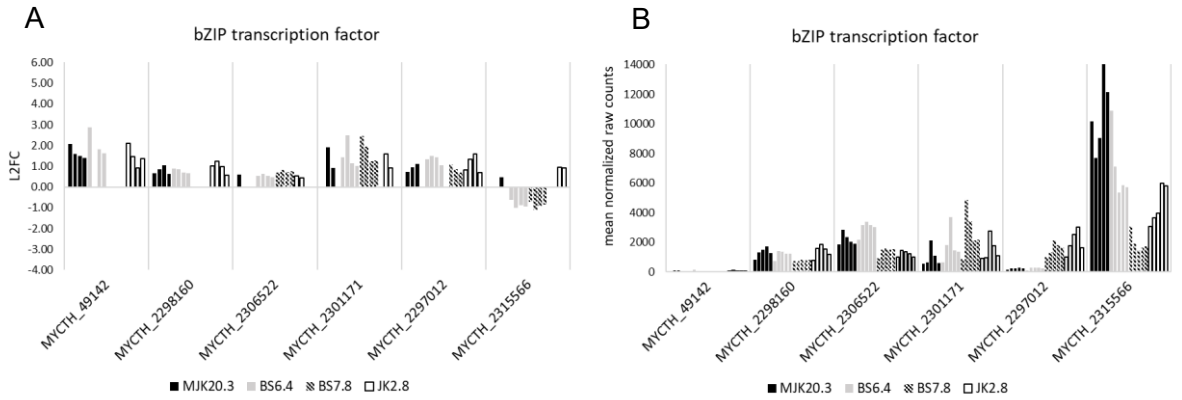

Figure 5.82: Differentially expressed genes encoding for predicted bZIP transcription factors in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty).

Among the genes in the bZIP transcription factor class (Figure 5.82), strong expression differences can only be observed for MYCTH_2315566 (cross pathway control protein 1), which is downregulated in BS6.4 and BS7.8 only and has high expression levels in MJK20.3 and BS6.4 only (MJK20.3 highest).

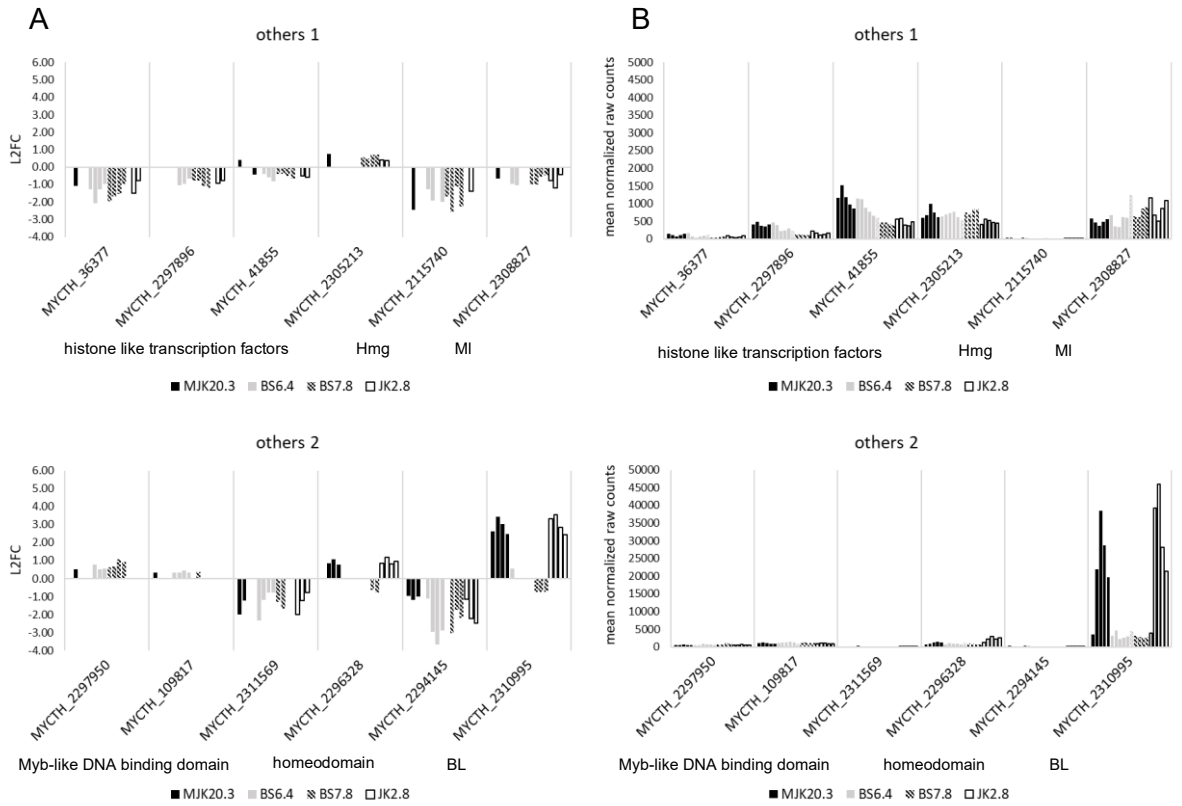

Figure 5.83: Other differentially expressed predicted transcription factor genes (part 1) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty). Hmg= high mobility group box, MI= Myb-like DNA binding domain, BL= basic region leucine zipper.

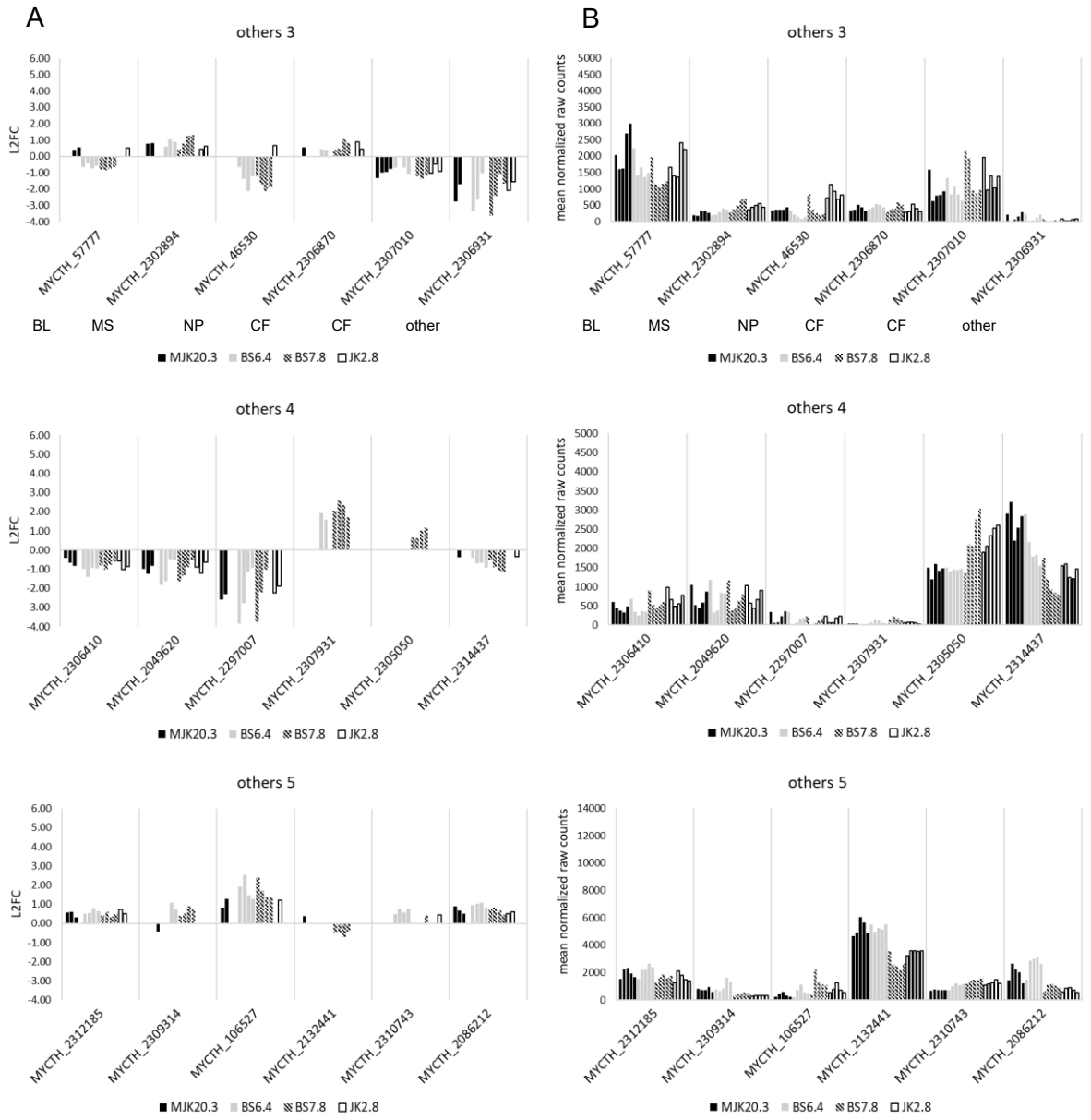

Figure 5.84: Other differentially expressed predicted transcription factor genes (part 2) in strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3 that belong to the t1-t4 intersection area. (A) Log2 fold change values (L2FC) of the single genes at t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty) in relation to the respective steady state condition. (B) Mean of the normalized counts from the two replicates of the single genes from steady state to t1, t2, t3, and t4 (left bar to right bar) for strains MJK20.3 (black, filled), BS6.4 (grey), BS7.8 (patterned), and JK2.8 (black, empty). BL= basic region leucine zipper, MS= MIZ/ SP-RING zinc finger, NP= NDT80/PhoG like DNA binding family, CF= copper fist DNA binding domain.

Among the “other” predicted transcription factor genes (Figures 5.79 and 5.84) some interesting differences for single genes are observable. MYCTH_2310995 (HacA/1), for example, which has the highest expression level of all transcription factors, is only upregulated in MJK20.3 and JK2.8 with extremely high expression levels (much higher at t1-t2 and only slightly higher at t3-t4 in JK2.8). In BS6.4 and BS7.8 much lower expression levels can be detected. MYCTH_57777 (regulatory protein Cys3) is upregulated at the later points in time (t3-t4) in MJK20.3 and JK2.8 only and downregulated in BS6.4 and BS7.8. The genes MYCTH_2305050 (forkhead protein Sep1/2/meiosis specific transcription

factor Mei4) and MYCTH_2314437 (multiprotein-bridging factor 1) have higher expression levels in BS7.8, JK2.8 and MJK20.3, BS6.4 respectively. Finally, MYCTH_2132441 (McmA/1) and MYCTH_2086212 (transcription initiation factor IIA subunit) are stronger expressed in MJK20.3 and BS6.4 compared to the BS7.8 and JK2.8.

Besides the comparison of the expression of predicted transcription factor genes between strains BS6.4, BS7.8, JK2.8 in comparison to the reference strain MJK20.3 as done above, also the expression differences of genes encoding for the orthologs of the known regulators of plant biomass degradation in comparison to the reference strain reveal valuable information about regulation processes in the investigated strains. Therefore, expression of genes encoding for those orthologs with medium to high expression levels, that were already described in chapter 5.3.1.4 were compared between strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3. The L2FC values as well as the mean of the normalized raw counts under the respective conditions are shown in Figure 5.85. If not mentioned specifically, expression of genes is regarded as similar.

		L2FC																mean normalized raw counts																					
		MJK20.3				BS6.4				BS7.8				JK2.8				MJK20.3					BS6.4				BS7.8					JK2.8							
gene	ortholog	t1	t2	t3	t4	t1	t2	t3	t4	t1	t2	t3	t4	t1	t2	t3	t4	SS	t1	t2	t3	t4	SS	t1	t2	t3	t4	SS	t1	t2	t3	t4	SS	t1	t2	t3	t4
MYCTH_2028011	Ace1	0.00	0.00	0.00	0.00	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.00	0.00	0.00	1572	1910	1889	1831	1446	1621	2194	2046	2029	1513	2101	2270	2101	2046	1783	1886	2754	2417	2150	2167
MYCTH_2308260	Ace2	0.00	0.00	-0.91	0.00	0.57	0.00	-4.14	-3.59	0.00	0.00	-3.74	-3.50	0.00	0.00	-3.22	-1.04	378	471	452	201	351	476	707	399	27	40	439	608	407	33	39	463	604	514	51	225
MYCTH_2301920	AmyR/BglR*	-1.84	-1.23	-1.10	-1.15	-1.71	-1.17	-0.83	-1.31	-1.22	-1.14	-0.84	-0.73	-1.30	-0.89	-0.76	-0.56	5142	1435	2186	2405	2310	5611	1712	2487	3160	2258	8213	3522	3725	4600	4938	6806	2766	3679	4020	4615
MYCTH_2307451	Ap3	0.00	0.71	0.00	0.00	0.75	0.85	0.49	0.00	0.87	0.63	0.33	0.28	0.42	0.45	0.36	0.00	304	386	496	368	334	286	472	506	392	321	319	582	494	400	387	355	470	482	453	341
MYCTH_113457	AreA/Nir2/Nit2	0.00	0.62	0.58	0.00	0.00	0.00	0.00	0.00	0.62	0.00	0.34	0.49	0.00	0.83	0.76	0.50	803	958	1227	1201	1068	772	800	886	809	817	903	1391	988	1146	1268	1059	1312	1889	1787	1497
MYCTH_2309867	AreB	-1.08	-0.82	0.00	0.00	-1.29	-1.15	0.00	0.00	-0.90	-0.74	0.00	0.00	-0.62	-0.95	0.00	0.00	527	249	299	519	623	535	219	240	441	491	651	349	390	560	800	659	430	342	594	706
MYCTH_2063030	ClbR/ClbR2/ClbR3	-0.78	-0.60	-0.68	0.00	0.00	0.77	1.30	0.64	0.00	0.59	0.75	0.53	0.00	0.00	0.00	0.00	489	284	322	304	333	452	473	773	1118	708	366	419	550	614	529	366	278	375	357	279
MYCTH_2298863	Clr1*	0.00	0.00	0.00	0.00	0.00	-1.03	-0.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2270	2072	2425	1929	2031	2371	1765	1158	1378	2161	4	3	7	7	6	2478	3120	2597	2230	2377
MYCTH_38704	Clr2*	5.12	6.31	5.52	5.47	0.00	0.00	0.00	0.00	1.04	0.00	0.00	1.60	7.28	7.46	6.45	6.22	80	2740	6258	3624	3510	0	0	1	0	0	100	207	107	124	304	98	15321	17345	8613	7339
MYCTH_2306730	ClrC	1.09	0.80	0.00	0.00	0.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.18	1.01	0.75	1.08	156	332	272	243	238	202	371	260	231	265	58	46	44	79	88	79	179	158	133	167
MYCTH_2296492	Clr4*	0.00	0.00	0.00	0.00	0.00	0.43	0.34	0.00	0.61	0.70	0.70	0.71	0.00	0.00	0.00	0.00	1294	1443	1225	1230	1112	1485	1825	2000	1879	1694	1077	1642	1753	1747	1757	1	0	0	1	0
MYCTH_2310085	Cre1*	0.49	0.96	1.37	0.79	-0.65	-0.79	0.00	0.00	-0.64	0.00	0.00	0.00	0.46	0.66	1.16	0.78	2245	3142	4354	5817	3873	2261	1443	1309	1816	1900	2284	1470	1824	2125	1701	2061	2843	3254	4615	3528
MYCTH_2055311	CreB	0.00	0.47	0.60	0.00	0.00	0.00	0.57	0.86	0.00	0.00	0.00	0.00	0.75	0.00	0.61	0.67	775	1006	1068	1173	1074	852	1070	956	1263	1541	2041	1863	1638	1806	2252	2111	3552	2705	3233	3362
MYCTH_2306444	CreC	0.00	0.00	0.00	0.00	0.34	0.40	0.30	0.52	0.42	0.44	0.65	0.80	0.00	0.51	0.39	0.00	499	457	582	615	491	528	681	706	661	766	495	661	674	777	864	544	630	774	712	609
MYCTH_2306452	CreD	0.00	-0.53	0.90	0.00	0.00	0.00	0.56	0.00	-0.48	0.00	0.55	0.49	-0.49	0.00	1.12	0.47	365	357	252	680	453	386	316	342	572	440	410	293	375	602	574	409	290	331	889	565
MYCTH_59287	Crz1	0.00	0.57	0.46	0.00	0.38	0.41	0.53	0.00	0.36	0.43	0.35	0.00	0.36	0.76	0.39	0.44	1361	1504	2018	1874	1633	1155	1498	1536	1661	1227	531	680	717	677	585	607	782	1028	798	827
MYCTH_46266	GaaR/Pdr2	1.60	2.90	1.53	1.65	1.24	1.16	0.95	1.15	0.86	1.05	0.90	0.70	2.19	3.17	1.67	1.48	826	2506	6143	2391	2582	768	1812	1722	1482	1708	856	1557	1771	1601	1393	814	3725	7328	2595	2271
MYCTH_46981	GaaX	1.68	2.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.72	1.87	1.84	0.00	0.00	261	835	1095	344	375	299	438	364	319	326	617	781	866	881	1018	615	2248	2196	790	664
MYCTH_2310995	HacA/1	2.64	3.45	3.03	2.49	0.57	0.00	0.00	0.00	0.00	-0.72	-0.76	-0.72	3.31	3.54	2.84	2.44	3522	21933	38565	28685	19763	3120	4626	2193	2551	2982	4454	3113	2699	2633	2696	3946	39188	46006	28173	21396
MYCTH_2297059	Hap2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1333	1463	1443	985	1177	1398	1523	1623	1185	1035	569	667	735	589	541	651	672	624	515	588
MYCTH_41855	Hap3	0.40	0.00	0.00	-0.44	0.00	-0.38	-0.57	-0.80	-0.41	-0.37	-0.51	-0.66	0.00	-0.51	-0.60	0.00	1153	1522	1186	977	851	1142	1126	883	772	658	604	457	465	423	383	558	576	392	369	488
MYCTH_67051	Hap5	0.00	0.00	-0.31	0.00	0.00	-0.25	-0.46	-0.43	0.00	0.00	-0.45	-0.44	0.00	0.00	-0.33	0.00	1365	1347	1207	1101	1376	1418	1475	1187	1030	1053	596	539	543	439	441	676	722	639	536	644
MYCTH_2309600	Hcr1	0.00	0.98	0.87	0.00	0.56	1.41	1.04	0.00	1.09	1.87	1.39	1.13	0.00	1.04	0.70	0.00	476	557	932	866	620	522	767	1379	1068	680	546	1160	1997	1430	1197	611	793	1260	995	743
MYCTH_2295635	Hep1/HP1	-1.68	0.00	0.00	0.00	-1.49	-1.25	0.00	0.00	-2.07	-2.60	0.00	0.00	0.00	0.00	0.00	0.00	94	29	55	112	100	85	30	36	79	72	21	5	3	15	13	16	11	10	14	20
MYCTH_2308921	Ire1	0.00	0.59	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.00	0.75	1.09	0.78	0.92	0.00	0.73	1517	1619	2275	1736	1989	1639	1603	1537	1206	1293	985	1424	1256	1658	2090	1408	2421	2677	1706	2346
MYCTH_2294559	Lae1/A	0.00	0.00	0.00	0.00	0.00	0.00	-1.47	-2.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	213	296	261	199	181	719	1471	421	260	171	334	334	251	231	180	198	404	256	145	221
MYCTH_2303067	MalR	0.00	0.00	0.84	0.00	0.00	1.26	1.49	1.16	0.53	1.19	1.40	1.24	0.00	0.47	0.49	0.00	191	166	231	342	200	185	231	445	525	417	171	248	391	451	403	188	182	261	265	207
MYCTH_2132441	McmA/1	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00	-0.45	-0.51	-0.72	-0.43	0.00	0.00	0.00	0.00	4639	4910	6047	5654	4900	5512	4966	5245	5139	5508	3525	2573	2473	2146	2625	3219	3576	3605	3524	3563
MYCTH_2303918	Mhr1	0.00	0.00	0.00	0.00	0.80	0.93	1.10	0.76	1.36	1.16	1.27	1.28	0.00	0.63	0.00	0.00	319	338	378	355	326	351	613	669	753	595	215	551	478	517	521	308	379	479	373	354
MYCTH_2298994	NirA/Nit4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.68	0.61	0.67	0.00	0.00	0.00	0.00	386	367	471	453	454	424	430	506	376	460	853	1201	1361	1303	1355	1097	1212	1155	1384	1165
MYCTH_2302460	NmrA/1	0.00	0.36	0.83	1.00	0.00	0.00	0.00	0.44	0.00	0.00	0.00	0.37	0.49	0.72	0.63	0.82	598	573	770	1066	1199	508	537	581	610	683	983	943	1039	1087	1268	1006	1409	1659	1550	1778
MYCTH_81165	PacC/1	0.00	1.40	0.00	0.00	0.53	1.08	0.00	0.00	0.89	1.11	0.00	0.00	1.23	1.19	0.00	0.00	1211	1652	3194	1664	1340	1183	1710	2500	1365	1275	1432	2652	3093	1707	1931	1508	3526	3430	1750	1689
MYCTH_2303559	Prk6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	192	202	202	182	182	196	195	221	212	227	233	233	208	247	273	231	253	204	204	212
MYCTH_2300719	Rca1	0.00	0.00	0.00	0.00	-0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1650	1455	1636	1607	1518	1768	1228	1455	1835	1382	1532	1385	1715	1719	1445	1607	1443	1443	1578	1607
MYCTH_2298696	Rce1	0.00	0.00	0.00	0.00	0.59	1.10	0.48	0.00	0.95	1.27	0.83	0.66	0.00	0.00	0.00	0.00	1025	970	1228	944	781	1113	1666	2384	1553	1286	1024	1983	2471	1825	1614	1077	930	1009	819	853
MYCTH_2302052	Res1	0.00	0.00	0.72	0.00	0.00	0.00	0.58	0.00	-0.66	-0.49	0.00	0.00	0.00	0.00	0.00	0.00	916	886	1096	1511	1312	868	742	879	1295	863	796	504	567	582	587	927	943	1035	1044	1342
MYCTH_53224	RhaR/Pdr1	0.00	0.00	0.00	0.00	0.60	1.19	0.74	0.84	0.71	1.06	0.50	0.57	0.00	0.62	0.00	0.00	225	223	310	243	282	202	304	458	335	360	321	524	667	454	477	325	381	498	364	296
MYCTH_2306768	Sah2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.54	0.51	0.63	0.72	0.00	0.00	0.00	0.00	730	620	925	827	800	728	793	805	890	822	1039	1509	1484	1605	1706	1262	1288	1600	1426	1306
MYCTH_2297068	Stk12	0.00	0.66	0.00	0.00	1.00	1.33	1.36	1.00	1.11	1.41	1.62	1.92	0.00	0.00	0.00	0.00	144	191	227	219	163	139	281	353	361	281	2087	4497	5557	6431	7884	3572	4027	5101	3854	5008
MYCTH_2312657	VeA/Vel1	0.00	0.82	0.63	0.52	0.00	0.00	0.00	0.00	-0.44	-0.48	0.00	0.45	0.41	0.98	0.58	0.89	828	924	1458	1280	1187	1132	1024	967	959	1387	616	453	441	591	839	479	639	949	718	890
MYCTH_113912	VelB	0.00	0.87	0.75	0.00	0.00	0.00	0.00	0.00	0.91	0.00	0.00	0.00	0.00	0.82	0.61	0.00	343	440	625	574	449	345	423	414	327	322	103	197	104	133	113	128	178	228	195	157
MYCTH_46530	Vib1	0.00	0.00	0.00	0.00	-0.64	-1.37	-2.13	-1.22	-1.21	-1.70	-2.13	-1.88	0.66	0.00	0.00	0.00	350	359	355	365	431	331	212	127	75	141	825	357	254	188	224	716	1131	926	689	811
MYCTH_108157	Vib2	0.00	0.44	0.67	0.00	0.00	0.00	0.00	0.39	0.00	0.00	0.00	0.44	-0.55	0.00	0.00	0.00	825	707	1119	1313	810	865	697	871	1021	1130	772	766	730	928	1049	767	522	761	796	601
MYCTH_2309330	Wc1/Blr1	0.00	0.57	0.00	0.00	0.86	1.08	0.89	0.90	1.25	1.27	1.53	1.69	0.57	0.91	0.61	0.00	588	687	871	751	571	460	839	974	858	860	448	1065	1082	1292	1442	633	939	1195	966	810
MYCTH_2294022	Wc2/Blr2	0.00	0.00	0.00	0.00	0.62	1.08	1.20	1.01	0.89	1.86	1.82	1.57	0.00	0.00	0.58	0.00	556	438	485	550	376	480	732	1007	1096	961	591	1092	2143	2079	1749	635	566	628	946	449
MYCTH_2310145	Xyr1*	3.15	4.33	3.85	3.67	3.28	3.22	3.77	4.39	0.00	0.00	0.00	0.00	4.81	5.11	4.71	4.56	494	4394	9959	7135	6289	340	3290	3139	4621	7097	579	782	520	576	714	365	10269	12669	9548	8664
MYCTH_2144297	Xpp1	0.63	0.93	0.92	0.83	0.00	0.00	0.00	0.00	-0.62	0.00	0.00	0.00	0.00	0.64	0.66	0.94	263	406	499	497	468	314	275	297	314	268	223	144	217	229	224	258	321	401	406	496

Figure 5.85: Differential expression of genes encoding for orthologs of known regulators of plant biomass degradation in *T. thermophilus* strains BS6.4, BS7.8, JK2.8, and the reference strain MJK20.3. Shown are the log2 fold change values (L2FC) at: 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after cellulose spike (t4) compared to the respective steady state condition as well as the mean of the normalized raw counts of the respective condition. Green color indicates an upregulation and red color indicates a downregulation. The dash separates the possible orthologs of this regulator if more than one was found. An asterisk marks the regulators that have already been investigated in *T. thermophilus*.

Among the genes encoding for orthologs of known regulators of plant biomass degradation that are upregulated in at least two points in time after the cellulose spike and have medium to high expression levels in strain MJK20.3 are MYCTH_113457 (AreA/Nir2/Nit2), MYCTH_38704 (Clr2), MYCTH_2310085 (Cre1), MYCTH_2055311 (CreB), MYCTH_59287 (Crz1), MYCTH_46266 (GaaR/Pdr2), MYCTH_46981 (GaaX), MYCTH_2310995 (HacA/1), MYCTH_2308921 (Ire1), MYCTH_2302460 (NmrA/1), MYCTH_81165 (PacC/1), MYCTH_2312657 (VeA/Vel1), MYCTH_108157 (Vib2), MYCTH_2309330 (Wc1/Blr1), MYCTH_2310145 (Xyr1), and MYCTH_2144297 (Xpp1).

Out of those, MYCTH_113457 (AreA/Nir2/Nit2), MYCTH_38704 (Clr2), MYCTH_2055311 (CreB), MYCTH_46266 (GaaR/Pdr2; t1-t3 only), MYCTH_46981 (GaaX), MYCTH_2310995 (HacA/1; t1, t2, t4), MYCTH_2308921 (Ire1; t1, t2, t4), MYCTH_2302460 (NmrA/1), MYCTH_81165 (PacC/1), MYCTH_2309330 (Wc1/Blr1), and MYCTH_2310145 (Xyr1) are stronger expressed in JK2.8. Out of the top 5 with the strongest expression in MJK20.3, all genes except MYCTH_2310085 (Cre1) are stronger expressed in JK2.8.

In BS6.4, MYCTH_113457 (AreA/Nir2/Nit2), MYCTH_38704 (Clr2), MYCTH_2310085 (Cre1), MYCTH_46266 (GaaR/Pdr2), MYCTH_46981 (GaaX), MYCTH_2310995 (HacA/1), MYCTH_2308921 (Ire1), and MYCTH_2310145 (Xyr1; only t1-t3) are weaker expressed, compared to MJK20.3. All other of the above-mentioned genes are equally expressed in both strains or slightly higher expressed in BS6.4.

Finally, in BS7.8, MYCTH_113457 (AreA/Nir2/Nit2; >BS6.4), MYCTH_38704 (Clr2; >BS6.4), MYCTH_2310085 (Cre1; t1-t3>BS6.4), MYCTH_2055311 (CreB; >BS6.4), MYCTH_59287 (Crz1;<BS6.4), MYCTH_46266 (GaaR/Pdr2; t1, t4<BS6.4), MYCTH_46981 (GaaX; >BS6.4), MYCTH_2310995 (HacA/1; t1, t4<BS6.4), MYCTH_2308921 (Ire1; t1, t2<BS6.4), MYCTH_2312657 (VeA/Vel1; <BS6.4), and MYCTH_2310145 (Xyr1; <BS6.4) are weaker expressed compared to MJK20.3. The remaining genes MYCTH_2302460 (NmrA/1), MYCTH_81165 (PacC/1), and MYCTH_2309330 (Wc1/Blr1) are slightly higher expressed in BS7.8 compared to MJK20.3 and BS6.4.

Among the genes that are downregulated with medium to high expression levels in SS, only MYCTH_2301920 (AmyR/BglR) could be detected. This gene has only slightly higher expression values in BS6.4 (t1-t3) as well as much higher expression levels in BS7.8 (highest expression at all points in time) and in JK2.8 (higher expression at all points in time but weaker expression compared to BS7.8) compared to MJK20.3. The L2FC values for this gene are almost identical across all strains.

Genes encoding for orthologs of known regulators of plant biomass degradation with medium to high expression levels that show no differential expression or only at one point in time after the cellulose spike in MJK20.3 include MYCTH_2028011 (Ace1), MYCTH_2298863 (Clr1), MYCTH_2296492 (Clr4), MYCTH_2297059 (Hap2), MYCTH_41855 (Hap3), MYCTH_67051 (Hap5), MYCTH_2132441 (McmA/1), MYCTH_2300719 (Rca1), MYCTH_2298696 (Rce1), MYCTH_2302052 (Res1), and MYCTH_2306768 (Sah2).

Out of these, MYCTH_2028011 (Ace1), MYCTH_2298863 (Clr1), and MYCTH_2306768 (Sah2) are stronger and MYCTH_2296492 (Clr4), MYCTH_2297059 (Hap2), MYCTH_41855 (Hap3), MYCTH_67051 (Hap5), and MYCTH_2132441 (McmA/1) are weaker expressed in JK2.8 (no expression of *clr4*) compared to MJK20.3.

In BS6.4, MYCTH_2028011 (Ace1), MYCTH_2296492 (Clr4), MYCTH_2297059 (Hap2), MYCTH_2132441 (McmA/1; t1, t4 only), and MYCTH_2298696 (Rce1) are stronger and MYCTH_2298863 (Clr1), MYCTH_41855 (Hap3), MYCTH_67051 (Hap5), and MYCTH_2302052 (Res1) are weaker expressed compared to MJK20.3.

Finally, in BS7.8, MYCTH_2028011 (Ace1; >BS6.4), MYCTH_2296492 (Clr4; <BS6.4), MYCTH_2298696 (Rce1; >BS6.4), and MYCTH_2306768 (Sah2; >BS6.4) are stronger and MYCTH_2298863 (Clr1; <BS6.4) MYCTH_2297059 (Hap2; <BS6.4), MYCTH_41855 (Hap3; <BS6.4), MYCTH_67051 (Hap5; <BS6.4), MYCTH_2132441 (McmA/1; <BS6.4), and MYCTH_2302052 (Res1; <BS6.4) are weaker expressed, compared to MJK20.3.

Other genes, that were having low expression levels in MJK20.3 and were therefore not mentioned yet, include MYCTH_2294022 (Wc2/Blr2) and MYCTH_2297068 (Stk12). MYCTH_2294022 (Wc2/Blr2) is much stronger expressed in BS6.4 and BS7.8 (highest in BS7.8) compared to MJK20.3. In JK2.8 no differences to MJK20.3 can be detected. The gene MYCTH_2297068 (Stk12) has much higher expression values in BS7.8 (up to ~25x higher) and JK2.8 (up to ~20x higher) compared to MJK20.3 and BS6.4. In BS6.4, expression of MYCTH_2297068 (Stk12) is similar to MJK20.3, where almost no expression is observable.

In summary, expression of the highest expressed predicted transcription factor genes and genes encoding for orthologs of known regulators of plant biomass degradation in MJK20.3 is reduced in BS6.4 and BS7.8. This effect is much stronger in BS7.8. The most obvious differences between BS6.4, BS7.8, and MJK20.3 are the expression of MYCTH_2121737 (sterol regulatory Cys6/Ara1), MYCTH_63778 (Sp3), MYCTH_38704 (Clr2), MYCTH_2310995 (HacA/1), MYCTH_2310085 (Cre1), MYCTH_46266 (GaaR/Pdr2), MYCTH_2132441 (McmA/1), MYCTH_2297068 (Stk12), and MYCTH_2310145 (Xyr1), whereas the most apparent differences between BS6.4 and BS7.8 are the expression of MYCTH_2132441 (McmA/1) and MYCTH_2297068 (Stk12). In JK2.8, in contrary, expression of the in MJK20.3 highest expressed predicted transcription factor genes and genes encoding for orthologs of known regulators of plant biomass degradation is higher, while the general expression pattern is almost identical compared to MJK20.3. The most obvious differences between JK2.8 and MJK20.3 are the expression of MYCTH_38704 (Clr2), MYCTH_2310995 (HacA/1), MYCTH_2132441 (McmA/1), MYCTH_2297068 (Stk12), and MYCTH_2310145 (Xyr1).

5.3.3 Response of the reference strain MJK20.3 (Δ*ku80*) to xylose

5.3.3.1 General and enrichment analysis

For differential expression analysis, the samples of the xylose cultivation taken in steady state (SS) and two different points in time in exponential state (ex1= 3 h, ex2 = 4 h after starting the wash out) were compared to the respective sample of the glucose cultivation. The number of the resulting differentially expressed genes with padj. ≤ 0.05 is shown in Table 5.25. Many genes are differentially expressed, more than 50 % in exponential state and less than 50 % in steady state. The ratio between up- and downregulated genes is similar. The numbers of differentially expressed genes for both exponential state samples are almost identical. In steady state the respective numbers are slightly lower.

Table 5.25: Number of differentially expressed genes in exponential state (ex1, ex2) and steady state (SS) with padj. ≤ 0.05 of strain MJK20.3 cultivated on xylose in relation to the respective condition during cultivation on glucose.

strain	condition	upregulated genes	downregulated genes
MJK20.3	ex1	2697	2469
	ex2	2678	2399
	SS	2170	2139

To narrow down the number of differentially expressed genes, those with a high differential expression (L2FC value > |2|) were separated. The respective numbers can be found in Table 5.26. Regarding the genes with L2FC value > |2|, much lower numbers can be detected in comparison to the total number of differentially expressed genes. In exponential state approximately 10 % of all genes are differentially expressed and in steady state only 5 %. The relation of up- to downregulated genes is in exponential state approximately 2:3 and in steady state 3:2. The numbers for the different points in time in exponential state are almost identical.

Table 5.26: Number of differentially expressed genes in exponential state (ex1, ex2) and steady state (SS) with padj. ≤ 0.05 and L2FC value > |2| of strain MJK20.3 cultivated on xylose in relation to the respective condition during cultivation on glucose.

strain	condition	upregulated genes	downregulated genes
MJK20.3	ex1	483	687
	ex2	475	660
	SS	310	233

To get an idea of the overall transcriptomic response to xylose, a GO enrichment analysis using the DAVID database (Huang et al. 2009a; Huang et al. 2009b) with the strongest differentially expressed genes (L2FC value > |2|) was performed as previously described (see chapter 5.3.1.1) The results are shown in Tables 5.27-5.32.

Table 5.27: GO term enrichment analysis of upregulated genes in exponential state (ex1) with L2FC value > |2| of strain MJK20.3 cultivated on xylose in relation to the respective condition during cultivation on glucose. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	carbohydrate metabolic process	3	22
	carbohydrate transport	6	4
	gluconeogenesis	9	3
	xylan catabolic process	10	5
CC	extracellular region	3	6
	integral component of membrane	1	96
	mitochondrion	2	20
MF	carbohydrate binding	4	5
	endo-1,4-beta-xylanase activity	8	5
	flavin adenine dinucleotide binding	2	10
	oxidoreductase activity	2	25
	ribosome binding	5	4

Table 5.28: GO term enrichment analysis of downregulated genes in exponential state (ex1) with L2FC value > |2| of strain MJK20.3 cultivated on xylose in relation to the respective condition during cultivation on glucose. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	carbohydrate metabolic process	3	12
	metabolic process	3	9
	transmembrane transport	3	16
CC	integral component of membrane	2	148
MF	hydrolase activity	2	18
	monooxygenase activity	3	6
	oxidoreductase activity	2	25
	oxidoreductase activity, acting on CH-OH group of donors	4	6
	phosphopantetheine binding	5	5
	protein kinase activity	3	16
	transferase activity	3	10

Table 5.29: GO term enrichment analysis of upregulated genes in exponential state (ex2) with L2FC value > |2| of strain MJK20.3 cultivated on xylose in relation to the respective condition during cultivation on glucose. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	carbohydrate metabolic process	3	22
	gluconeogenesis	9	3
	xylan catabolic process	11	5
CC	extracellular region	5	9
	integral component of membrane	1	95
	mitochondrion	2	19
MF	carbohydrate binding	4	5
	cellulose binding	5	5
	endo-1,4-beta-xylanase activity	8	5
	hydrolase activity, hydrolyzing O-glycosyl compounds	2	10
	oxidoreductase activity	2	26

Table 5.30: GO term enrichment analysis of downregulated genes in exponential state (ex2) with L2FC value > |2| of strain MJK20.3 cultivated on xylose in relation to the respective condition during cultivation on glucose. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	carbohydrate metabolic process	3	12
	chitin catabolic process	12	3
	metabolic process	2	9
	transmembrane transport	3	15
CC	integral component of membrane	2	143
MF	hydrolase activity	2	16
	monooxygenase activity	4	7
	O-methyltransferase activity	10	3
	oxidoreductase activity	2	20
	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	3	8
	phosphopantetheine binding	5	5
	protein kinase activity	3	13
	transferase activity	4	11

Table 5.31: GO term enrichment analysis of upregulated genes in steady state with L2FC value > |2| of strain MJK20.3 cultivated on xylose in relation to the respective condition during cultivation on glucose. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	carbohydrate metabolic process	4	23
	carbohydrate transport	8	4
	metabolic process	3	11
	polysaccharide catabolic process	8	4
	xylan catabolic process	18	6
CC	extracellular region	13	11
MF	carbohydrate binding	8	6
	cellulose binding	11	7
	endo-1,4-beta-xylanase activity	15	6
	hydrolase activity	2	14
	hydrolase activity, hydrolyzing O-glycosyl compounds	4	11
	phosphopantetheine binding	6	5
	substrate-specific transmembrane transporter activity	4	5
	transferase activity, transferring acyl groups other than amino-acyl groups	6	4

Table 5.32: GO term enrichment analysis of downregulated genes in steady state with L2FC value > |2| of strain MJK20.3 cultivated on xylose in relation to the respective condition during cultivation on glucose. The fold enrichment and the corresponding number of genes of single GO terms that belong to the different GO term categories “biological process” (BP), “cellular component” (CC), and “molecular function” (MF) using a p-value cutoff of 0.05 are shown.

category	GO term	fold enrichment	genes
BP	actin nucleation	40	2
CC	integral component of membrane	1	51
	old growing cell tip	45	2
MF	3-oxo-arachidoyl-CoA synthase activity	60	2
	3-oxo-cerotoyl-CoA synthase activity	60	2
	3-oxo-lignoceronyl-CoA synthase activity	60	2

Biological processes that are enriched among the upregulated genes in exponential state (Tables 5.27 and 5.29) are carbohydrate metabolic processes including “xylan catabolic process”, “carbohydrate transport”, and “gluconeogenesis”, whereas “carbohydrate transport” is only enriched in ex1. The GO terms of the category “cellular component” that are enriched include “integral component of membrane”, “extracellular region”, and genes related to mitochondria. Regarding molecular function, “xylanase activity”, “oxidoreductase activity” as well as “carbohydrate binding” are enriched in both exponential state samples. “Ribosome binding” (ex1), “flavin adenine dinucleotide binding” (ex1), “cellulose binding” (ex2), and “hydrolase activity, hydrolyzing O-glycosyl compounds” (ex2) are exclusively enriched in the respective exponential state samples.

The fold enrichment among the downregulated genes during the exponential state (Tables 5.28 and 5.30) is mainly low. Enriched biological processes are “transmembrane transport” and “metabolic process”, including “carbohydrate metabolic process”. Exclusive for ex2 are “chitin catabolic process”. The only GO term of the category “cellular component” among the downregulated genes is “integral component of membrane”. Among the GO terms in the molecular function category, “protein kinase activity”, “oxidoreductase activity”, “hydrolase activity”, “transferase activity”, “monooxygenase activity”, and “phosphopantetheine binding” are enriched. Exclusive for ex2 is “O-methyltransferase activity”.

Biological processes that are enriched among the upregulated genes in steady state (Table 5.31) are “carbohydrate metabolic process” including “xylan catabolic process” and “carbohydrate transport”. In the “cellular component” category only the GO term “extracellular region” is enriched. In the “molecular function” category mainly processes that are linked to carbohydrate metabolism, including “xylanase activity”, “carbohydrate binding” (e.g. “cellulose binding modules”), “transferase activity”, “transporter”, and “hydrolase activity” are enriched. In contrast to the exponential state, “phosphopantetheine binding” is enriched among the upregulated genes in steady state. The GO terms “oxidoreductase” and “gluconeogenesis” are exclusively enriched in exponential state.

Among the downregulated genes in steady state (Table 5.32) the fold enrichment is extremely high, but the number of genes that are enriched is very low. Here, mainly processes that are connected to hyphal tip growth like “actin nucleation” and “old growing cell tip” (both GO terms share one gene) are enriched as well as various “CoA synthase activities” (the same two genes).

5.3.3.2 Differential expression of proteases

The purpose of the investigation of protease expression has already been mentioned (see chapter 5.3.1.2). An overview of predicted *T. thermophilus* proteases was provided in this chapter as well.
The total number of predicted protease genes that were differentially expressed at exponential state and steady state during cultivation on xylose is shown in Figure 5.86. Both exponential state samples, are very similar, with a slightly lower number of up- and downregulated predicted protease genes in ex2 compared to ex1. In steady state, a lower number of predicted protease genes is up- and downregulated compared to the exponential state samples. In total, about 65 % of all predicted protease genes are differentially expressed in exponential state and roughly 50 % in steady state. The relation of up- to downregulated genes is with approximately 1.5:1 slightly higher in steady state.

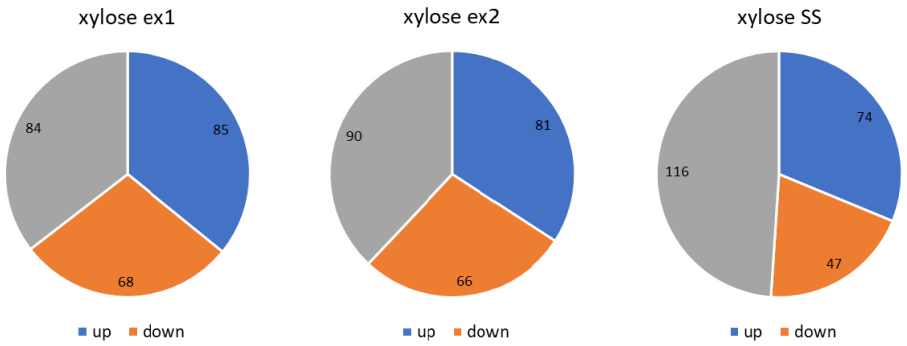

Figure 5.86: Number of differentially expressed predicted protease genes of strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. Shown are the numbers of up- (blue) and downregulated (orange) genes as well as genes with no differential expression (grey) in exponential state (ex1, ex2) and steady state (SS) compared to the respective condition during cultivation on glucose.

To get a detailed overview of the expression profiles of the predicted protease genes on xylose, a heatmap (Figure 5.87) was created, showing the differential expression levels of genes belonging to the different classes of proteases at exponential state and steady state during cultivation of xylose compared to glucose. Generally, L2FC values are for most of the genes rather low. Differential expression in the exponential state samples is almost identical, except a small number of single genes that have higher or lower L2FC values in either ex1 or ex2. The expression profile in steady state is completely different compared to the exponential state. Here, a much weaker differential expression is observable compared to the exponential state. The strongest differential expression in exponential state can be seen for genes encoding for predicted serine peptidases. During steady state, predicted protease genes with differential expression are evenly distributed among the different classes.

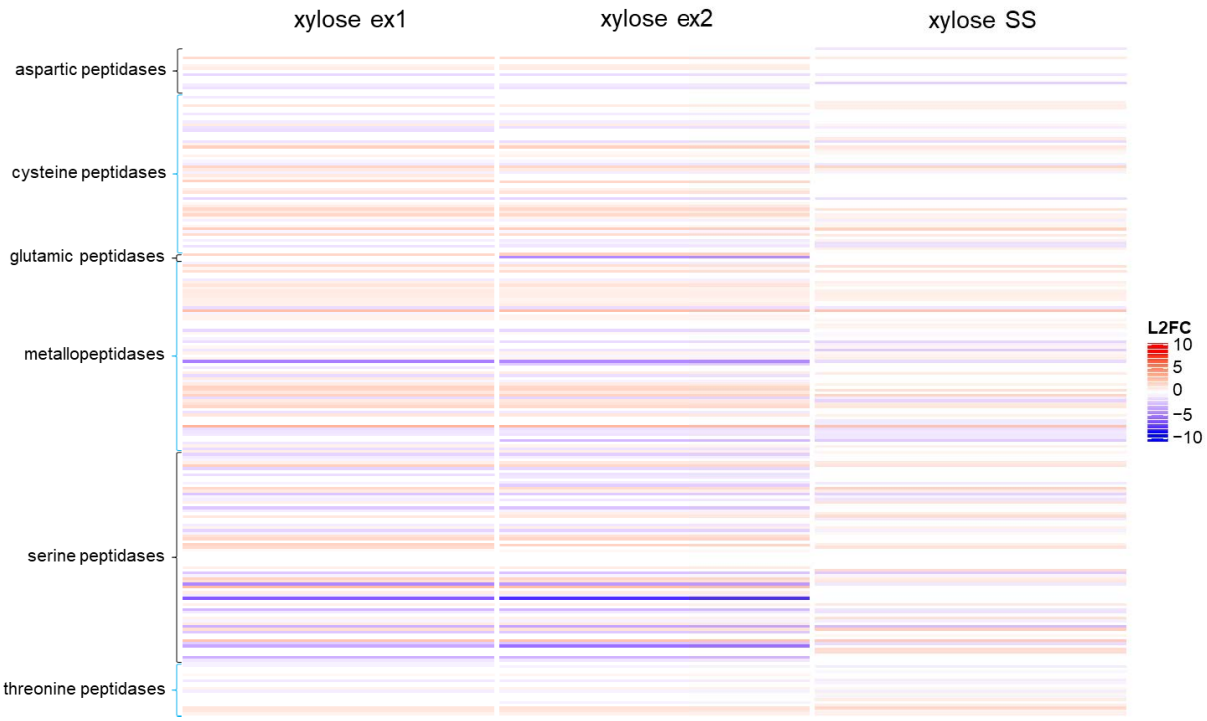

Figure 5.87: Heatmap with differential expression values of all predicted protease genes in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. Shown are the log2 fold change values (L2FC) of single predicted protease genes belonging to the different protease classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes, respectively. The examined conditions include two points in time during the exponential state (ex1, ex2) and one point in time for steady state.

To narrow down the number of differentially expressed genes and filter out the most important differentially expressed genes, predicted protease genes with L2FC >|2| were analyzed next. A gene was included, when L2FC >|2| at least at one condition during cultivation on xylose compared to the respective condition during cultivation on glucose was detected. In Table 5.33, the numbers of differentially expressed predicted protease genes that meet that criteria are shown. A further heatmap (Figure 5.88) visualizes the differential expression levels of these genes.

Table 5.33: Number of genes belonging to different protease classes that are differentially expressed (L2FC >|2|) in exponential state (ex1, ex2) and steady state (SS) during cultivation on xylose compared to the respective condition during cultivation on glucose.

		MJK20.3 xylose					
class	genes	ex1 up	ex1 down	ex2 up	ex2 down	SS up	SS down
aspartic peptidases	17	0	0	0	0	0	1
cysteine peptidases	56	6	0	6	0	2	0
glutamic peptidases	2	1	0	1	1	0	0
metallopeptidases	68	6	1	5	3	3	2
serine peptidases	75	7	11	6	13	3	2
threonine peptidases	19	0	0	0	0	0	0
total	237	20	12	18	17	8	5

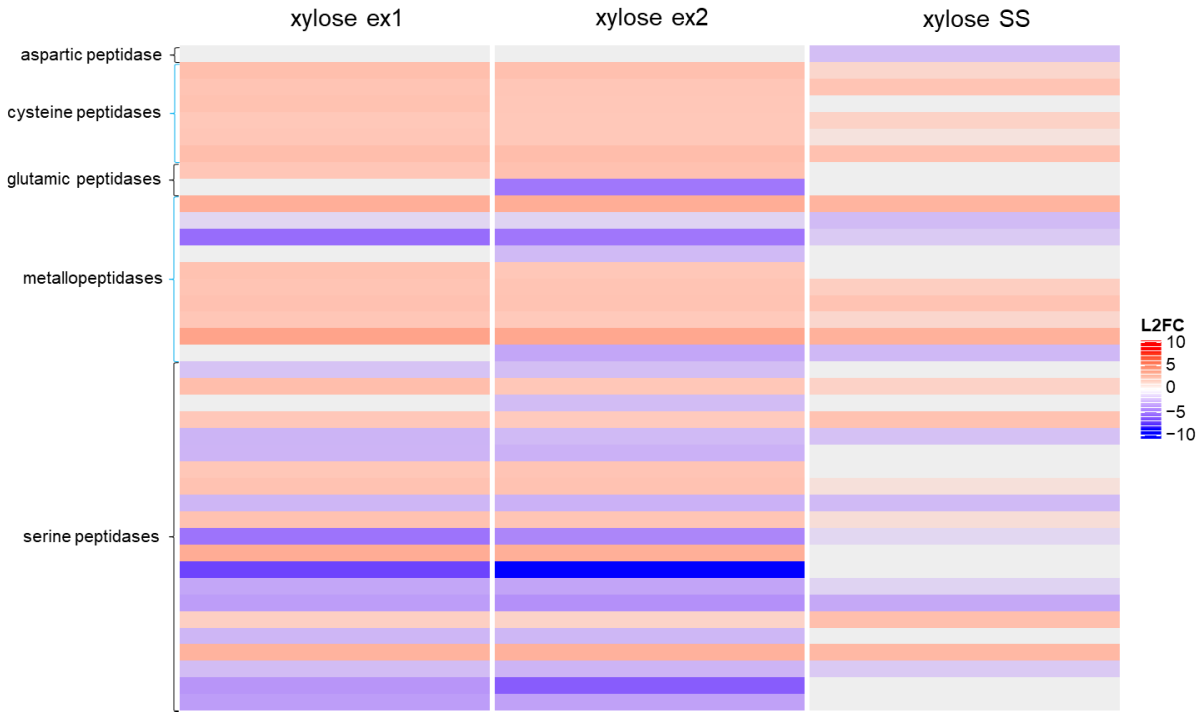

Figure 5.88: Heatmap with differential expression values of predicted protease genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. Shown are the log2 fold change values (L2FC) of single predicted protease genes belonging to the different protease classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes, respectively. The examined conditions include two points in time in exponential state (ex1, ex2) and one point in time in steady state.

The heatmap of the differentially expressed genes with L2FC>|2| confirms the results of the previous figures. The number of differentially expressed genes with L2FC >|2| is very low. The expression pattern between the exponential state samples is almost identical with some single genes that are stronger downregulated in ex2. Differentially expressed genes in steady state have almost identical L2FC values compared to the exponential state. The main difference between steady state and exponential state, is a much lower number of differentially expressed genes with L2FC >|2| during steady state. The most downregulated genes in exponential state can be found in the serine peptidases class and the most upregulated genes in the cysteine peptidases class. In steady state, predicted protease genes with differential expression are evenly distributed among the different classes.

To investigate and compare gene expression in detail, graphs showing L2FC values as well as graphs showing the mean of the normalized raw count values, were created as previously (see chapter 5.3.1.2) described (Figures 5.89-5.91). In the following the highly and differentially expressed proteases during growth on xylose in comparison to growth on glucose will be discussed briefly.

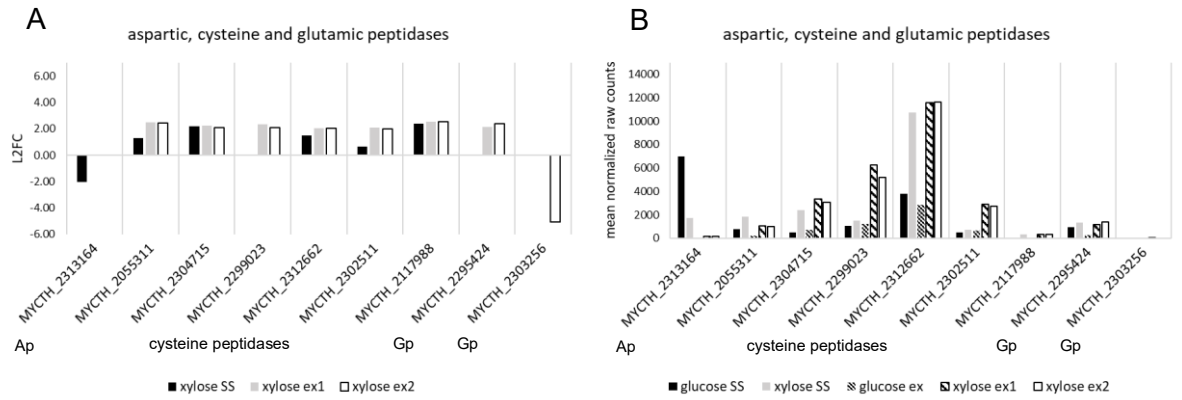

Figure 5.89: Differentially expressed predicted aspartic, cysteine and glutamic peptidase genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty). Ap= aspartic peptidase, Gp= glutamic peptidase.

In steady state, only the predicted aspartic peptidase gene MYCTH_2313164 is strongly expressed on glucose (Figure 5.89) and has only medium expression levels on xylose, while in exponential state no expression is observable. In contrast, all genes encoding predicted cysteine peptidases (Figure 5.89) have much higher expression levels (up to 4 x higher) on xylose in steady state as well as during the exponential state. The predicted cysteine peptidase gene with the highest expression levels on xylose in exponential state as well as in steady state is MYCTH_2312662. Generally, expression levels of predicted cysteine peptidase genes in exponential state are much higher as in steady state. Differences between the two exponential state samples are only for MYCTH_2304715 and MYCTH_2299023 observable, where expression levels are slightly lower in ex2.

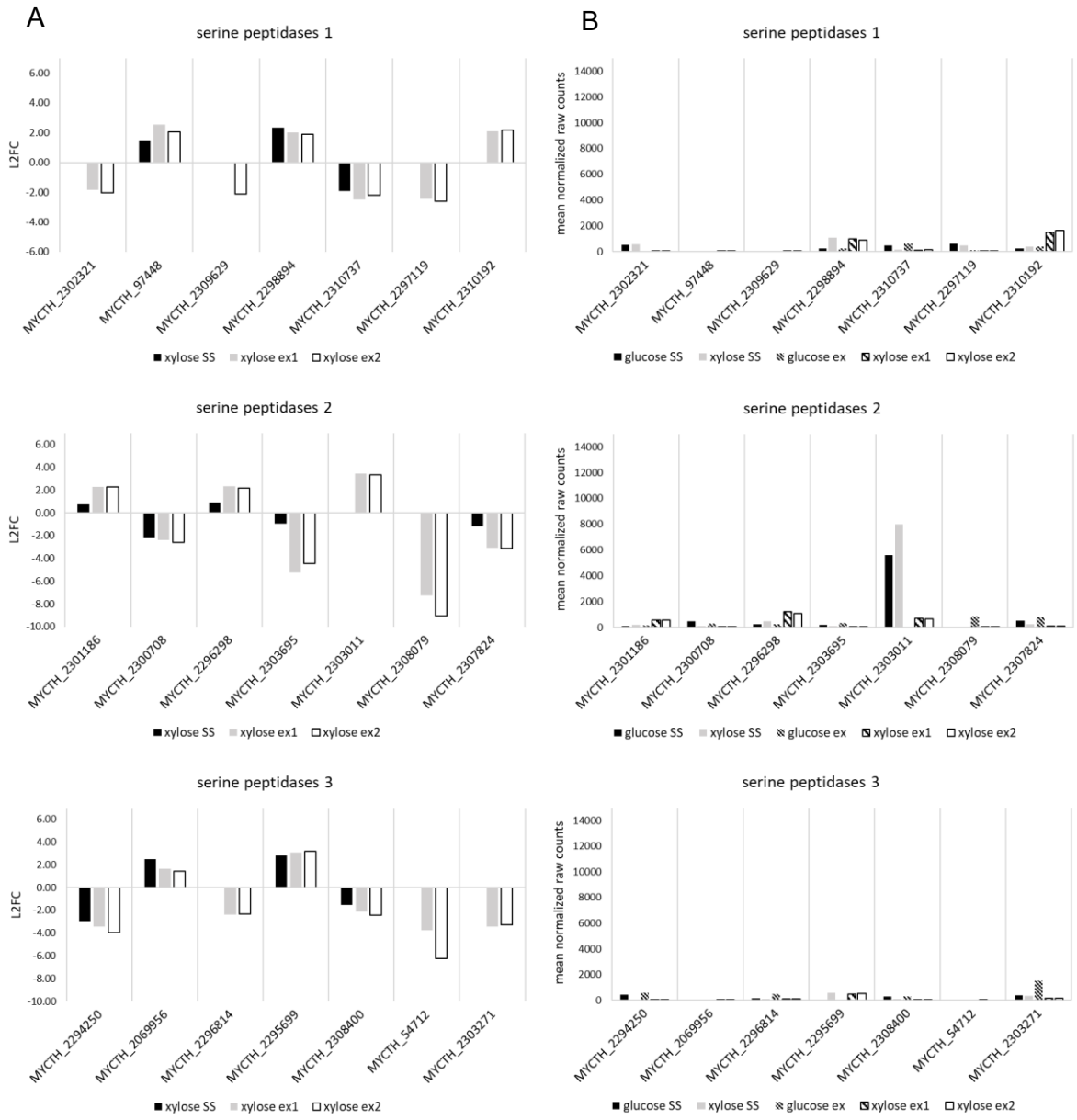

Figure 5.90: Differentially expressed predicted serine peptidase genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

In the serine peptidase category (Figure 5.90), many genes with low raw count values can be detected. Only MYCTH_2298894, MYCTH_2310192, MYCTH_2303011, and MYCTH_2303271 have medium to high expression levels for at least one condition. MYCTH_2298894 is much higher expressed on xylose in steady state as well as in exponential state, MYCTH_2310192 only in exponential state, MYCTH_2303011 only in steady state and MYCTH_2303271 only on glucose in exponential state. The predicted serine peptidase gene with the highest expression levels on xylose in steady state is MYCTH_2303011 and in exponential state MYCTH_2310192. Differences between the two exponential state samples are for MYCTH_2298894 (slightly lower in ex2) and MYCTH_2310192 (slightly higher in ex2) observable.

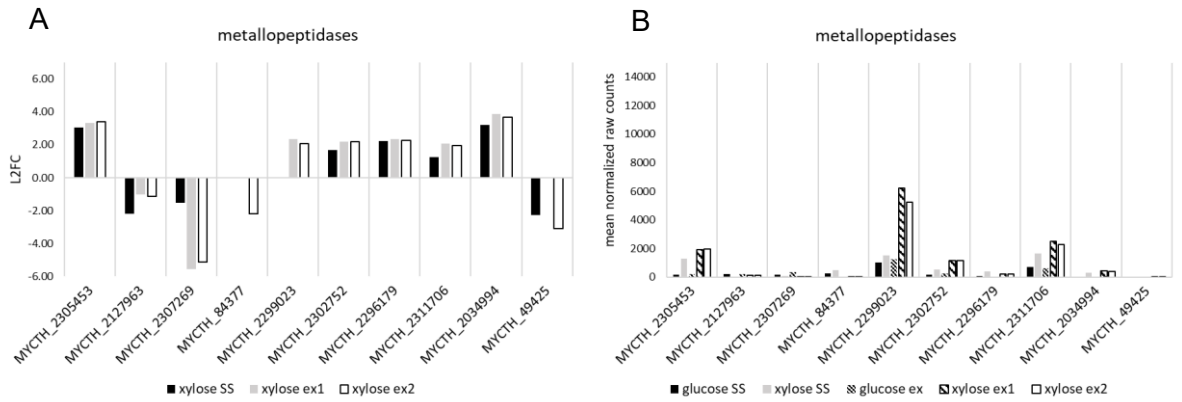

Figure 5.91: Differentially expressed predicted metallopeptidase genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

Regarding predicted metallopeptidase genes (Figure 5.91), only MYCTH_2299023, MYCTH_2311706, and MYCTH_2305453 have medium to high expression levels. All these genes are much higher expressed on xylose, compared to the respective condition on glucose, especially in the exponential state. The predicted metallopeptidase gene with the highest expression levels on xylose in steady state with relatively low expression levels on glucose is MYCTH_2311706. The same observations can be made for MYCTH_2299023 in exponential state. Differences between the two exponential state samples are for MYCTH_2299023 (lower in ex2) and MYCTH_2310192 (slightly lower in ex2) observable.

To identify further interesting highly expressed proteases, that were not included in the previous figures, the same graphs, as seen in Figures 5.89-5.91 were created using all genes encoding for proteases with a signal peptide. The results are shown in Figures 5.92 and 5.93.

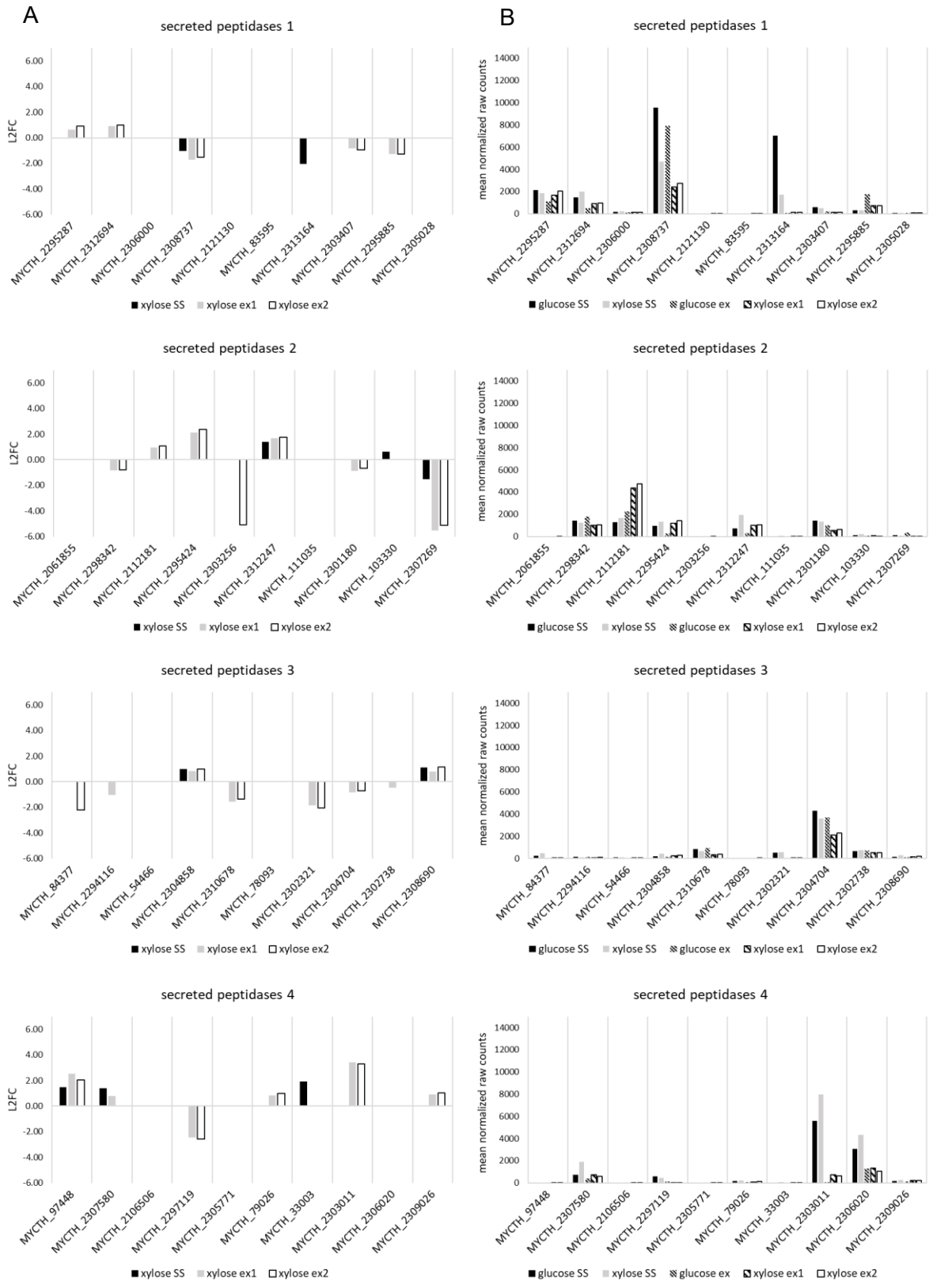

Figure 5.92: Differentially expressed genes encoding for predicted peptidases with a signal peptide (part 1) with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

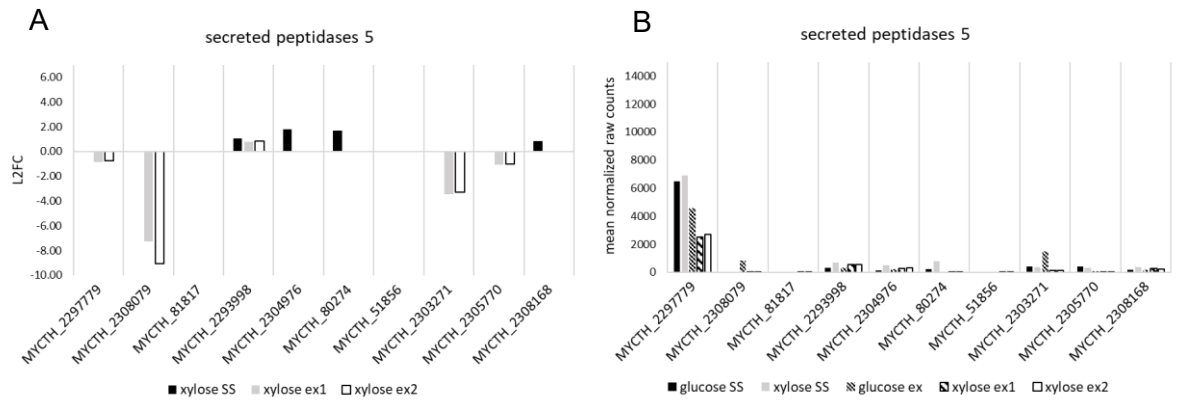

Figure 5.93: Differentially expressed genes encoding for predicted peptidases with a signal peptide (part 2) with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

The genes encoding for predicted peptidases with a signal peptide (Figures 5.92 and 5.93) include MYCTH_2308737, MYCTH_2313164, and MYCTH_2304704, which have high expression levels and are stronger expressed on glucose, whereas the genes MYCTH_2112181, MYCTH_2312247, MYCTH_2303011, and MYCTH_2306020 have high expression levels and are stronger expressed on xylose in exponential state and in steady state. MYCTH_2297779 is stronger expressed on xylose in steady state and on glucose in exponential state. Out of these genes, the highest expression levels could be observed for MYCTH_2297779, MYCTH 2303011, and MYCTH 2308737 on xylose in steady state and for MYCTH_2299023 and MYCTH_2112181 on xylose in exponential state. In ex2 expression values for almost all above-mentioned predicted protease genes, except MYCTH_2306020 are slightly higher.

In summary, some predicted protease genes are stronger expressed on glucose and some higher on xylose. Slightly more predicted protease genes with a high expression level have higher expression values during growth on xylose. As already mentioned, expression must be examined individually according to the strain and conditions that will be used in future experiments to choose candidates for deletion.

5.3.3.3 Differential expression of carbohydrate active enzymes (CAZYs)

The purpose of the investigation of CAZY expression has already been mentioned (see chapter 5.3.1.3). An overview of predicted *T. thermophilus* CAZYs was provided in this chapter as well. The total number of predicted CAZY genes that were differentially expressed at exponential state and steady state during cultivation on xylose is shown in Figure 5.94. Both exponential state samples, are very similar, with a slightly higher number of upregulated predicted CAZY genes and a slightly lower number of downregulated predicted CAZY genes in ex2 compared to ex1. In steady state, a much higher number of predicted CAZY genes is upregulated and a much lower number of predicted CAZY genes is downregulated compared to the exponential state samples. In total approximately 50 % of all predicted CAZY genes are differentially expressed in exponential state and 46 % in steady state. The relation of up- to downregulated genes is with roughly 3:1 much higher in steady state, compared to 1:1 in exponential state.

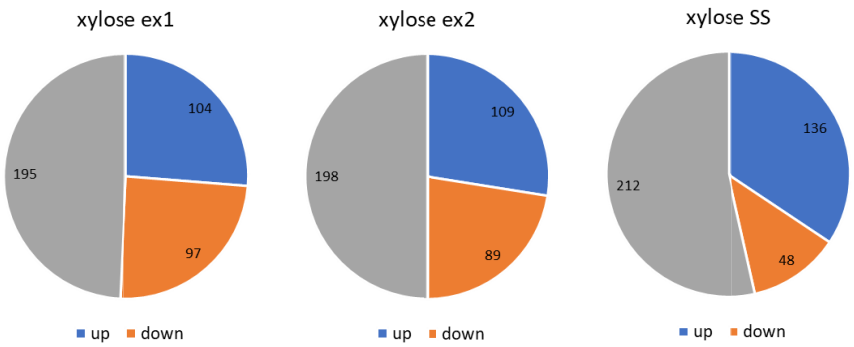

Figure 5.94: Number of differentially expressed predicted CAZY genes of strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. Shown are the numbers of up- (blue) and downregulated (orange) genes as well as genes with no differential expression (grey) in exponential state (ex1, ex2) and steady state (SS) compared to the respective condition during cultivation on glucose.

To get a detailed overview of the expression profiles of predicted CAZY genes on xylose, a heatmap (Figure 5.95) was created, showing the differential expression levels of the genes belonging to the different classes of CAZYs at exponential state and steady state during cultivation on xylose compared to the respective condition during cultivation on glucose. Generally, L2FC values are for most of the genes rather low. Differential expression in the exponential state samples is almost identical, except for some single genes that have higher or lower L2FC values in either ex1 or ex2. The expression profile in steady state is different compared to the exponential state. Except that, differential expression is mostly evenly distributed among the different classes. In steady state, a stronger upregulation across all classes, especially for predicted hemicellulase genes, is observable compared to the exponential state. In contrast, in exponential state, a stronger downregulation especially for predicted cell wall remodeling, hemicellulase and glycosyltransferase genes as well as for single genes belonging to the "other" class is detectable. The strongest differential expression in exponential state and in steady state can be observed for predicted hemicellulase and cellulase genes.

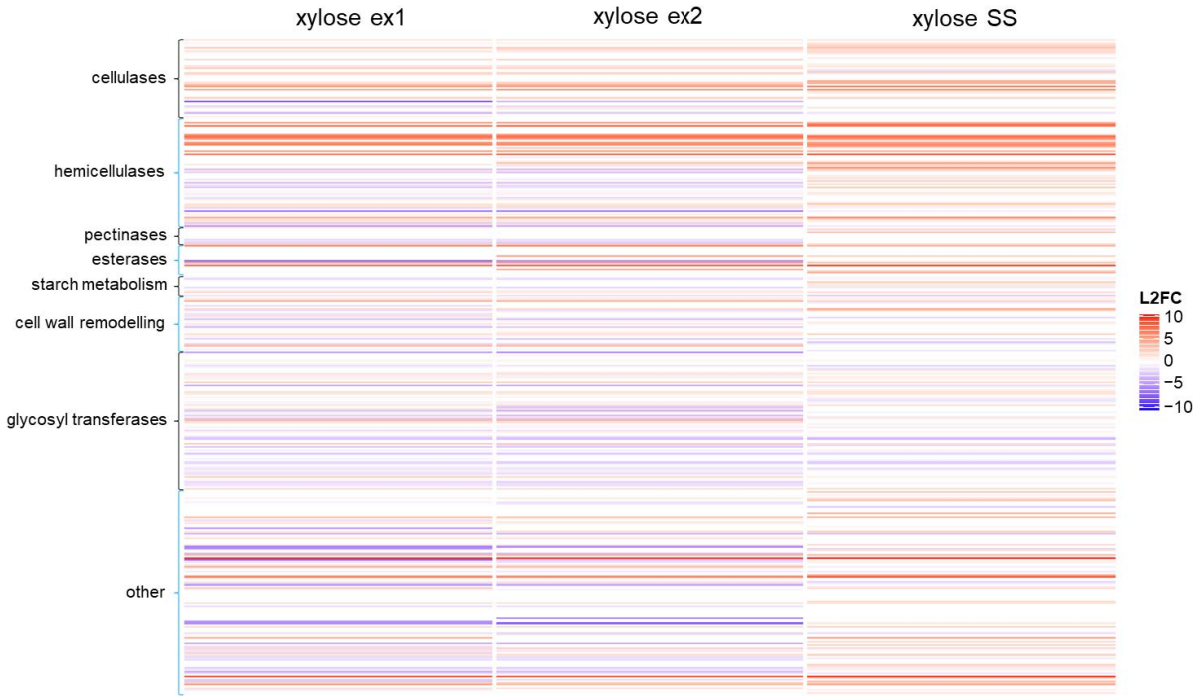

Figure 5.95: Heatmap with differential expression values of all predicted CAZY genes in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. Shown are the log2 fold change values (L2FC) of single predicted CAZY genes belonging to the different CAZY classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes, respectively. The examined conditions include two points in time in exponential state (ex1, ex2) and one point in time in steady state.

To narrow down the number of differentially expressed genes and filter out the most important differentially expressed genes, predicted CAZY genes with L2FC >|2| were analyzed next. A gene was included, when L2FC >|2| in at least one condition during cultivation on xylose compared to the respective condition during cultivation on glucose was detected. In Table 5.34, the numbers of differentially expressed genes encoding for predicted CAZYs that meet that criteria are shown. A further heatmap (Figure 5.96) visualizes the differential expression levels of those genes.

Table 5.34: Number of genes belonging to different CAZY classes that are differentially expressed (L2FC >|2|) in exponential state (ex1, ex2) and steady state (SS) during cultivation on xylose compared to the respective condition during cultivation on glucose.

			MJK20.3 xylose					
class	type	genes	ex1 up	ex1 down	ex2 up	ex2 down	SS up	SS down
cellulases	endoglucanases	9	1	0	1	0	2	0
	cellobiohydrolases	7	0	0	1	0	0	0
	ß-glucosidases	9	2	0	2	0	1	0
	LPMOs	24	4	2	5	1	4	0
hemicellulases	xylanases	12	5	0	5	0	6	0
	xylosidases	4	3	0	3	0	3	0
	endoarabinases	3	0	0	1	0	1	0
	exoarabinases/ arabinofuranosidases	11	2	0	3	0	5	0
	mixed-linked glucanase	5	0	1	0	2	1	0
	mannanases	10	0	2	0	1	2	0
	mannosidases	11	0	1	0	1	1	0
	galactanases	2	0	0	0	0	0	0
	galactosidases	7	1	2	1	1	2	0

pectinases	polygalacturonases	2	0	0	0	0	0	0
	rhamnosidases	1	0	0	0	0	0	0
	pectin lyases	8	1	0	1	1	2	0
esterases	feruloyl esterases	4	1	0	1	0	1	0
	acetyl esterases	9	2	1	3	1	3	0
	pectin esterases	4	0	0	1	0	2	0
	glycuronoyl esterases	2	0	0	0	0	0	0
starch metabolism	alpha amylases	4	0	0	0	0	1	0
	alpha glucosidases	4	0	0	0	0	0	0
	glucoamylases	2	0	0	0	0	0	0
	glycogen debranching enzymes	2	0	0	0	0	0	1
cell wall remodeling	glucanases	12	2	0	2	0	2	0
	transglucosylases	4	0	1	0	1	0	0
	chitosanases	2	0	0	0	0	0	0
	diacetylmuramidase	1	0	1	0	1	0	0
	glucosaminidase	3	0	0	0	0	0	0
	chitinases	8	1	0	1	1	1	1
	crosslinking transglycosidase	3	0	0	0	0	0	0
glycosyltransferases	glycosyltransferases	83	4	7	3	10	0	2
other	other	124	10	15	9	10	19	2
	total	396	39	33	43	31	59	6

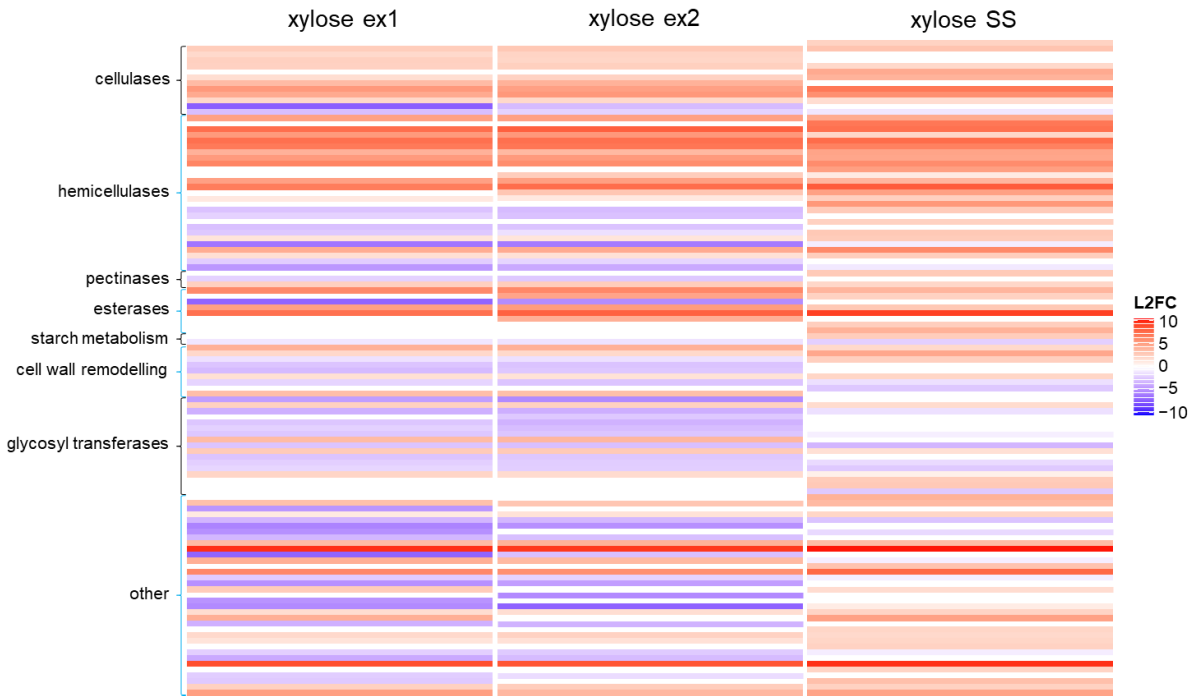

Figure 5.96: Heatmap with differential expression values of predicted CAZY genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. Shown are the log2 fold change values (L2FC) of single predicted CAZY genes belonging to the different CAZY classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes, respectively. The examined conditions include two points in time in exponential state (ex1, ex2) and one point in time in steady state.

The heatmap of the differentially expressed genes with L2FC>|2| confirms the results of the previous figures. The expression pattern between the exponential state samples is very similar with some genes, especially in the “other” category, less up- and downregulated in ex2 compared to ex1. Furthermore, some single genes have higher or lower differential expression values in either ex1 or ex2. In steady

state, genes that are differentially expressed, mostly have identical differential expression trends compared to the respective differentially expressed genes in exponential state. Nevertheless, differential expression is stronger for most of the genes in steady state. The main difference between steady state and exponential state, that can be observed here, is the much higher number of upregulated genes and the much lower number of downregulated genes with L2FC >|2| in steady state. The relation of up- to downregulated genes is as already seen in previous figures much higher in SS compared to the exponential state. Downregulated genes in exponential state are evenly distributed among the different CAZY classes and the most and strongest upregulated genes can be found in the hemicellulases and cellulases classes as well as in single genes of the “other” class. In steady state the most and strongest upregulated genes can also be found in the hemicellulases and cellulases classes as well as in single genes of the “other” class. Downregulation is here only for single genes in the categories starch metabolism, cell wall remodeling, glycosyltransferases, and “other” observable.

To investigate and compare gene expression in detail, graphs showing L2FC values as well as graphs showing the mean of the normalized raw count values created as previously (see chapter 5.3.1.2) described (Figures 5.97-5.104). The expression of predicted CAZY genes in MJK20.3 during growth on glucose, especially the top genes of each CAZY class have already been discussed in a previous chapter. Focus of the following is on describing the differences in expression between cultivation on glucose and xylose especially for the medium and highly expressed predicted CAZY genes of each class. If not mentioned specifically, expression of genes is regarded as similar.

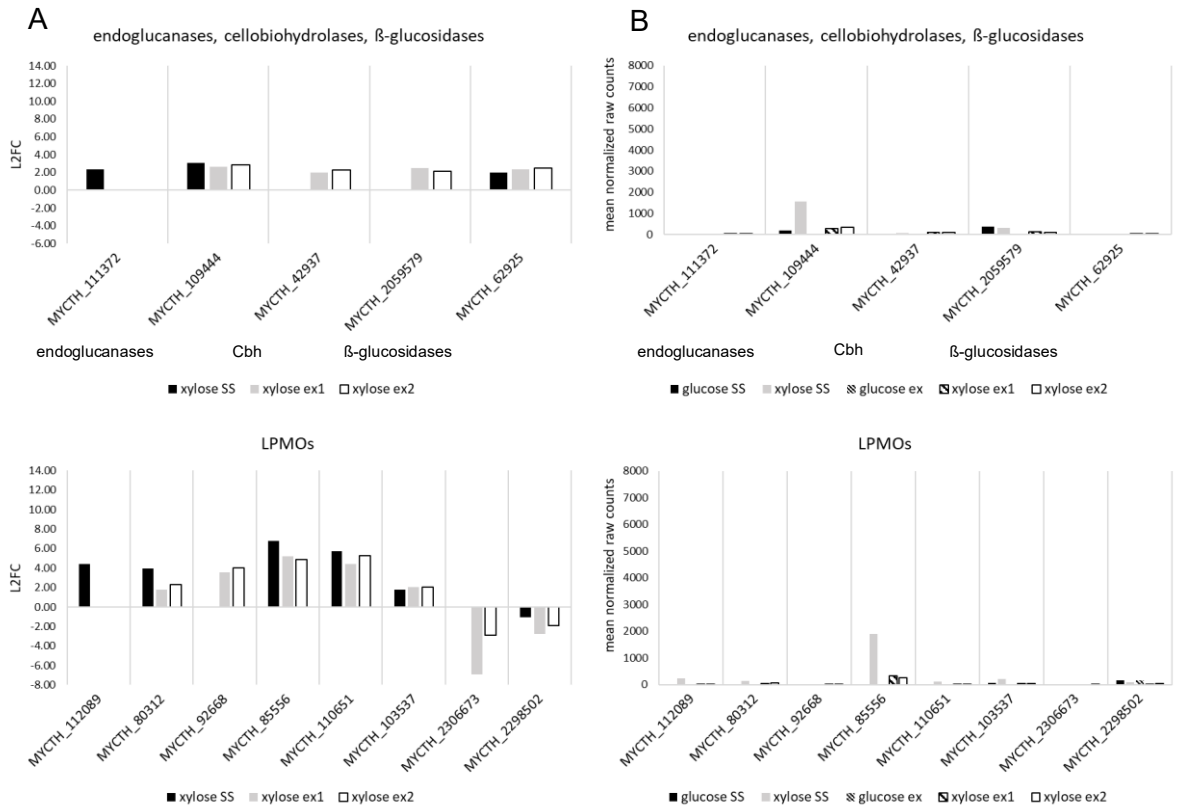

Figure 5.97: Differentially expressed predicted cellulase genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty). Cbh= cellobiohydrolase.

Among the predicted cellulase genes (Figure 5.97), only the predicted endoglucanase gene MYCTH_109444 and the predicted LPMO gene MYCTH_85556 are highly expressed. This is only the case during growth on xylose in steady state. At the respective condition during growth on glucose, almost no expression is detectable for these genes. All other genes have extremely low expression levels and are, therefore, not discussed.

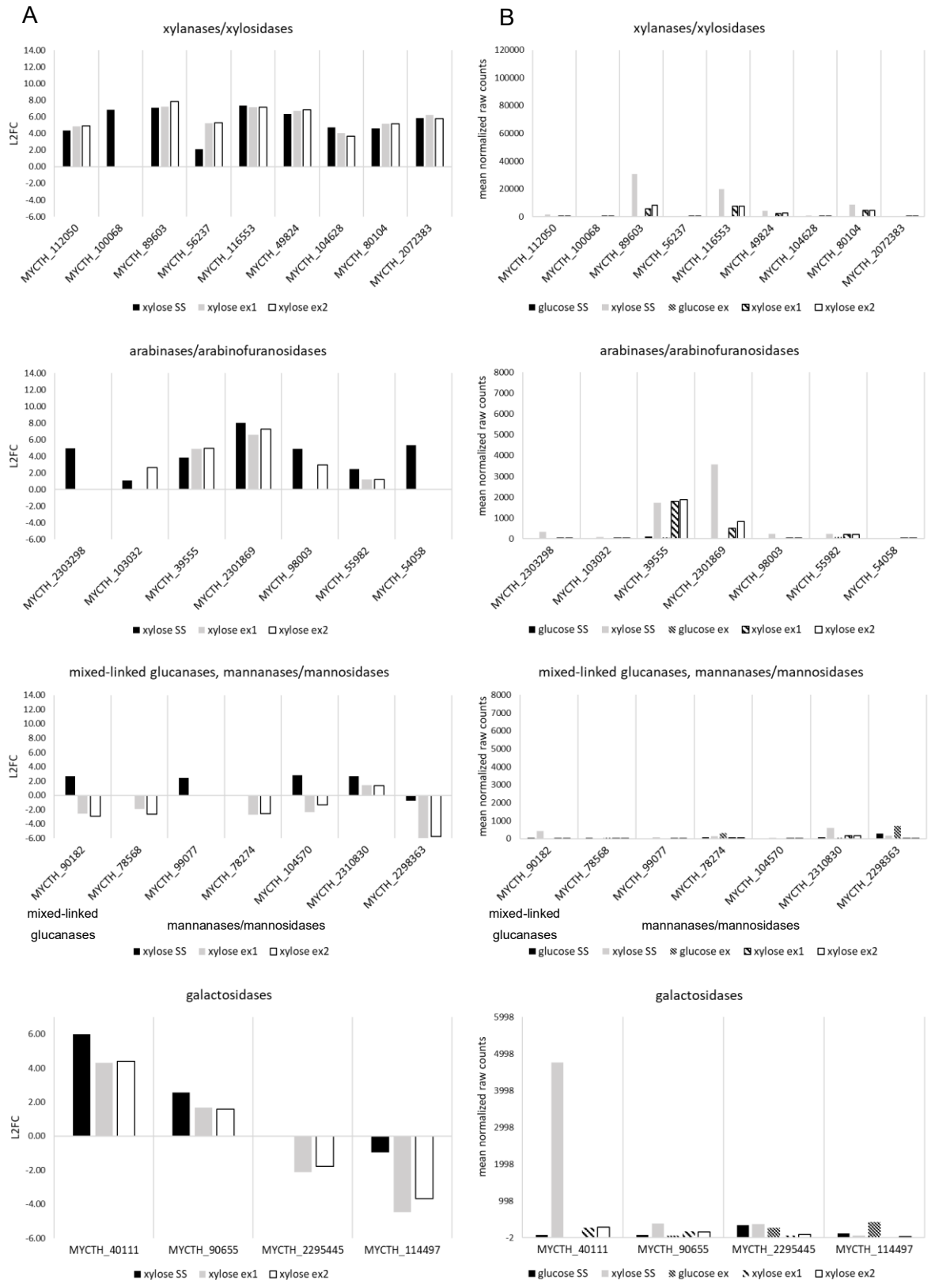

Figure 5.98: Differentially expressed predicted hemicellulase genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

Among the genes encoding for predicted hemicellulases (Figure 5.98), also many genes with low raw count values can be detected. Only the predicted xylanase/xylosidase genes MYCTH_89603, MYCTH_116553, MYCTH_80104, and MYCTH_49824, the predicted arabinase/arabinofuranosidase genes MYCTH_39555 and MYCTH_2301869, and the predicted galactosidase gene MYCTH_40111 have medium to high expression levels at least at one condition. The predicted xylanase/xylosidase genes are much higher expressed on xylose during steady state compared to the exponential state. They are not expressed on glucose. The predicted arabinase/arabinofuranosidase gene MYCTH_39555 is equally expressed in steady state and exponential state and the predicted arabinase/arabinofuranosidase gene MYCTH_2301869 has much higher expression levels in steady state during growth on xylose. For both of the genes almost no expression on glucose is observable. Finally, during growth on xylose, the predicted galactosidase gene MYCTH_40111 is highly expressed in steady state with no expression on glucose and almost no expression in exponential state. The predicted hemicellulase genes with the highest expression levels on xylose in steady state and in exponential state are the above-mentioned predicted xylanase/xylosidase genes. Differences between the two exponential state samples among the highly expressed genes are only for MYCTH_89603 (higher in ex2) and MYCTH_39555 (slightly higher in ex2) observable.

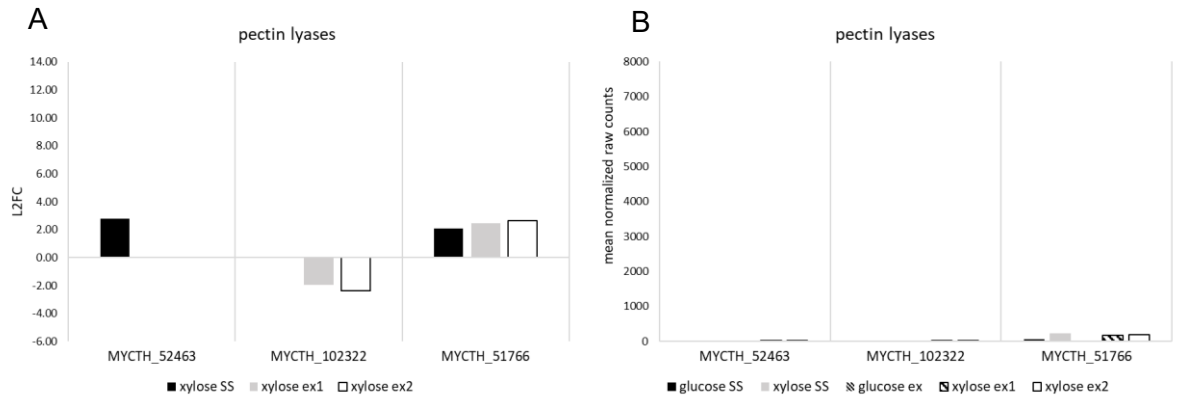

Figure 5.99: Differentially expressed predicted pectinase genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

For all predicted pectinase genes (Figure 5.99) only extremely low expression levels are detectable.

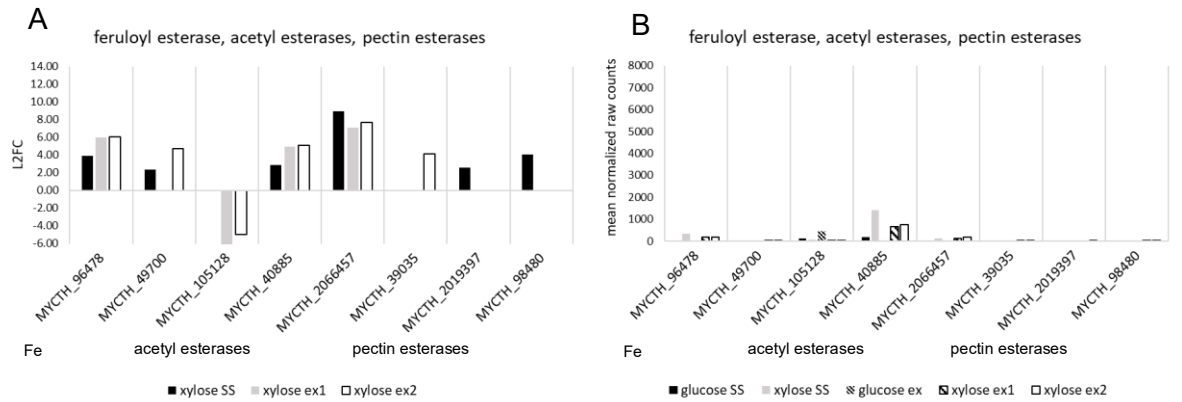

Figure 5.100: Differentially expressed predicted esterase genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty). Fe= feruloyl esterase.

The only predicted esterase gene (Figure 5.100) with medium to high expression levels at least at one condition is the predicted acetyl esterase gene MYCTH_40885. The expression is much higher during steady state compared to exponential state during growth on xylose, on glucose almost no expression is detectable.

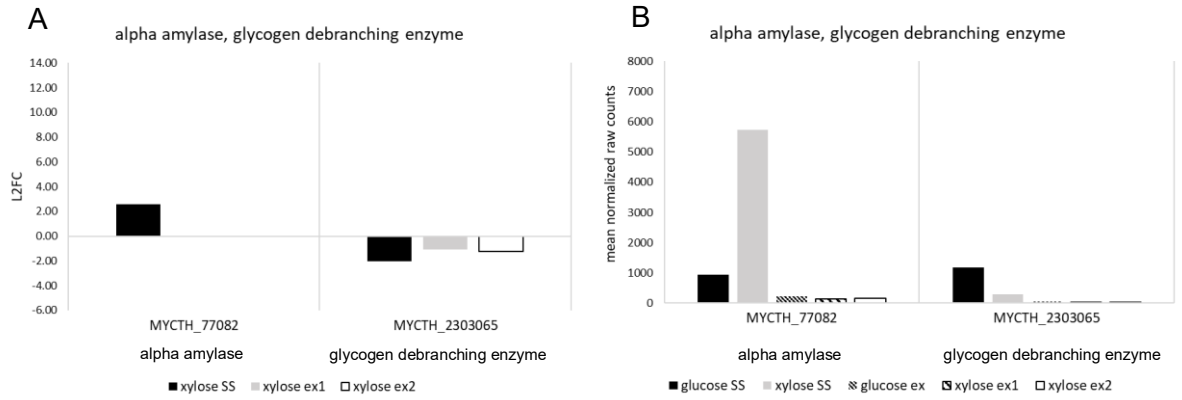

Figure 5.101: Differentially expressed predicted starch metabolism genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

Within the starch metabolism category (Figure 5.101), only the predicted alpha amylase gene MYCTH_77082 is highly expressed in steady state. Here, expression values during growth on xylose are much higher compared to the respective value during growth on glucose. At all other conditions, almost no expression is observable. The predicted glycogen debranching enzyme gene MYCTH_2303065 has medium expression levels only in steady state during growth on glucose. At all other conditions, almost no expression is observable.

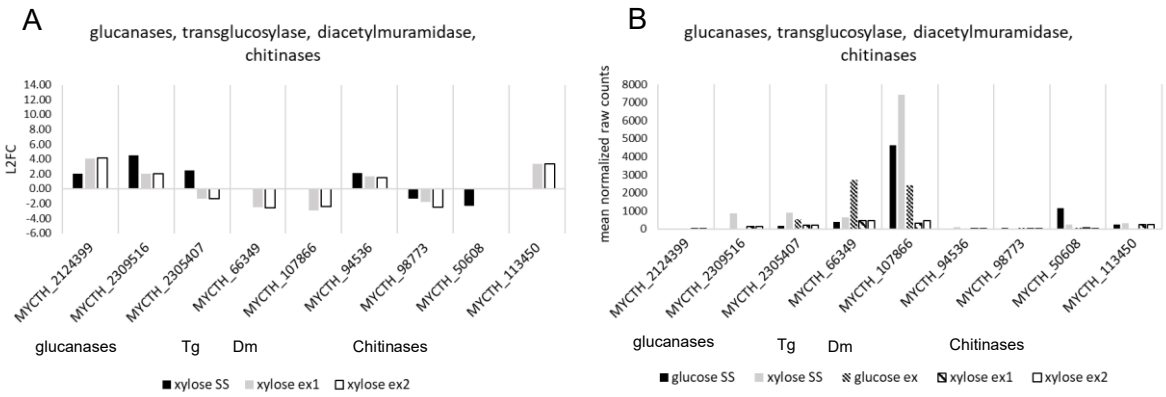

Figure 5.102: Differentially expressed predicted cell wall remodeling genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty). Tg= transglucosylase, Dm= diacetylmuramidase.

Regarding the predicted cell wall remodeling genes (Figure 5.102), the predicted transglucosylase gene MYCTH_66349 has high expression levels only during exponential state when grown on glucose and the predicted diacetylmuramidase gene MYCTH_107866 has higher expression levels in steady state (higher during growth on xylose) compared to the exponential state. In exponential state high expression is only observed during growth on glucose with almost no expression on xylose for this gene.

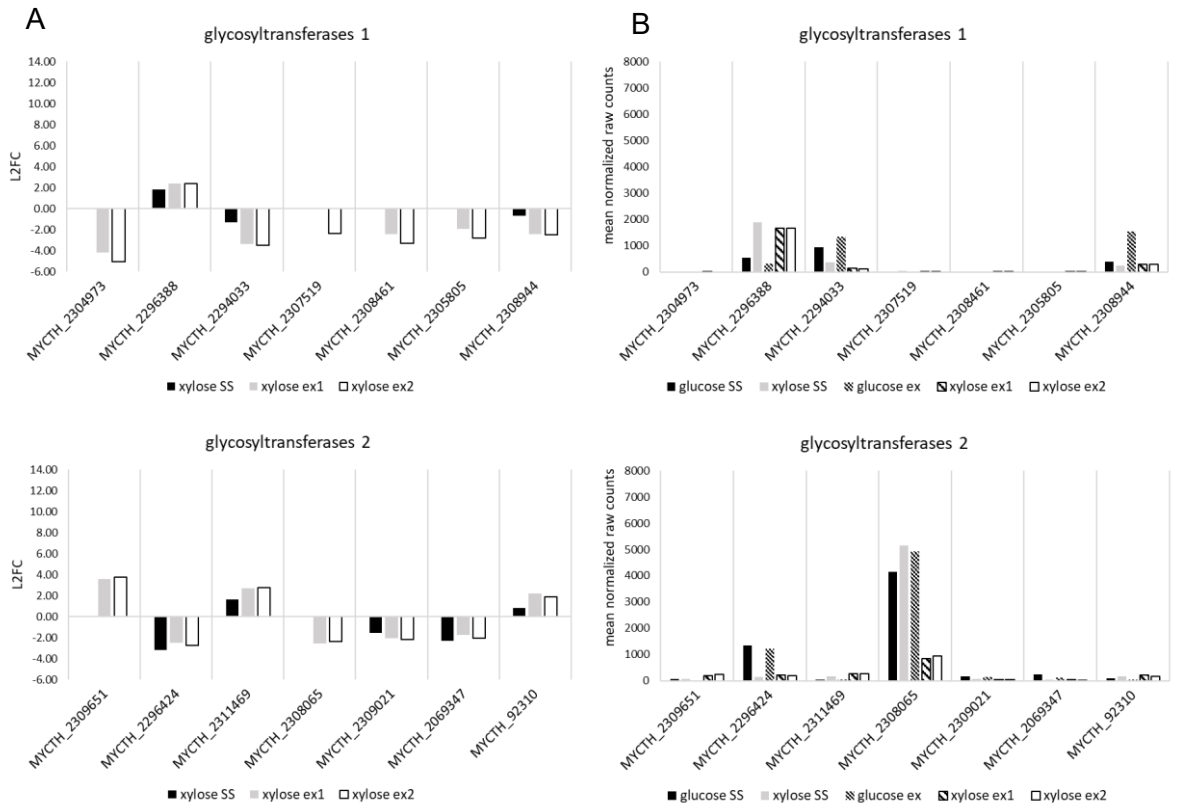

Figure 5.103: Differentially expressed predicted glycosyltransferase genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

Among the predicted glycosyltransferase genes (Figure 5.103) that have medium to high expression levels at least at one condition are MYCTH_2296388, MYCTH_2294033, MYCTH_2308944, MYCTH_2308065, and MYCTH_2296424. The genes MYCTH_2294033 and MYCTH_2296424 have a much stronger expression in both exponential state and steady state during growth on glucose, whereas for MYCTH_2308944 only in exponential state high expression, which is much stronger during growth on glucose, can be observed. MYCTH_2296388 is stronger expressed during growth on xylose in both exponential and steady state (slightly higher in steady state) with low expression levels on glucose. Finally, MYCTH_2308065 is highly expressed in steady state during growth on xylose and glucose (higher for xylose) and in exponential state only during growth on glucose (slightly lower compared to steady state on xylose). The predicted glycosyltransferase gene with the highest expression levels on xylose in steady state is MYCTH_2308065 and in exponential state MYCTH_2296388. Differences between the two exponential state samples among the highly expressed genes are not observed.

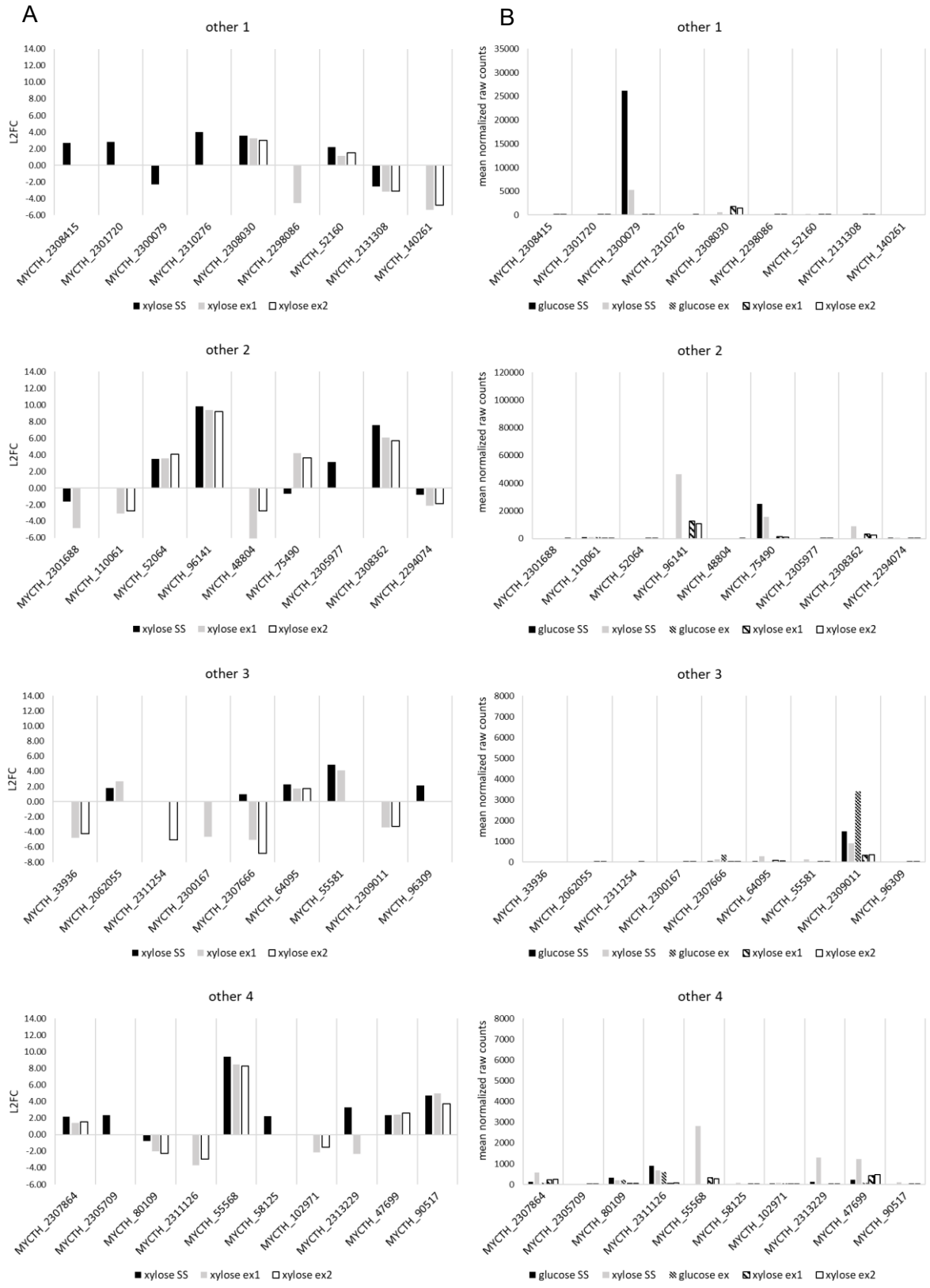

Figure 5.104: Differentially expressed "other" predicted CAZY genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

Genes belonging to the “other” category have already been categorized in a previous chapter (see Table 5.10 chapter 5.3.1.3) and are described in the following according to this characterization. Within this category (Figure 5.104) only the genes MYCTH_2300079, MYCTH_96141, MYCTH_75490, MYCTH_2308362, MYCTH_2309011, MYCTH_55568, MYCTH_2313229, and MYCTH_47699 have medium to high expression levels at least at one condition. The predicted hemicellulase gene MYCTH_55568, predicted starch degrading LPMO gene MYCTH_2313229, and predicted pectinase gene MYCTH_47699 only have high expression levels in steady state during growth on xylose. At all other conditions, no or only a very low expression can be observed. For the predicted hemicellulase gene MYCTH_96141 expression levels are extremely high (much higher in steady state compared to exponential state) during growth on xylose with no expression during growth on glucose. This is also the case for the predcited hemicellulase gene MYCTH_2308362, but with much lower expression levels. The predicted trehalase gene MYCTH_75490 is only expressed in steady state (much higher during growth on glucose) and the predicted cellulase gene MYCTH_2309011 is in steady state higher expressed and in exponential state much higher expressed during growth on glucose. Finally, the predicted hemicellulase gene MYCTH_2300079 is extremely strong expressed in steady state during growth on glucose with much lower expression levels on xylose, while no expression in exponential state during growth on both carbon sources can be observed. The gene with the highest expression levels on xylose in steady state as well as in exponential state is by far the predicted hemicellulase gene MYCTH_96141. Differences between the two exponential state samples among the highly expressed genes are only seen in MYCTH_96141, where expression is slightly lower in ex2.

In summary, differences between the cultivation on xylose and glucose can be mainly found in the expression of predicted hemicellulase genes, as well as for single genes belonging to the cellulase category, especially for the predicted LPMO gene MYCTH_85556 and the predicted endoglucanase gene MYCTH_109444.

5.3.3.4 Differential expression of transcription factors of plant biomass degradation

The purpose of the investigation of the expression of transcription factors has already been mentioned (see chapter 5.3.1.4). An overview of predicted *T. thermophilus* transcription factors was provided in this chapter as well.

The total number of predicted transcription factor genes that were differentially expressed at exponential state and steady state during cultivation on xylose is shown in Figure 5.105. Both exponential state samples, are very similar, with a slightly higher number of downregulated and a slightly lower number of upregulated predicted transcription factor genes in ex2 compared to ex1. In steady state, a lower number of predicted transcription factor genes are up- and downregulated compared to the exponential state samples. In total approximately 55 % of all predicted transcription factor genes are differentially expressed in exponential state and 45 % in steady state. The relation of up- to downregulated genes is with roughly 1:1.25 in steady state and exponential state very similar.

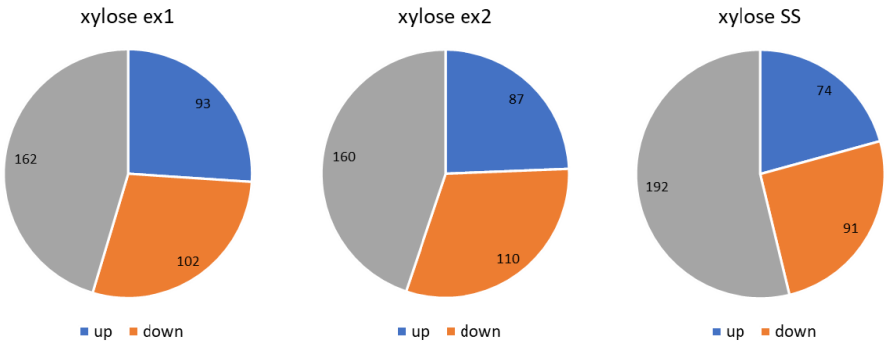

Figure 5.105: Number of differentially expressed predicted transcription factor genes of strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. Shown are the numbers of up- (blue) and downregulated (orange) genes as well as genes with no differential expression (grey) in exponential state (ex1, ex2) and steady state (SS) compared to the respective condition during cultivation on glucose.

To get a detailed overview of the expression profiles of the predicted transcription factors genes on xylose, a heatmap (Figure 5.106) was created, showing the differential expression levels of genes belonging to the different classes of transcription factors at exponential state and steady state during cultivation on xylose compared to the respective condition during cultivation on glucose. Generally, L2FC values are for most of the genes rather low. Differential expression between exponential state samples is almost identical, except for some single genes that have higher or lower L2FC values in either ex1 or ex2. The expression profile in steady state is different compared to the exponential state. Here, the lower number and weaker differential expression of up- and downregulated genes across all classes is, as already seen in Figure 5.105, observable compared to the exponential state. Nevertheless, the genes that are differentially expressed are mostly showing the same expression trend compared to the exponential state. In contrast, in exponential state, especially a stronger downregulation of genes belonging to the fungal Zn(2)-Cys(6) binuclear cluster domain, zinc finger, C2H2 type and fungal specific transcription factor domain classes as well as for single genes in the “other” class is detectable. Except for that differential expression is evenly distributed among the different classes.

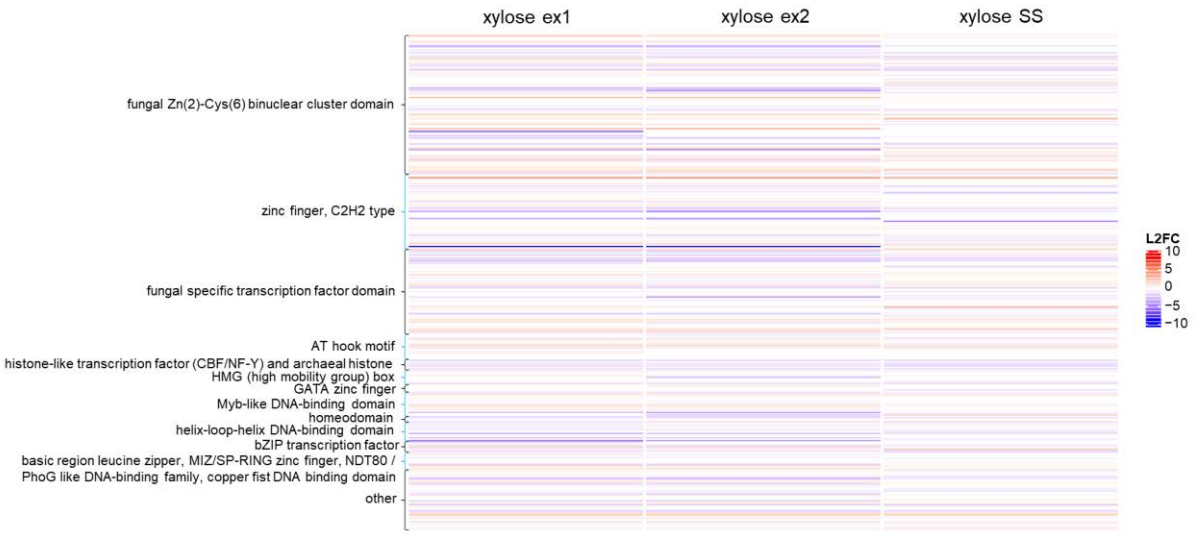

Figure 5.106: Heatmap with differential expression values of all predicted transcription factor genes in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. Shown are the log2 fold change values (L2FC) of single predicted transcription factor genes belonging to the different transcription factor classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes, respectively. The examined conditions include two points in time in exponential state (ex1, ex2) and one point in time in steady state.

To narrow down the number of differentially expressed genes and filter out the most important ones, predicted transcription factors genes with L2FC >|2| were analyzed next. A gene was included, when L2FC >|2| at least at one condition during cultivation on xylose compared to the respective condition during cultivation on glucose was detected. In Table 5.35, the numbers of differentially expressed predicted transcription factor genes that meet that criteria are shown. A further heatmap (Figure 5.107) visualizes the differential expression levels of these genes.

Table 5.35: Number of genes belonging to different transcription factor classes that are differentially expressed (L2FC >|2|) in exponential state (ex1, ex2) and steady state (SS) during cultivation on xylose compared to the respective condition during cultivation on glucose.

		MJK20.3 xylose					
class	genes	ex1 up	ex1 down	ex2 up	ex2 down	SS up	SS down
fungal Zn(2)-Cys(6) binuclear cluster domain	101	5	7	3	7	2	0
zinc finger, C2H2 type	53	2	3	1	3	1	2
fungal specific transcription factor domain	61	2	3	1	6	3	0
AT hook motif	18	1	0	0	0	0	0
histone-like transcription factor (CBF/NF-Y) and archaeal histone	7	0	0	0	0	0	0
HMG (high mobility group) box	10	0	0	0	1	0	0
GATA zinc finger	6	0	0	0	0	0	0
Myb-like DNA-binding domain	17	0	1	0	1	0	0
homeodomain	5	0	0	0	0	0	0
helix-loop-helix DNA-binding domain	13	0	3	0	3	0	0
bZIP transcription factor	8	0	1	0	0	1	0
basic region leucine zipper	4	0	0	0	0	0	0
MIZ/SP-RING zinc finger	3	0	0	0	0	0	0
NDT80 / PhoG like DNA-binding family	3	1	0	0	0	0	0
copper fist DNA binding domain	2	0	0	0	0	0	0
other	46	1	3	2	2	1	1
total	357	12	21	7	23	8	3

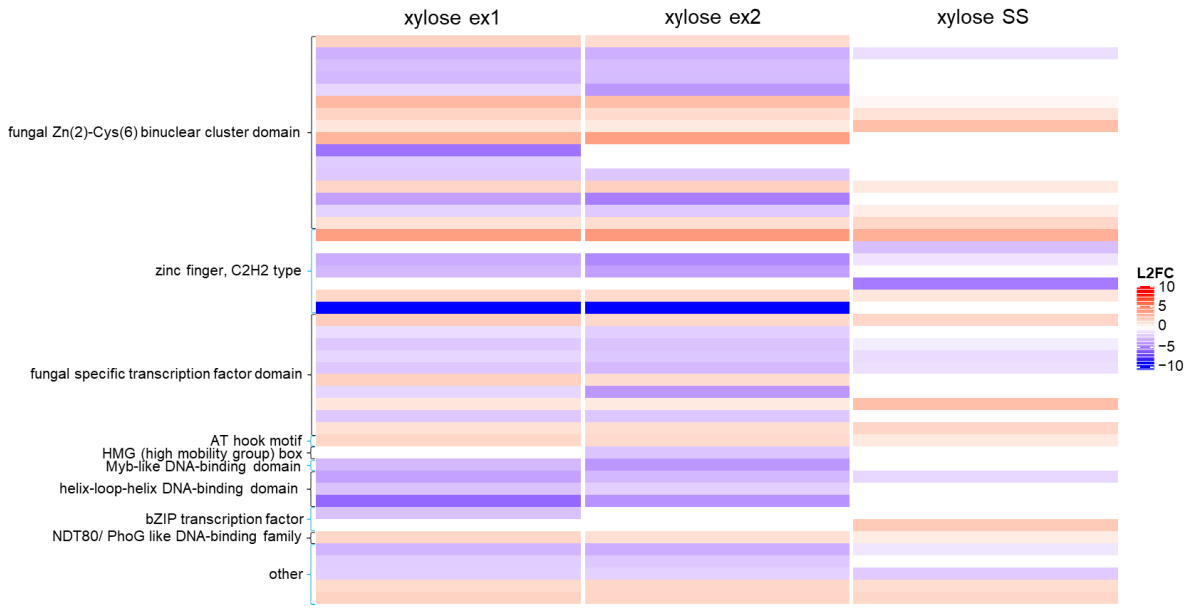

Figure 5.107: Heatmap with differential expression values of predicted transcription factor genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. Shown are the log2 fold change values (L2FC) of single predicted transcription factor genes belonging to the different transcription factor classes as a color scale. Negative values (blue) represent downregulated genes and positive values (red) upregulated genes, respectively. The examined conditions include two points in time in exponential state (ex1, ex2) and one point in time in steady state.

The heatmap of the differentially expressed genes with L2FC>|2| confirms the results of the previous figures. The number of differentially expressed genes with L2FC >|2| is very low in general. The expression pattern between the exponential state samples is almost identical, some genes are stronger up- or downregulated in ex1 or ex2, respectively. In steady state, expression of genes that are differentially expressed, have mostly the same expression trend compared to the exponential state. The main difference between steady state and exponential state is a much lower number of differentially expressed genes, especially the downregulated genes with L2FC >|2| during steady state. In exponential state more genes are downregulated than upregulated. Differentially expressed genes are evenly distributed among the different classes in exponential state as well as in steady state.

To get an idea on the possible function of these genes, the protein sequences of the transcription factors with L2FC >|2| during growth on xylose were blasted as previously described (see chapter 5.3.1.4). The descriptions of the blast hits with the lowest E-value or the names of some respective regulator orthologs as well as the general expression trends were already provided in Tables 5.13 and 5.24. Interesting genes that first appeared during the following analysis are shown in Table 5.36.

Table 5.36: Medium and highly expressed predicted transcription factor genes that are among the interesting genes in the following analysis and have not been described yet. Shown are the descriptions of the best protein blast hits (lowest E-values), performed using the NCBI database. The color of the gene indicates the general expression trend (up- or downregulation) during growth on xylose compared to growth on glucose in steady state (color of MYCTH_) and exponential state (color of number); green= upregulated, red= downregulated, black= no differential expression.

gene	description best blast hits
MYCTH_2303596	sterol uptake control protein 2/beauvericin cluster specific repressor
MYCTH_2310586	hypothetical
MYCTH_2089183	Nit4/NirA
MYCTH_73708	Moc3/control of sterol uptake
MYCTH_2306627	transcriptional activator Dal81
MYCTH_2301577	hypothetical
MYCTH_92449	hypothetical
MYCTH_2297012	all development altered
MYCTH_2304200	centromere binding protein
MYCTH_2311995	transcription initiation factor

To investigate and compare gene expression in detail, graphs showing L2FC values as well as graphs showing the mean of the normalized raw count values were created as previously (see chapter 5.3.1.2) described (Figures 5.108-5.111). The expression of predicted transcription factor genes in MJK20.3 during growth on glucose, especially the top genes of each transcription factor class have already been discussed in a previous chapter. Focus of the following is on describing the differences in expression between cultivation on glucose and xylose especially for the medium and highly expressed predicted transcription factor genes of each class. If not mentioned specifically, expression of genes is regarded as similar.

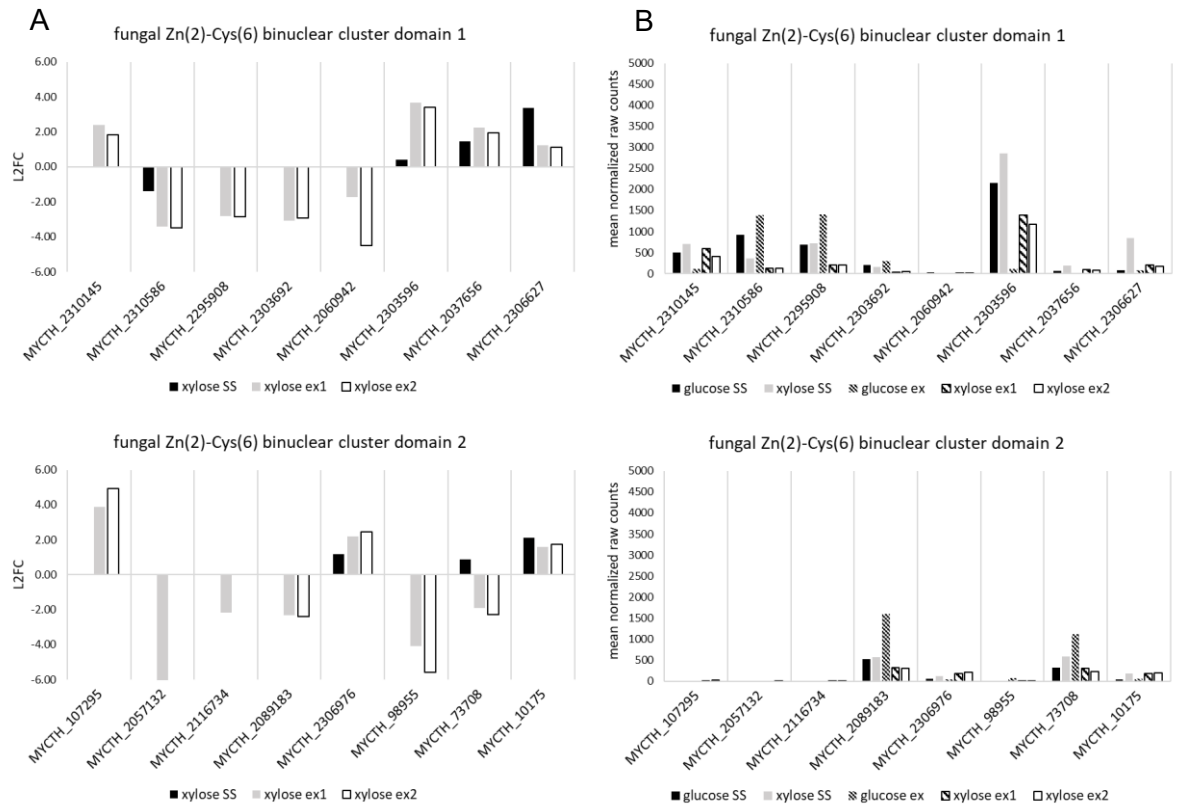

Figure 5.108: Differentially expressed genes encoding for predicted transcription factors with fungal Zn(2)-Cys(6) binuclear cluster domain with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

Among the genes encoding for predicted transcription factors with fungal Zn(2)-Cys(6) binuclear cluster domain (Figure 5.108), only MYCTH_2303596 (sterol uptake control protein 2/beauvericin cluster specific repressor), MYCTH_2295908 (Pro1), MYCTH_2310586 (hypothetical), MYCTH_2089183 (Nit4/NirA), MYCTH_73708 (Moc3/control of sterol uptake), and MYCTH_2306627 (transcriptional activator Dal81) are medium or highly expressed at least at one condition. For MYCTH_2303596 (sterol uptake control protein 2/beauvericin cluster specific repressor) expression in exponential and steady state during growth on xylose is higher compared to growth on glucose, whereas in steady state expression on glucose is still high and in exponential state almost no expression on glucose can be identified. Regarding MYCTH_2306627 (transcriptional activator Dal81), medium expression levels can be observed only in steady state during growth on xylose. At all other conditions, almost no expression is observed. For MYCTH_2295908 (Pro1) and MYCTH_2089183 (Nit4/NirA) low expression levels in steady state (similar for glucose and xylose) and high expression levels in exponential state during growth on glucose can be observed (very low expression levels on xylose). MYCTH_2310586 (hypothetical) has medium to high expression levels (higher in exponential state) only during growth on glucose. Finally, MYCTH_73708 (Moc3/control of sterol uptake) has medium expression levels only on glucose in exponential state. At all other conditions, only very low expression is observed. The gene

encoding for a predicted transcription factor with fungal Zn(2)-Cys(6) binuclear cluster domain with the highest expression level on xylose in steady state and in exponential state is MYCTH_2303596 (sterol uptake control protein 2/beauvericin cluster specific repressor). Differences between the two exponential state samples among the medium to highly expressed genes are only seen for MYCTH_2303596 (slightly lower in ex2).

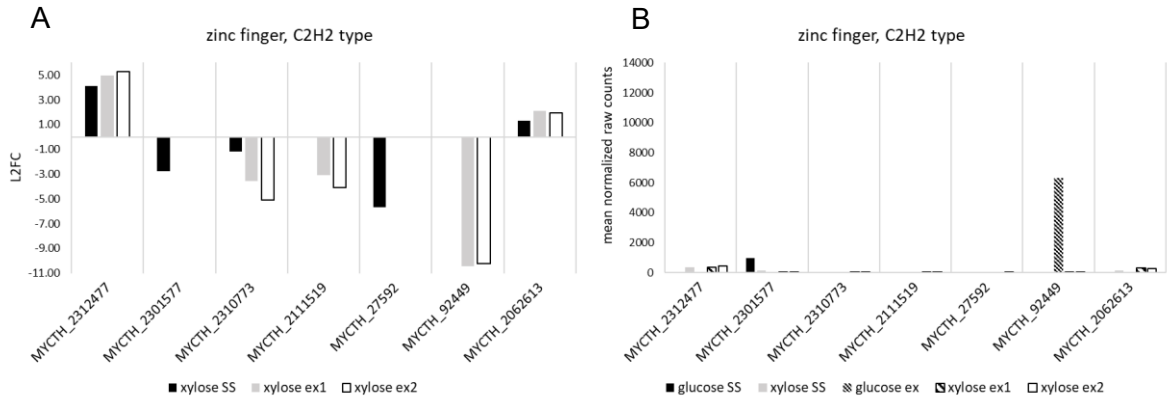

Figure 5.109: Differentially expressed genes encoding for predicted zinc finger, C2H2 type transcription factors with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

The only genes encoding for predicted zinc finger, C2H2 type transcription factors (Figure 5.109) that have medium to high expression levels at least at one condition are MYCTH_2301577 (hypothetical) and MYCTH_92449 (hypothetical). The gene MYCTH_2301577 (hypothetical) has medium expression levels only during growth on glucose in steady state and MYCTH_92449 (hypothetical) has very high expression levels but only in exponential state during growth on glucose. At all other conditions, almost no expression is observable.

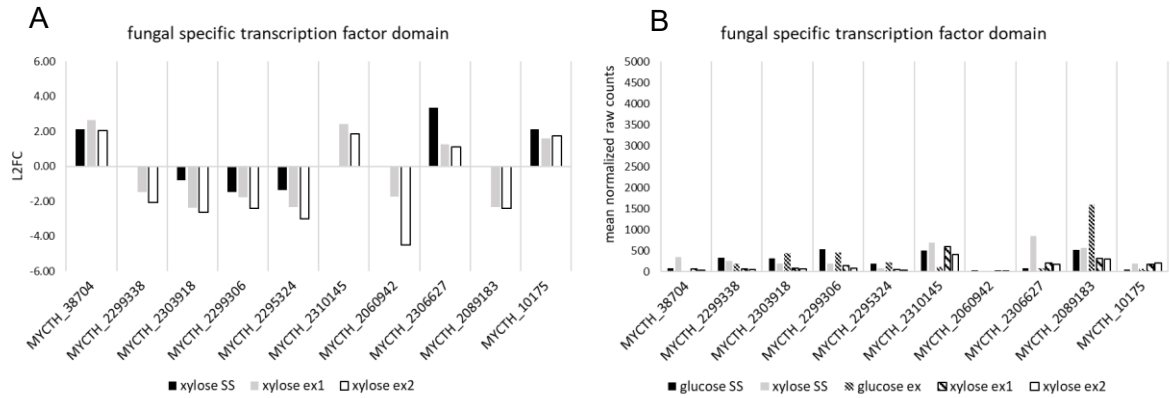

Figure 5.110: Differentially expressed genes encoding for predicted transcription factors with fungal specific transcription factor domain with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty).

Among the genes encoding for predicted transcription factors with fungal specific transcription factor domain (Figure 5.110) only MYCTH_2306627 (transcriptional activator Dal81) and MYCTH_2089183 (Nit4/NirA), which have already been discussed among the genes encoding for predicted transcription factors with fungal Zn(2)-Cys(6) binuclear cluster domain, have medium to high expression levels at least at one condition.

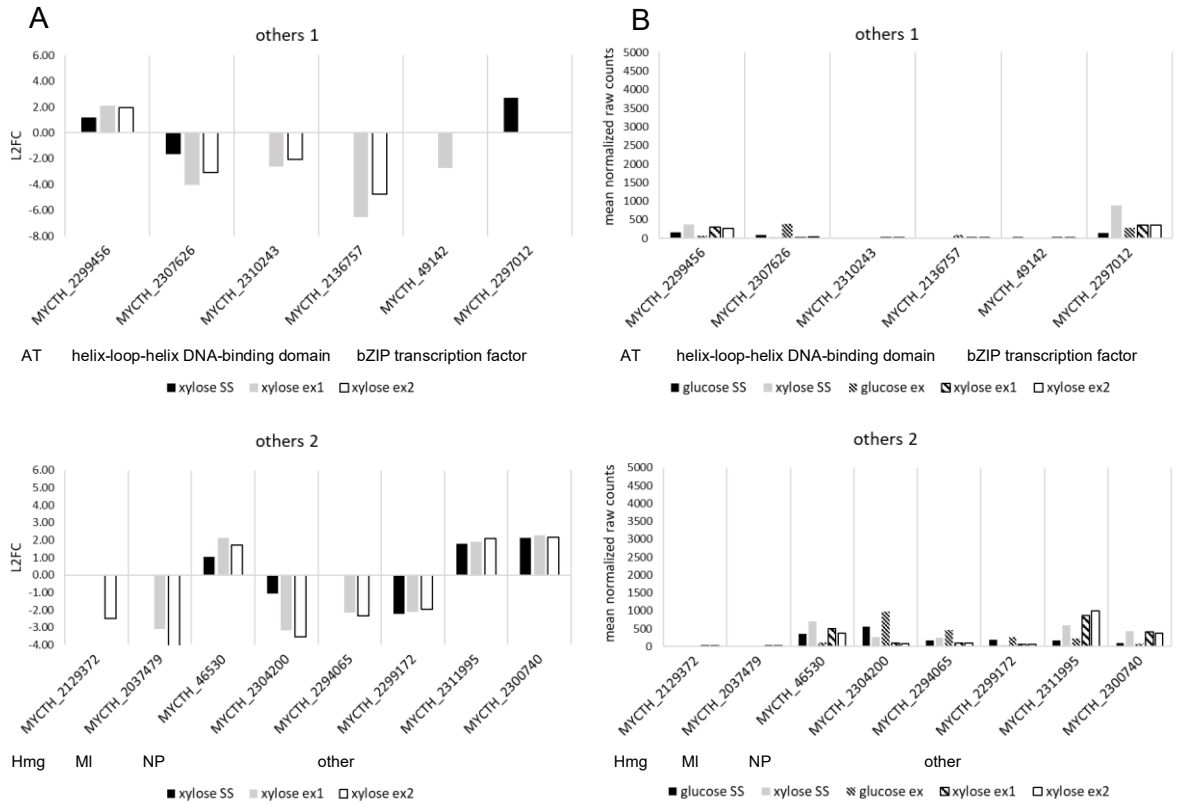

Figure 5.111: Differentially expressed other predicted transcription factors genes with L2FC >|2| in strain MJK20.3 during cultivation on xylose compared to the respective condition during cultivation on glucose. (A) Log2 fold change values (L2FC) of the single genes according to the respective condition: xylose SS (black, filled), xylose ex1 (grey), and xylose ex2 (black, empty). (B) Mean of the normalized counts from the two replicates of the single genes according to the respective condition: glucose SS (black, filled), xylose SS (grey), glucose ex (small diagonal stripes), xylose ex1 (big diagonal stripes), and xylose ex2 (black, empty). AT= AT hook motif, Hmg= high mobility group box, MI= Myb-like DNA-binding domain, NP= NDT80/PhoG like DNA binding family.

Finally, in the “others” category (Figure 5.111), only the genes MYCTH_2297012 (all development altered), MYCTH_2304200 (centromere binding protein), and MYCTH_2311995 (transcription initiation factor) have medium expression levels at least at one condition. For MYCTH_2297012 (all development altered) medium expression levels can only be detected in steady state during growth on xylose and in MYCTH_2304200 (centromere binding protein) only in exponential state during growth on xylose. At all other conditions, expression levels are very low. MYCTH_2311995 (transcription initiation factor) is stronger expressed during growth on xylose, with medium expression levels only during exponential state (slightly higher in ex2).

Besides the comparison of the expression differences during growth on glucose or xylose according to the transcription factor classes and best blast hits, as done above, also the expression differences of genes encoding for orthologs of the known regulators of plant biomass degradation in comparison to the reference strain reveal valuable information about regulation processes regarding the different conditions. Therefore, expression levels of genes encoding for orthologs of regulators with medium to high expression levels, that were already described in chapter 5.3.1.4 were also analyzed among the dataset obtained from cultivations on xylose. The L2FC values as well as the mean of the normalized

raw counts at the respective conditions are shown in Figure 5.112. If not mentioned specifically, expression of genes is regarded as similar.

		L2FC xylose vs. glucose			mean normalized raw counts				
					steady state		exponential state		
gene	ortholog	SS	ex1	ex2	glucose	xylose	glucose	xylose 1	xylose 2
MYCTH_2028011	Ace1	0.00	0.00	0.00	1572	1565	1028	1159	1176
MYCTH_2308260	Ace2	0.00	0.00	0.00	378	377	588	456	515
MYCTH_2301920	AmyR/BglR*	0.00	1.73	1.75	5142	5404	826	2735	2772
MYCTH_2307451	Ap3	0.35	0.00	0.00	304	388	447	471	540
MYCTH_113457	AreA/Nir2/Nit2	0.00	0.00	0.00	803	865	595	576	571
MYCTH_2309867	AreB	0.00	0.00	0.00	527	684	664	770	804
MYCTH_2063030	ClbR/ClbR2/ClbR3	0.00	0.00	0.00	489	392	437	336	339
MYCTH_2298863	Clr1*	0.00	0.00	0.00	2270	1799	330	256	248
MYCTH_38704	Clr2*	2.13	2.64	2.03	80	346	10	62	40
MYCTH_2306730	ClrC	-1.87	-2.95	-3.07	156	43	132	17	16
MYCTH_2296492	Clr4*	0.00	0.00	0.00	1294	1140	1091	1043	1039
MYCTH_2310085	Cre1*	0.00	0.50	0.50	2245	1823	950	1346	1346
MYCTH_2055311	CreB	1.27	2.48	2.42	775	1865	187	1049	1001
MYCTH_2306444	CreC	0.00	-0.79	-0.92	499	527	542	314	284
MYCTH_2306452	CreD	0.00	0.00	0.49	365	341	580	678	817
MYCTH_59287	Crz1	-1.53	-1.32	-1.29	1361	471	773	309	316
MYCTH_46266	GaaR/Pdr2	0.61	0.00	0.00	826	1259	189	181	154
MYCTH_46981	GaaX	1.47	1.27	1.07	261	720	136	329	286
MYCTH_2310995	HacA/1	0.00	-0.58	0.00	3522	5034	5415	3620	4312
MYCTH_2297059	Hap2	-1.21	-0.99	-1.28	1333	575	1066	535	440
MYCTH_41855	Hap3	-1.33	-0.85	-0.67	1153	459	1139	630	714
MYCTH_67051	Hap5	-1.37	-1.35	-1.42	1365	526	1618	633	604
MYCTH_2309600	Hcr1	0.00	0.00	0.00	476	374	384	338	343
MYCTH_2295635	Hep1/HP1	-1.95	-1.07	-1.11	94	24	130	62	60
MYCTH_2308921	Ire1	-0.79	-1.01	-1.12	1517	878	1052	522	483
MYCTH_2294559	Lae1/A	0.00	0.00	0.00	213	115	197	97	104
MYCTH_2303067	MalR	0.00	-1.27	-1.39	191	146	181	75	69
MYCTH_2132441	McmA/1	-0.45	-0.86	-1.00	4639	3401	5620	3087	2808
MYCTH_2303918	Mhr1	-0.81	-2.35	-2.63	319	182	437	86	70
MYCTH_2298994	NirA/Nit4	1.15	1.29	0.87	386	853	177	430	320
MYCTH_2302460	NmrA/1	0.79	1.30	1.16	598	1037	476	1174	1064
MYCTH_81165	PacC/1	0.00	0.71	0.66	1211	1345	794	1298	1257
MYCTH_2303559	Prk6	0.00	0.00	0.00	192	210	258	320	337
MYCTH_2300719	Rca1	0.00	0.69	0.59	1650	1414	389	626	588
MYCTH_2298696	Rce1	0.00	0.00	0.00	1025	963	619	584	582
MYCTH_2302052	Res1	-0.53	0.00	0.00	916	635	545	444	427
MYCTH_53224	RhaR/Pdr1	0.00	0.00	0.00	225	280	112	145	138
MYCTH_2306768	Sah2	0.00	0.00	0.00	730	780	834	682	620
MYCTH_2297068	Stk12	4.10	3.71	3.55	144	2447	88	1146	1029
MYCTH_2312657	VeA/Vel1	-1.24	-0.60	-0.62	828	351	219	145	142
MYCTH_113912	VelB	-1.84	-1.17	-1.23	343	96	234	104	100
MYCTH_46530	Vib1	1.02	2.11	1.70	350	712	117	508	382
MYCTH_108157	Vib2	-0.56	0.00	0.00	825	561	351	400	388
MYCTH_2309330	Wc1/Blr1	0.00	0.00	-0.49	588	540	365	307	259
MYCTH_2294022	Wc2/Blr2	0.00	-1.54	-1.30	556	643	534	184	217
MYCTH_2310145	Xyr1*	0.00	2.40	1.84	494	696	113	597	404
MYCTH_2144297	Xpp1	0.00	-1.03	-0.98	263	202	852	418	432

Figure 5.112: Differential expression of genes encoding for orthologs of known regulators of plant biomass degradation in *T. thermophilus* during growth on xylose compared to growth on glucose. Shown are the log2 fold change values (L2FC) of differentially expressed genes during growth on xylose compared to growth on glucose as well as the mean of the normalized raw counts in steady state and exponential state during growth on glucose and xylose. Green color indicates an upregulation and red color indicates a downregulation. The dash separates the possible orthologs of this regulator if more than one was found. An asterisk marks the regulators that have already been investigated in *T. thermophilus*.

The genes encoding for orthologs of known regulators of plant biomass degradation that have medium to high expression levels and are upregulated during growth on xylose in comparison to growth on glucose in steady state include MYCTH_2055311 (CreB), MYCTH_46266 (GaaR/Pdr2), MYCTH_2302460 (NmrA/1), and MYCTH_2297068 (Stk12). Out of these, the genes with the strongest expression are: MYCTH_2055311 (CreB) and MYCTH_2297068 (Stk12).

Genes that have medium to high expression levels and are upregulated during growth on xylose in comparison to growth on glucose in exponential state are MYCTH_2301920 (AmyR/BglR), MYCTH_2310085 (Cre1), MYCTH_2055311 (CreB), MYCTH_2302460 (NmrA/1), MYCTH_81165 (PacC/1), and MYCTH_2297068 (Stk12). Out of these, the regulators with the strongest expression are: MYCTH_2301920 (AmyR/BglR) and MYCTH_2310085 (Cre1).

Genes with medium to high expression levels that are downregulated during growth on xylose in comparison to growth on glucose in steady state include: MYCTH_59287 (Crz1), MYCTH_2297059 (Hap2), MYCTH_41855 (Hap3), MYCTH_67051 (Hap5), MYCTH_2308921 (Ire1), MYCTH_2132441 (McmA/1), MYCTH_2302052 (Res1), MYCTH_2312657 (VeA/Vel1), and MYCTH_108157 (Vib2). Out of these, the genes with the strongest downregulation are: MYCTH_59287 (Crz1), MYCTH_2297059 (Hap2), MYCTH_41855 (Hap3), and MYCTH_67051 (Hap5).

Genes with medium to high expression levels that are downregulated during growth on xylose in comparison to growth on glucose in exponential state include MYCTH_2310995 (HacA/1), MYCTH_2297059 (Hap2), MYCTH_41855 (Hap3), MYCTH_67051 (Hap5), MYCTH_2308921 (Ire1), and MYCTH_2132441 (McmA/1). Out of these, the gene with the strongest downregulation is MYCTH_67051 (Hap5).

Genes with medium to high expression levels that show no differential expression in steady state during growth on xylose compared to growth on glucose include: MYCTH_2028011 (Ace1), MYCTH_2301920 (AmyR/BglR), MYCTH_2298863 (Clr1), MYCTH_2296492 (Clr4), MYCTH_2310085 (Cre1), MYCTH_2310995 (HacA/1), MYCTH_81165 (PacC/1), MYCTH_2300719 (Rca1), and MYCTH_2298696 (Rce1). Out of these, the ones with the highest expression are: MYCTH_2301920 (AmyR/BglR), MYCTH_2310995 (HacA/1), MYCTH_2298863 (Clr1), and MYCTH_2310085 (Cre1).

Genes with medium to high expression levels that show no differential expression in both exponential state samples during growth on xylose compared to growth on glucose include: MYCTH_2028011 (Ace1) and MYCTH_2296492 (Clr4) with equal expression levels.

All other, not separately mentioned genes encoding for orthologs of known regulators of plant biomass degradation have either low L2FC or normalized raw count values or both and will, therefore, not be mentioned.

In summary, genes putatively involved in sterol metabolism, nitrogen metabolism, development, and chromatin accessibility are differentially expressed during growth on xylose, whereas differences in expression levels are relatively low between expression on glucose compared to growth on xylose. Regarding genes encoding for orthologs of known regulators of plant biomass degradation, most notable differences between growth on glucose and xylose are the expression of MYCTH_2297059 (Hap2), MYCTH_41855 (Hap3), MYCTH_67051 (Hap5), MYCTH_2310085 (Cre1; exponential state only), MYCTH_2301920 (AmyR/BglR; exponential state only), MYCTH_2055311 (CreB), MYCTH_2132441 (McmA/1), MYCTH_2297068 (Stk12), and MYCTH_2302460 (NmrA/1).

6. Discussion

In this part the obtained results will be discussed. At first, the establishment of molecular biological methods and bioreactor cultivations is discussed. Afterwards, the transcriptomic analysis of the reference strain and all deletion mutants including protease, CAZY and transcription factor gene expression in response to cellulose and the respective differential expression during growth on xylose in comparison to growth on glucose are discussed separately. The results of the biochemical analysis and bioreactor cultivations are integrated in this part. Finally, the physiological response to a pH change is discussed.

6.1 Establishment and application of molecular biological methods

Deletions via split marker approach and CRISPR/Cas12a based approach were very effective in combination with the transformation into the Δ*ku80* background. Deletions with the split marker approach were always successful in the first transformation round except for the deletion of *clr1* (*clr1*: 0/5; *clr2*: 6/11; *vib2*: 4/12; data not shown). The CRISPR/Cas12a based approach was successfully used for *clr1*, *clr4, vib1*, and *pyr5* targeting (*clr1*: 5/5; *clr4*: 9/9; *vib1*: 5/6; *pyr5*: 2/2; data not shown) illustrating the high transformation efficiency while using CRISPR/Cas12a in comparison to the split marker approach. For the deletion of *clr1* transformation efficiency increases while using the CRISPR/Cas12a based approach. Further direct comparisons between split marker and CRISPR/Cas12a based approach are not possible due to the different targets. The establishment of the Tet-On/Off system in *T. thermophilus* failed. The reason for this is most likely the fact that the used integration mechanism according to Arentshorst et al. 2015 is not suitable for *T. thermophilus*. When the plasmid is integrated, a functional and a non-functional *pyr5* gene are resulting, which are flanking the Tet-On/Off system. Due to the homology, those two genes could recombine which causes the plasmid including the Tet-On/Off system to be removed while only a functional or non-functional *pyr5* gene remains. The described problem is known, but is not observed frequently in *A. niger* (Arentshorst et al. 2015; Wanka et al. 2016). The observations in *T. thermophilus* fit to the results of the luciferase assays and diagnostic PCRs, where the wildtype strain or mixed clones could be detected (data not shown). Furthermore, this matches the observations made during the deletion of the cellulase regulators. Here it was observed that the loop-out of *amdS* in many cases started already without the addition of FAA (data not shown). The recommendations for future experiments are therefore the integration of the Tet-On/Off system without the possibility of its removal via recombination of surrounding flanks.

6.2 Establishment of bioreactor cultivations

A detailed description of conditions and media used for bioreactor cultivations of *T. thermophilus*, especially chemostat cultivations was not available in literature. Within this work, protocols to cultivate *T. thermophilus* were established and reproducibly applied. This enables investigation of *T. thermophilus* under industrially relevant conditions. Several unexpected problems occurred during the establishment (data not shown). If $CaCl_2$ concentration is too high, sporulation will be observed when cultivating at 37 °C, which is not the optimum growth temperature of 45 to 50 °C described for *T. thermophilus* (Van

Oorschot 1977; Emalfarb et al. 1998; Maheshwari et al. 2000; Visser et al. 2011). The induction of sporulation via $CaCl_2$ was already observed in other fungi like *Penicillium notatum* and *Blastocladiella emersonii* (Hadley and Harrold 1958; Pitt and Ugalde 1984; Coutinho and Corrêa 1999). A high concentration of $CaCl_2$ and the non-optimal growth temperature could thus have triggered sporulation as a stress response independently from each other due to the fact that setting the necessary temperature to 45 °C alone does not prevent the fungus from sporulating if the $CaCl_2$ concentration is still too high. Nevertheless, regarding the established protocols, sporulation is always observed when switching from batch to continuous cultivation, but a filamentous morphology could always be reached after about 12 to 17 residence times (see Figures 5.4-5.12, Pages 69 ff.).

In contrast to the wildtype strain, which did not need externally supplied $CaCl_2$ (natural $CaCl_2$ release from glassware during cultivation was sufficient), all strains derived from the marker free Δ*ku80* strain required an addition of 40 mg/L $CaCl_2$ .This phenomenon is so far not known and could be further investigated via a comparative genome analysis of the Δ*ku80* and wildtype strain. One possible explanation could be that FAA plating during the strain development caused a mutation due to the mutagenic nature of FAA (Van Leeuwe et al. 2019). This could be tested via repeating FAA plating of the Δ*ku80*, *amdS*$^+$ strain with following analysis of the growth behavior of several clones regarding $CaCl_2$.

6.3 Transcriptomic analysis of different *T. thermophilus* strains, cultivated using different conditions

6.3.1 Hydrolase and transcription factor expression in the reference strain

A broad spectrum of carbon sources can be utilized by *T. thermophilus*. The results of the plate growth assay (see Figure 5.2, Page 66) visualize this ability. Here the reference strain was able to grow on every tested carbon source. The best growth could be observed using glucose, cellobiose, cellulose, xylan, and mannose as carbon source. Regarding the results of the bioreactor cultivations with the reference strain MJK20.3 (see Figure 5.4, Page 69) the ability of the fungus to grow on cellulose was confirmed. Here a decrease in dry biomass (plant and fungal biomass) as well as an increased base addition and offgas values (O_2 consumed and CO_2 produced; after an initial spiking related drop of both values) after spiking cellulose is observable. The increase in protein secretion (up to 3-fold, see Figure 5.15, Page 80) and an altered secretion pattern as seen in the SDS PAGEs (see Figure 5.16, Page 81) after spiking cellulose indicates the production of enzymes required for the degradation of cellulose. The GO term enrichment analysis (see Tables 5.4 and 5.5, Page 89) of the transcriptome also illustrates that lots of genes related to cellulose utilization like cellulose binding genes and genes with cellulase activity as well as other CAZY genes like xylanase, xylosidase, and pectate lyase genes as well as general secretion processes are upregulated after the cellulose spike. The PCA plot (see Figure 5.17, Page 86) shows a massively altered expression of genes after spiking cellulose. This also fits to the results of the SDS PAGEs.

Only a few predicted proteases genes are differentially expressed at all points in time after the cellulose spike (see Figure 5.23, Page 95). The genes MYCTH_2295287, MYCTH_2303011, and MYCTH_2312662 are downregulated, while MYCTH_2308812 and MYCTH_2294116 are upregulated.

When looking for genes encoding for predicted proteases with a signal peptide (see Figure 5.24, Page 96) and the highest expression levels in steady state as well as after spiking cellulose MYCTH_2308737, MYCTH_2304704, and MYCTH_2297779 were found. In the publication of Li et al. 2020b, the effects of the deletion of the protease genes MYCTH_2303011, MYCTH_2308737, MYCTH_2295424, MYCTH_111035, and MYCTH_2308168 on protein secretion have been investigated in *T. thermophilus*. Deletion of the single genes as well as multiple deletions enhanced cellulase expression drastically. Among the highest expressed predicted protease genes that are differentially expressed at all points in time after the cellulose spike in the reference strain only two of those can be found: MYCTH_2303011 and MYCTH_2308737. For MYCTH_2295424 high but only slightly increased expression levels at 1 h to 4 h (counts before spike: ~950; counts after spike: ~700-2100) can be observed after the cellulose spike. The genes encoding MYCTH_111035 (counts before spike: ~30; counts after spike: ~30-200) and MYCTH_2308168 (counts before spike: ~200; counts after spike: ~110-180) were all showing almost no expression in steady state as well as after spiking cellulose. Table 6.1 lists the highest expressed predicted protease genes and their expression values during steady state and after the cellulose spike. The table includes both, protease genes that have already been studied and the ones worth studying. A deletion of them might enhance yield and purity of secreted proteins.

Table 6.1: Predicted proteases, recommended for deletion to possibly enhance yield and purity of secreted proteins in *T. thermophilus*. Shown are the respective genes encoding for the predicted proteases with (+SP), and without (-SP) a signal peptide as well as the normalized raw count values in steady state (SS), 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after spiking cellulose (t4). Protease genes where the effect of deletion on protein expression in *T. thermophilus* was already examined are written in bold letters.

gene		mean normalized raw counts	
-SP	+SP	**SS**	**t1, t2, t3, t4**
MYCTH_2312662		3797	683, 636, 1490, 1609
MYCTH_2308812		243	1073, 1954, 847, 857
	MYCTH_2295287	2108	1830, 1533, 1589, 1361
	MYCTH_2294116	125	963, 828, 377, 336
	MYCTH_2304704	4282	7931, 6049, 3838, 3256
	MYCTH_2297779	6493	9994, 15625, 8718, 6756
	MYCTH_2303011	5596	1288, 1977, 1516, 1357
	MYCTH_2308737	9578	15052, 11294, 9111, 7143

The expression of predicted CAZY genes (see Figures 5.29-5.36, Pages 103 ff.) in response to cellulose in *T. thermophilus* is very strong. One quarter of all predicted CAZY genes are differentially expressed at all points in time after the spike. Here especially cellulase, hemicellulase, pectinase, and esterase genes are strongly upregulated. The highest expression can be observed 1 h to 2 h after the cellulose spike, especially for genes encoding predicted cellulases, including LPMOs. The beneficial effects of LPMOs on enzyme cocktails used in lignocellulolytic biorefineries has already been demonstrated and might increase the profitability of those biorefineries via greatly enhancing the efficiency and thereby reducing the costs of enzyme cocktails (Chylenski et al. 2019). Due to the fact, that cellulose degradation of *T. thermophilus* is very effective and its expression of predicted LPMO genes under inducing conditions very high, the application of LPMOs from *T. thermophilus* in enzyme cocktails should be investigated in future experiments. Besides those observations, the upregulated and highly expressed LPMO *Mt*LPMO9A (MYCTH_85556) is the only LPMO known to be able to cleave xylan in addition to

cellulose (Frommhagen et al. 2015). This LPMO and possibly further LPMOs could contribute to the very effective use of xylan as a carbon source, as observed in the plate growth assay. A strong upregulation for predicted cellulase genes after addition of cellulose was expected but surprisingly, predicted hemicellulase genes, especially xylanase/xylosidase, mannanase/mannosidase, pectin lyase, and acetyl esterase genes were among the highly expressed genes as well. Synergistic effects of hemicellulases and pectinases when applied together with cellulases on cellulose degradation have already been demonstrated (Zhang et al. 2019). An explanation of the coupled expression of cellulases with hemicellulases and pectinases could have evolutionary origin, since besides cellulose, pectin, and hemicellulose are available in nature. Furthermore, hemicellulase and cellulase expression can be coupled due to shared regulators. The most prominent of those regulators is Xyr1/XlnR/Xlr1 (Hasper et al. 2000; Stricker et al. 2006; Li et al. 2015). A directly coupled expression of pectinases and cellulases is so far unknown. A further explanation could be the actual need of those enzymes for cellulose degradation, although this have not been reported yet. This could easily be investigated via deletion of the respective genes and following cultivation on cellulose. A few further predicted CAZY genes have been found to be highly expressed in response to cellulose in this study. These were not described as cellulase genes or genes encoding for lignocellulolytic enzymes in previous publications, like Berka et al. 2011 and Karnaouri et al. 2014a. Those CAZY genes (see Figure 5.36, Page 110) include the predicted cellobiose dehydrogenase genes MYCTH_111388 (counts before spike: ~63; counts after spike: ~30000-112000) and MYCTH_81925 (counts before spike: ~70; counts after spike: ~21000-73000) as well as the predicted CBM1 genes MYCTH_112399 (counts before spike: ~42; counts after spike: ~2500-10000) and MYCTH_33936 (counts before spike: ~3; counts after spike: ~1700-7800). According to the CAZY database the GH families those proteins belong to have cellulose as their most common substrate, which fits to the observed results (CAZY database). With the knowledge gained from the analysis of the expression of predicted CAZY genes on cellulose, the strongest expressed genes of each CAZY class can be identified (see chapter 5.3.1.3). With this information, single highly expressed CAZYs could be further examined for applications in cellulosic enzyme cocktails. Furthermore, the deletion of the respective genes could reduce the secretory burden and therefore enhance secretion of the protein of interest. The latter is much less effective compared to the deletion of regulators that control the expression of many genes. That is why the differential expression of transcription factors and known regulators of plant biomass degradation was investigated as well.

Generally, only a low number of predicted transcription factor genes (see Figures 5.40-5.43, Pages 117 ff.) were differentially expressed at all points in time after the spike (~10 %). Here many more genes were upregulated than downregulated (~2:1) and belong to the fungal Zn(2)-Cys(6) binuclear cluster domain, zinc finger C2H2 type and fungal specific transcription factor domain class. Generally, the strongest upregulation can be observed 1 h after the spike and correlates with the strongest upregulation of the predicted CAZY genes 1 h to 2 h after the spike. The upregulated genes with the strongest expression and without an ortholog of known regulators of plant biomass degradation are summarized in Table 6.2. According to their differential expression trend those genes are potential targets for deletion (upregulated genes) or overexpression (downregulated genes) to enhance yield and purity of proteins of interest similar to, e.g. Haefner et al. 2017b and Haefner et al. 2017a. Alternatively, this knowledge

can also be used to enhance expression of lignocellulolytic enzymes, e.g. for applications in lignocellulolytic biorefineries.

Table 6.2: Predicted transcription factor genes, recommended for deletion (DE) or overexpression (OE) to possibly enhance yield and purity of secreted proteins while using *T. thermophilus* as fungal production host. Shown are the descriptions of the best protein blast hits (lowest E-values), performed using the NCBI database as well as the normalized raw count values in steady state (SS), 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after spiking cellulose (t4).

			mean normalized raw counts	
DE	OE	description best blast hits	**SS**	**t1, t2, t3, t4**
MYCTH_2121737		sterol regulatory/Ara1	876	2341, 3991, 1799, 2230
MYCTH_2310586		hypothetical	926	1639, 2379, 2074, 1651
MYCTH_2295908		Pro1	684	1016, 4020, 3694, 1992
MYCTH_63778		Sp3	372	2423, 5156, 4245, 3751
	MYCTH_2305551	pyrimidine pathway regulatory 1/Ada6/citrinin CtnR	7831	1540, 2459, 2131, 1298
	MYCTH_2300024	ArcA like	5611	3880, 4221, 3442, 2592
	MYCTH_2300537	Acu15	3591	2426, 2403, 1974, 2579
	MYCTH_115802	SteA/Fst12/Ste12	8857	4962, 6028, 6595, 6401

For none of those genes an involvement in plant biomass degradation is described. In the following, the functions of the putative orthologs of the above-mentioned predicted transcription factors are shortly discussed. This will not be done for orthologs, whose description can be clearly assigned to a function as it is the case for, e.g. the description “pyrimidine pathway regulatory 1”. The general involvement of a sterol regulatory pathway protein in cellulase expression is already known. In *T. reesei* and *N. crassa* a deletion of *sah2* resulted in cellulase hyper-production phenotypes (Reilly et al. 2015). For the gene MYCTH_2121737 (sterol regulatory/ Ara1) the opposite trend is expected due to its upregulation after the cellulose spike. The transcription factor Sp3 is a member of the specificity protein/Krüppel-like factor (Sp/KLF) family as well as a ubiquitous and bifunctional transcription factor. It is mainly investigated in mammalian cells where it is involved in a variety of cellular processes with no clear assignment (Chen et al. 2020a). Pro1 is known to be involved in fungal development, e.g. sexual development and symbiotic or pathogenic host interaction (Steffens et al. 2016). The regulator Ada6 is involved in conidiation, sexual development and oxidative stress in, e.g. *N. crassa* (Sun et al. 2019) and very similar to the regulators SteA, Ste12, and Fst12 which are also known to be involved in fungal development, especially sexual development (Borneman et al. 2001; Lee et al. 2012). The biosynthesis pathway of the secondary metabolite citrinin is regulated via CtnR in, e.g. *A. nidulans* (Chiang et al. 2009). Involvement in the catabolic pathway of arginine is known for the gene *arcA* (Empel et al. 2001). The regulator Acu15 is involved in acetate utilization (Bibbins et al. 2002). As shown above, many genes encoding for orthologs of regulators that are involved in fungal (sexual) development seem to be involved in the regulation of cellulose degradation, which is known in literature. As an example, the light dependent regulator VeA/Vel1, which is known to be involved in fungal sexual development, is essential for the expression of (hemi)-cellulase encoding genes in *T. reesei* (Kim et al. 2002; Karimi-Aghcheh et al. 2014). Furthermore, regulators of secondary metabolism are also known to be involved in plant biomass degradation. For instance, Xpp1, a regulator responsible for a switch between primary and secondary metabolism, does repress xylanase-encoding genes in *T. reesei* (Derntl et al. 2015; Derntl et al. 2017).

To clarify the role of known regulators of plant biomass degradation during the cellulolytic response in *T. thermophilus*, expression of the *T. thermophilus* orthologs of those regulators was investigated as well (see Figure 5.44, Page 120) and will be discussed in the following. Among the upregulated genes, the strongest and most constant upregulation can be observed for the genes encoding for the orthologs of the regulators GaaR/Pdr2, GaaX, and HacA/1 as well as the already in *T. thermophilus* investigated regulators Clr2, Cre1, and Xyr1. An important regulator of cellulase expression in most filamentous fungi is ClrB/2 (Benocci et al. 2017). This seems also to be the case in *T. thermophilus* due to the extremely high expression levels of *clr2* (counts before spike: ~80; counts after spike: ~2750-6200) after the cellulose spike. The very important role of Clr2 for the expression of lignocellulolytic enzymes has been already shown in the patent of Haefner et al. 2017b, where the deletion of *clr2* massively lowered the amount of secreted enzymes in an industrial isolate of *T. thermophilus* and therefore enhanced the yield as well as purity of heterologously expressed proteins. The upregulation of *cre1* (counts before spike: ~2250; counts after spike: ~3100-5800) is explainable via the release of glucose during the cellulose degradation process. Glucose acts as an inducer of CreA/1, which then enables fine tuning of the cellulolytic response. Therefore, the role of Cre1 in *T. thermophilus* seems to be similar compared to other filamentous fungi (Orejas et al. 1999; Orejas et al. 2001; Tamayo et al. 2008; Sun and Glass 2011). Via deleting *cre1* this fine tuning of the cellulolytic response can be eliminated toward a higher expression of cellulases, which has been shown for *T. thermophilus* by Yang et al. 2015 and Liu et al. 2017. GaaR/Pdr2 (counts before spike: ~800; counts after spike: ~2400-6150) and GaaX (counts before spike: ~260; counts after spike: ~350-1100), both regulators specifically involved in pectin degradation are not known to be involved in cellulose degradation in other filamentous fungi (Alazi et al. 2016; Niu et al. 2017). The expression of these might be coupled to the observed upregulation of predicted pectin lyase genes. Whether the genes encoding for these regulators are necessary, influence cellulase expression, or are just evolutionary coupled to cellulase expression needs to be investigated via further deletion experiments. The extremely strong expression of the gene encoding for the ortholog of HacA/1 (counts before spike: ~3500; counts after spike: ~19800-38500) is explainable due to the fact that during the cellulolytic response the expression and secretion of enzymes (especially lignocellulolytic enzymes) is massively enhanced. This was already observed during protein analysis and predicted CAZY gene expression analysis. Given the high expression, the unfolded protein response is needed to ensure correct folding but also degradation of misfolded proteins (Mori et al. 1996; Kaufman 1999). The role of HacA/1 in balancing protein secretion during the lignocellulolytic response is described in Huberman et al. 2016 and seems to be important in *T. thermophilus* as well. The light dependent regulator VeA/Vel1 (counts before spike: ~830; counts after spike: ~920-1450) also seems to play a role in regulation of cellulase degradation in *T. thermophilus*. A similar function as in *T. reesei*, where Vel1 is considered as essential for the expression of (hemi)-cellulase encoding genes, is possible (Seiboth et al. 2012; Karimi-Aghcheh et al. 2014). XlnR/Xyr/GalR is known to regulate cellulase expression directly in *T. reesei*, *P. oxalicum*, and *A. niger* as well as indirectly via Vib1 in *N. crassa* (Hasper et al. 2000; Stricker et al. 2006; Sun et al. 2012; Li et al. 2015). In *T. thermophilus*, *xyr1* is presumably not involved in cellulose degradation since deletion of *xyr1* did not lead to growth defects on cellulose and expression of cellulolytic genes has not been significantly affected (Dos Santos Gomes et al. 2019). Nevertheless, a very high expression level of *xyr1* (counts before spike: ~500; counts after spike: ~4400-9950) could be

observed in *T. thermophilus* after spiking cellulose. This also fits to the observed strong expression of predicted hemicellulase and acetyl esterase genes, whose expression could be regulated by Xyr1. The regulation of especially predicted xylanase genes via Xyr1 in *T. thermophilus* has also been shown in the publication of Dos Santos Gomes et al. 2019. The observed coupled expression of predicted cellulase and hemicellulase genes could have evolutionary reasons, similar to the observed coupled expression of predicted pectinase genes.

The only relevant downregulated gene is *amyR*/*bglR* (counts before spike: ~5150; counts after spike: ~1450-2400). This fits to the observed repressing function of AmyR/BglR during cellulase expression in many other filamentous fungi mainly via glucose sensing for CCR (Nitta et al. 2012; Li et al. 2015; Benocci et al. 2017). In *T. thermophilus* this function for AmyR/BglR has been confirmed, in this case the deletion of *amyR/bglR* increased lignocellulolytic activities on cellulose (Xu et al. 2018).

Finally, genes encoding for orthologs of two regulators with high expression levels but no significant differential expression after the cellulose spike could be identified: *clr1* (counts before spike: ~2300; counts after spike: ~1900-2400) and *mcmA/1* (counts before spike: ~4650; counts after spike: ~4900-6050). ClrA/1 is besides ClrB/2 regarded as the main regulator of cellulase expression in many filamentous fungi (Benocci et al. 2017). A similar role of *clr1* compared to *clr2* for the expression of lignocellulolytic enzymes in *T. thermophilus* has been shown in the patent of Haefner et al. 2017a, where the deletion of *clr1* massively lowered the amount of secreted enzymes in an industrial isolate of *T. thermophilus* and therefore, enhanced the yield as well as purity of heterologously expressed proteins analogous to the deletion of *clr2* in the patent of Haefner et al. 2017b. Based on the observation in this thesis, *clr1* seems to be constitutively expressed. This suggests a regulatory function on scouting enzymes via Clr1 that are releasing inducers like, e.g. cellobiose, which activates the main lignocellulolytic response via Clr2. A more likely explanation (due to observations in other filamentous fungi) could be that Clr1 is constitutively expressed in an inactive form and only activated via inducers. This has been shown for *N. crassa* and *A. nidulans*, where Clr1/A has a role in cellulose sensing and is activated only in the presence of inducers (Coradetti et al. 2013). The fact that an activation of Clr1 is required was also shown via an overexpression experiment of *clr1* in *N. crassa*, where no expression of Clr1 target genes was observed (Coradetti et al. 2012; Craig et al. 2015). Similar functions could be possible for the ortholog of McmA/1 in *T. thermophilus* whose gene has an even stronger constitutive expression compared to *clr1* (counts before spike: ~4600; counts after spike: ~4900-6000). McmA/1 is known to positively control cellulase expression in *A. nidulans*, presumably via interaction with *clrB/2*, but has almost no effect on cellulase production in the fungus *T. cellulolyticus* (Yamakawa et al. 2013; Tani et al. 2014a; Fujii et al. 2015).

Generally, the strongest upregulation of the genes encoding for the discussed regulator orthologs can be observed 1 h after the spike, which correlates with the strongest upregulation of predicted CAZY genes 1 h to 2 h after the spike. Table 6.3 summarizes the genes encoding for orthologs of known regulators and already investigated regulators in *T. thermophilus* that are recommended for deletion or overexpression to either enhance or lower lignocellulolytic enzyme expression.

Table 6.3: Genes encoding for orthologs of known regulators and already investigated regulators of plant biomass degradation, recommended for deletion (DE) or overexpression (OE) to possibly enhance yield and purity of secreted proteins in *T. thermophilus*. Shown are the respective genes as well as the normalized raw count values in steady state (SS), 0.5 h (t1), 1 h (t2), 2 h (t3), and 4 h after spiking cellulose (t4). Regulators, where the effect of deletion on lignocellulolytic enzyme expression was already examined in *T. thermophilus* are written in bold letters.

			mean normalized raw counts	
DE	OE	regulator/ortholog	**SS**	**t1, t2, t3, t4**
MYCTH_38704		**Clr2**	80	2740, 6258, 3624, 3510
MYCTH_46266		GaaR/Pdr2	826	2506, 6143, 2391, 2582
MYCTH_46981		GaaX	261	835, 1095, 344, 375
MYCTH_2312657		VeA/Vel1	828	924, 1458, 1280, 1187
MYCTH_2310145		**Xyr1**	494	4394, 9959, 7135, 6289
MYCTH_2296663		**Clr1**	2270	2072, 2425, 1929, 2031
MYCTH_2132441		McmA/1	4639	4910, 6047, 5654, 4900
	MYCTH_2301920	**AmyR/BglR**	5142	1435, 2186, 2405, 2310
	MYCTH_2310085	**Cre1**	2245	3142, 4354, 5817, 3873

6.3.2 The effects of the deletion of cellulase regulator *clr1*

The effects of the deletion of *clr1* regarding growth on various carbon sources as well as the effects on the transcriptome while growing on glucose and cellulose in comparison to the reference strain MJK20.3 could be analyzed in this work. In the plate growth assay with 16 different carbon sources (see Figure 5.2, Page 66), growth defects were observed on cellulose and cellobiose. In addition, this growth defect on cellulose was confirmed in bioreactor cultivations. The strain showed no measurable physiological response to cellulose, offgas values (O_2 consumed, CO_2 produced) were decreasing, and base addition stopped completely (see Figure 5.10, Page 75). A GO term enrichment (see Tables 5.17 and 5.18, Pages 126 ff.) analysis of the transcriptome revealed upregulation of genes connected to nucleophagy, mitophagy, and autophagy upon switching the carbon sources to cellulose. At the same time genes that enable essential functions of the cell like transcription, translation, and respiration were downregulated. This clearly indicates the initiation of cell death (Galluzzi et al. 2018), which is likely caused by the inability of the *clr1* mutant to express the required enzymes needed for growth on cellulose. This is confirmed by the results of the Bradford assay (see Figure 5.15, Page 80) and SDS PAGEs (see Figure 5.16, Page 81), where a reduced protein content (~100 µg/mL) was observed. While protein secretion increases in the reference strain up to 3-fold after cellulose addition, secretion in the *clr1* mutant remains unchanged. The secretion pattern of the *clr1* mutant after spiking cellulose is weaker and completely different compared to the reference strain. An industrial isolate of *T. thermophilus* deleted in *clr1* gave a similar response to cellulose (Haefner et al. 2017a). In addition, deletion of *clr1* in *N. crassa* resulted in an inability to grow on cellulose (Coradetti et al. 2012). Growth and secretion of enzymes on glucose is unaffected by the deletion of *clr1*, which can be seen, e.g. in the plate growth assay, the physiology of the bioreactor cultivations, and the SDS PAGEs. The fact that the transcriptome during growth on glucose is unaffected by deletion of *clr1* is confirmed by the PCA plot (see Figure 5.17, Page 86). Here, a clustering between all steady state and exponential state samples is visible.

Regarding the expression of predicted protease genes (see Figures 5.49-5.54, Pages 133 ff.), many protease genes with very high expression values (higher compared to the reference strain) were observed in the *clr1* mutant, especially after spiking cellulose. This is presumably connected to the

starting cell death, where proteases are involved in the degradation of cellular components (Sukharev et al. 1997). Expression during steady state is similar to the reference strain, meaning that the same predicted protease genes shown in Table 6.1 could be deleted to possibly enhance yield and purity of secreted proteins while using this strain as a production host.

The expression of predicted CAZY genes in the *clr1* mutant (see Figures 5.58-5.70, Pages 146 ff.) is also completely different compared to the reference strain. Almost all predicted cellulase, hemicellulase, pectinase and esterase genes that are upregulated in the reference strain and therefore, presumably required for growth on cellulose, are showing no or only a very low expression in the *clr1* mutant. Especially genes encoding predicted CAZYs that have the highest expression values in the reference strain show no expression in the *clr1* mutant. The observations made regarding the expression of predicted CAZY genes explain the above-mentioned results of Haefner et al. 2017a.

The expression of genes encoding for orthologs of known regulators and already investigated regulators of plant biomass degradation upon deletion of *clr1* (see Figure 5.85, Page 175), was investigated in this work as well. The results indicate both a positive and negative regulatory role of Clr1. Among the genes that are indirectly or directly positively regulated by Clr1 and therefore, show a weaker expression compared to the reference strain are, e.g. the orthologs to the regulators GaaR/Pdr2, HacA/1, VeA/Vel1, and McmA/1 as well as the already investigated regulator genes *clr2*, *cre1*, and *xyr1*. Clr1 is so far only known to regulate *clr2* expression in *N. crassa* (Coradetti et al. 2012; Coradetti et al. 2013). The importance and involvement of *clr2* for cellulase expression in *T. thermophilus* is confirmed by the transcriptome data, where a high expression (normalized counts up to 6250) is observed in the reference strain and only a low (normalized counts up to 300) expression in the *clr1* mutant. A weaker expression of *cre1* (counts after spike: ~3100-5800 vs. 1450-2100) is observed and most likely due to the fact that cellulose cannot be degraded in the *clr1* mutant, subsequently no glucose is released, which itself acts as an inducer of CreA/1 (Orejas et al. 1999; Orejas et al. 2001; Tamayo et al. 2008; Sun and Glass 2011). GaaR/Pdr2 is a regulator specifically involved in pectin degradation and not known to be regulated via ClrA/1 in other filamentous fungi (Alazi et al. 2016; Niu et al. 2017). The gene encoding for the ortholog of GaaR/Pdr2 is lower expressed compared to the reference strain (counts after spike: ~2400-6150 vs. 1400-1800). This is also the case for predicted pectin lyase genes and might indicate an evolutionary dependent coupling between pectinase and cellulase gene expression. The weaker expression of the gene encoding for the ortholog of HacA/1 (counts after spike: ~19800-38500 vs. 2600-3100) is explainable due to the fact that, as already discussed, fewer proteins are secreted in the *clr1* mutant and therefore, the unfolded protein response is not needed to the same extent as in the reference strain. The regulation of *xlr1* expression via Clr1 as observed in the *clr1* mutant, is also known for *N. crassa*, whereas in *Aspergilli* regulation in the opposite manner can be observed (Craig et al. 2015; Raulo et al. 2016). For an industrial isolate of *T. thermophilus* the same response was noticed (Haefner et al. 2017a). The lower expression of *xyr1* (counts after spike: ~4400-9950 vs. 500-800) fits to the observed lower expression of predicted hemicellulase genes compared to the reference strain and the possibility of an evolutionary dependent coupling between hemicellulase and cellulase gene expression. Interestingly the gene encoding for the ortholog of VeA/Vel1 (counts after spike: ~920-1450 vs. 450-850) also seems to be regulated via Clr1, which has not been reported so far. Nevertheless, it is known that Vel1 is essential for the expression of (hemi)-cellulase encoding genes in *T. reesei* (Seiboth et al.

2012; Karimi-Aghcheh et al. 2014). Expression of *mcmA/1* (counts after spike: ~4900-6050 vs. 2150-2600) is not known to be regulated via ClrA/1. Due to the potential similar roles and functions of Clr1 and the ortholog of McmA/1 in *T. thermophilus* as discussed in the previous chapter, a transcriptional regulation of *mcmA/1* via Clr1 could be possible to enable fine tuning of cellulase expression.

Among the genes that are negatively regulated by Clr1 is *amyR*/*bglR*, which is expressed stronger (counts before spike:~8200; counts after spike: ~3500-5000) compared to the reference strain (counts before spike:~5150; counts after spike: ~1450-2400). This illustrates, that Clr1 might also play a role in coordinating and controlling CCR via direct regulation of *amyR/bglR* in *T. thermophilus,* which has not been observed in other filamentous fungi so far. The enhancement of cellulolytic activities on cellulose in a *T. thermophilus amyR/bglR* deletion strain and other filamentous fungi like *N. crassa* and *A. nidulans* was already observed (Benocci et al. 2017; Xu et al. 2018). Other genes encoding for orthologs of regulators like Sah2, Wc2/Blr2, and Stk12 are also stronger expressed in the *clr1* deletion strain. The role of Stk12 as a transcriptional brake could be similar compared to *N. crassa* where deletion of *stk12* resulted in a 7-fold higher cellulase production compared to the wildtype (Lin et al. 2019). In the *clr1* mutant the expression of the gene encoding the ortholog of Stk12 is massively enhanced (counts before spike:~2100; counts after spike: ~4500-7900) compared to the reference strain (counts before spike:~150; counts after spike: ~160-230), suggesting that Clr1 could also be involved in coordinating and controlling cellulase expression via regulation of the gene expression of the ortholog of Stk12. The same suggestions can be made for the gene encoding for the ortholog of Sah2 (counts after spike: ~600-900 vs. 1500-1700), where deletions in *T. reesei* and *N. crassa* resulted in cellulase hyper-production phenotypes (Reilly et al. 2015). The role of *wc2*/*blr2* (counts after spike: ~400-550 vs. 1100-2150), in plant biomass degradation has not been studied, yet.

Among the predicted transcription factor genes (see Figures 5.74-5.84, Pages 166 ff.) that could be involved in cellulase gene expression are MYCTH_2121737 (sterol regulatory Cys6/Ara1), MYCTH_63778 (Sp3) and MYCTH_57777 (regulatory protein Cys3), because they are upregulated in the reference strain but show no or a very low expression in the *clr1* mutant. The involvement of a sterol regulatory pathway protein in cellulase expression is already known for Sah2 (as previously described). Here MYCTH_2121737 and the gene encoding for the ortholog of Sah2 could play opposing roles in the regulation of cellulase expression. The known function of the transcription factor Sp3 has already been discussed in the previous chapter. The transcription factor Cys3 is known to regulate acquisition and utilization of sulfur in filamentous fungi as, e.g. in *N. crassa* (Paietta 2008). The importance and involvement of especially MYCTH_2121737 (sterol regulatory cys6/ara1) and MYCTH_63778 (Sp3) for cellulase expression in *T. thermophilus* gets stressed since in the reference strain these genes are highly expressed and upregulated (MYCTH_2121737: counts before spike:~880; counts after spike: ~1800-4000, MYCTH_63778: counts before spike:~370; counts after spike: ~2500-5100) and only lowly expressed without differential expression in the *clr1* mutant (MYCTH_2121737: counts before spike:~1000; counts after spike: ~750-1100, MYCTH_63778: counts before spike:~260; counts after spike: ~240-450).

6.3.3 The effects of the deletion of cellulase regulator *clr2*

In the plate growth assay with 16 different carbon sources (see Figure 5.2, Page 66), growth defects were only observed for cellulose and pectin for the *clr2* mutant. In contrast to the *clr1* mutant cellobiose could be utilized. This will be discussed later in this text. The growth defect on cellulose becomes clearer regarding the physiology of the *clr2* mutant during chemostat cultivation after spiking cellulose. The strain showed no physiological response to cellulose, offgas (O_2 consumed, CO_2 produced) values were decreasing, and base addition stopped completely similar to the *clr1* mutant (see Figure 5.9, Page 74). The results of the GO term analysis (see Tables 5.15 and 5.16, Pages 124 ff.) were very similar to those in the *clr1* mutant with the exception that in the *clr2* mutant, residual expression of some genes involved in cellulose catabolic processes and carbohydrate transport is visible. The same growth phenotype for a *clr2* deletion strain was described in *N. crassa* (Coradetti et al. 2012). The fact that the enzymes needed for growth on cellulose could not be secreted could be confirmed via Bradford assay (see Figure 5.15, Page 80) and SDS PAGEs (see Figure 5.16, Page 81). Here, the results are identical to the *clr1* mutant. A completely different and weaker expression pattern after deletion of *clr2* could also be observed in an industrial isolate of *T. thermophilus* during growth on cellulose (Haefner et al. 2017b). Growth and secretion of enzymes on glucose is unaffected by the deletion of *clr2*, which can be seen, e.g. in the plate growth assay, the physiology of the bioreactor cultivations, and the SDS PAGEs. The fact that the transcriptome during growth on glucose is unaffected by deletion of *clr2* is confirmed by the PCA plot (see Figure 5.17, Page 86). Here a clustering between all steady state and exponential state samples is visible.

Expression of predicted protease genes (see Figures 5.49-5.54, Pages 133 ff.) is very similar to the *clr1* mutant but with fewer predicted protease genes that are higher expressed compared to the reference strain. Expression during steady state is similar to the reference strain, meaning that the same predicted protease genes shown in Table 6.1 could be deleted to possibly enhance yield and purity of secreted proteins while using this strain as a production host.

The expression of predicted CAZY genes in the *clr2* mutant (see Figures 5.58-5.70, Pages 146 ff.) is also almost identical to the *clr1* mutant. Nevertheless, in contrast to the *clr1* mutant, moderate to high expression for some predicted ß-glucosidase (including the highest expressed predicted ß-glucosidase gene MYCTH_115968), hemicellulase, and pectin lyase genes can be observed in the *clr2* mutant. Therefore, the ability of the *clr2* mutant to grow on cellobiose is most likely due to residual ß-glucosidase activity (Karnaouri et al. 2014a; Lange 2017). A further explanation for this observation could be the expression of cellobiose transporters in the *clr2* mutant but not in the *clr1* mutant which has not been investigated within this work, but could be clarified by comparing the expression of predicted cellobiose transporter genes in both mutants or by deletion of these genes in the *clr2* mutant with a following cultivation on cellobiose. The growth defect on pectin does not have an obvious reason. In contrast to the *clr1* mutant, which is able to grow on pectin, a higher expression of some predicted pectinase genes was observed. It is possible that transporters required to take up pectin degradation products are not expressed in the *clr2* mutant. However, a transporter analysis is beyond the scope of this work. The general observations made regarding the expression of predicted CAZY genes explain the results of Haefner et al. 2017b, where the deletion of *clr2* in an industrial isolate of *T. thermophilus* drastically

lowered the amount of enzymes that are secreted along the protein of interest during recombinant polypeptide expression.

The expression of genes encoding for orthologs of known regulators and already investigated regulators of plant biomass degradation (see Figure 5.85, Page 175) is similar to the *clr1* mutant except for the genes encoding for the orthologs of VeA/Vel1 (counts after spike: ~450-850 (*clr1* mutant) vs. 950-1400 (*clr2* mutant)) and McmA/1 (counts after spike: ~2150-2600 (*clr1* mutant) vs. 5000-5500 (*clr2* mutant)). Additionally, *clr1* is slightly weaker expressed 0.5 h to 2 h after the cellulose spike (counts ~1900-2400 vs. 1150-1800) compared to the reference strain. A positive regulation of *clrA/1* expression via Clr2/B is not known so far. It seems that in *T. thermophilus* this is possible in the early response to cellulose, whereas at later points in time *clr1* gets expressed independently from *clr2*. Furthermore, regarding the residual expression of some predicted cellulase, hemicellulase, and pectinase genes and the ability to grow on cellobiose, Clr1 is likely responsible for a basic expression of lignocellulolytic enzymes that is not coupled to *clr2* expression. Expression of the genes encoding these enzymes is presumably directly regulated by Clr1. A direct regulation of lignocellulolytic enzyme gene expression via Clr1 was already observed in *N. crassa* (Coradetti et al. 2012; Coradetti et al. 2013). Nevertheless, the main and essential lignocellulolytic response (including regulators and CAZYs) seems to get triggered via Clr1 dependent expression of *clr2*. Orthologs of VeA/Vel1 and McmA/1 whose gene expressions seem to be regulated by Clr1 in *T. thermophilus* are apparently not regulated by Clr2, the same expression levels compared to the reference strain can be observed in the *clr2* mutant. This fits to the thesis that Clr1 and the ortholog of McmA/1 might have similar functions and are possibly regulating the gene expression of each other, as discussed in the previous chapter, to enable fine tuning and coordination of cellulase expression. The importance and involvement of all mentioned regulators for cellulase expression in *T. thermophilus*, whose expressions are affected by Clr1 or Clr2 get stressed, since in the reference strain these genes are highly expressed and only lowly expressed in either the *clr2* mutant or the *clr1* mutant or both.

The expression of predicted transcription factors genes (see Figures 5.74-5.84, Pages 166 ff.) is also very similar to the *clr1* mutant. The importance and involvement of the already mentioned genes MYCTH_2121737 (sterol regulatory cys6/ara1) and MYCTH_63778 (Sp3) for cellulase expression in *T. thermophilus* gets stressed, since in the reference strain these genes are highly expressed (see previous chapter) and only lowly expressed in the *clr2* mutant (MYCTH_2121737: counts before spike:~1050; counts after spike: ~500-1300/ MYCTH_63778: counts before spike:~450; counts after spike: ~700-2300) and the *clr1* mutant (see previous chapter). Therefore, expression of these transcription factors seems to be directly or indirectly regulated via *clr2* or *clr1*. Nevertheless, the observed Clr1 dependent regulation is most likely indirectly via Clr2.

6.3.4 The effects of the deletion of cellulase regulator *clr4*

Regarding the plate growth assay (see Figure 5.2, Page 66), no growth defect could be observed. Growth on all tested carbon sources was identical compared to the reference strain. During chemostat cultivation the *clr4* mutant behaved similar to the reference strain (see Figure 5.11, Page 76). This was confirmed by the GO term enrichment analysis, where the results also showed no differences (see Tables 5.19 and 5.20, Page 128). The protein concentration determined by the Bradford assay (see

Figure 5.15, Page 80) showed notably lower (~2 fold) concentration when comparing the data 4 h after cellulose addition between the *clr4* mutant and the reference strain. Similar trends were observed in Liu et al. 2019a, where the deletion of *clr4* in *T. thermophilus* lowered the total protein expression and the band pattern of the SDS PAGE after 4 days of growth on cellulose, in comparison to the wildtype strain. In contrast to these results, the results of the SDS PAGEs (see Figure 5.16, Page 81) in steady state and 4 h after the spike were identical between the reference strain and the *clr4* mutant. The similarity of transcriptome of the *clr4* mutant and the reference strain during growth on glucose and after spiking cellulose is clarified by the PCA plot results (see Figure 5.17, Page 86), where a clustering is visible.

The expression of predicted protease genes in the *clr4* mutant (see Figures 5.49-5.54, Pages 133 ff.) is very similar compared to the reference strain, meaning the same conclusions as for the reference strain can be made.

The expression of predicted CAZY genes in the *clr4* mutant (see Figures 5.58-5.70, Pages 146 ff.) is also very similar compared to the reference strain with the exception that a high number of predicted cellulase, hemicellulase, pectinase, and esterase genes that are upregulated in the reference strain and are, therefore, presumably required for growth on cellulose, are showing a higher expression in the *clr4* mutant. Especially the genes that have the highest expression values in the reference strain are affected. These observations are also very contrary to those of Liu et al. 2019a. They conducted a transcriptomic analysis with samples gained from cultures 4 h after switching the carbon source from glucose to cellulose in a shaking flask experiment. A massively lower expression of especially the major cellulase and hemicellulase genes in the *clr4* deletion strain compared to the wildtype strain (total RPKM: ~1000 vs. ~16000) was observed. However, only the GH family numbers of those genes are supplied, which is not sufficient to compare their results to ones from this study. Four genes were exactly described in the manuscript: the predicted endoglucanase gene MYCTH_86753, the predicted cellobiohydrolase genes MYCTH_109566 and MYCTH_66729 as well as the predicted ß-glucosidase gene MYCTH_115968. Regarding those genes, MYCTH_86753 (counts after spike: ~13500-56500 vs. 27500-102500) and MYCTH_115968 (counts after spike: ~25700-54800 vs. 35150-90300) have much higher expression levels in the *clr4* mutant compared to the reference strain, MYCTH_109566 (counts after spike: ~71500-214000 vs. 72350-180700) is expressed lower in the *clr4* mutant and MYCTH_66729 (counts after spike: ~22000-96600 vs. 32900-99300) is slightly higher expressed compared to the reference strain. These observations are in strong contrast to the results of Liu et al. 2019a. According to Liu et al. 2019a LPMO genes are also much lower expressed in the *clr4* deletion strain. Again, no gene identifiers are supplied and therefore, no comparison can be made. There are several predicted CAZY genes (including LPMO genes) with lower expression levels in the *clr4* mutant compared to the reference strain, but the difference is not as strong as described by Liu et al. 2019a, where almost no expression in the *clr4* deletion strain could be observed. In summary, the results from Liu et al. 2019a regarding predicted CAZY gene expression in a *clr4* deletion strain could not be confirmed in this study using bioreactor cultivations. In contrary, for most of the especially strongest expressed genes, a much higher or at least similar expression compared to the reference strain could be observed.

Regarding the expression of genes encoding for orthologs of known regulators or already identified regulators of plant biomass degradation (see Figure 5.85, Page 175), the effects of the deletion of *clr4*

on the gene expression of those regulators was investigated in this work as well. A positive and negative regulatory role of Clr4 was observed. Generally, many of the genes that show the strongest upregulation in the reference strain do have stronger expression values (e.g. *clr2*: counts after spike: ~2800-6300 vs. ~7300-17300) in the *clr4* mutant, but the expression profile is very similar. This observation is also made for the expression of predicted CAZY genes. Among the genes that are indirectly or directly positively regulated by Clr4 and therefore show a weaker expression compared to the reference strain are *cre1* (counts after spike: ~3100-5800 vs. 2850-4600) and the ortholog of *mcmA/1* (counts after spike: ~4900-6000 vs. 3500-3600). In the *clr4* mutant a lower expression of *cre1* is coupled with a higher expression of the ortholog of *creB* (counts after spike: ~1000-1200 vs. 2700-3550). In literature CreB is described as a deubiquitinating enzyme that enables higher CreA activity via preventing its proteasomal degradation in *A. nidulans* (Lockington and Kelly 2002). The observed lower expression of *cre1* and higher expression of the ortholog of *creB* therefore does not react as described in literature. It might be that in *T. thermophilus* deletion of *clr4* influences *cre1* expression and activity negatively (instead of positively) via the ortholog of CreB, which lowers carbon catabolite repression and therefore, enhances expression of genes encoding for predicted lignocellulolytic enzymes as observed. Expression of the ortholog of *mcmA/1* seems to be regulated via Clr4 similar to Clr1. A complex interaction between those regulators to enable fine tuning of cellulase expression could be possible.

Among the genes that might be negatively regulated by Clr4 and therefore, show a much stronger expression in the *clr4* mutant compared to the reference strain, the most notable are the orthologs of *gaaR*/*pdr2*, *gaaX*, *hacA*/*1*, and *stk12* as well as the already investigated regulator genes *amyR*/*bglR*, *clr2*, and *xyr1*. Particularly, *amyR*/*bglR* (counts after spike: ~1450-2400 vs. 2800-4600) has higher expression levels in the *clr4* mutant compared to the reference strain. This indicates, that Clr4 might also play a role in coordinating and controlling CCR via direct regulation of *amyR*/*bglR* expression in *T. thermophilus* besides Clr1. Together with the orthologs of *mcmA*/*1* and *stk12*, *amyR*/*bglR* expression trend is similar between the *clr4* and *clr1* mutant and stresses the possible function of Clr4 and Clr1 in coordinating CCR and fine tuning of lignocellulolytic enzyme expression possibly via those regulators. Nevertheless, to clarify the exact role of the genes encoding for the orthologs of McmA/1 and Stk12 (positive or negative influence on cellulase expression) further deletion experiments are necessary. Especially the much higher expression of *clr2* which was proven to be one of the main regulators of cellulase expression in *T. thermophilus* and the higher expression of the gene encoding for the ortholog of HacA/1 (counts after spike: ~19800-38500 vs. 21400-46000) confirm the observed higher expression of the main predicted cellulase, hemicellulase, and pectinase genes in the *clr4* mutant. Higher gene expression did not result in higher protein secretion (see SDS PAGEs and especially the Bradford assay). The reason for this could be that an increased translation of CAZYs was leading to secretion and folding stress (therefore activating UPR via, e.g. HacA/1) that was itself leading to secretion problems although a higher gene expression was observed. It has already been shown that bottlenecks in the secretory pathway of *A. awamori* resulted in partial intracellular retention of heterologous proteins (Lombraña et al. 2004). Furthermore, it was observed that an increased expression of secretory pathway genes can slow down secretion in order to deal with the increased protein load in *A. nidulans* (Schalén et al. 2016). The application of those observations in *T. thermophilus* could be investigated via analyzing the intracellular protein concentration or the investigation of protein expression longer than the 4 h within

this thesis. The assumption that the gene expression of orthologs of GaaR/Pdr2 and GaaX as well as the regulator Xyr1 is coupled to lignocellulolytic enzyme expression in *T. thermophilus* is confirmed by the fact that besides a higher expression of these genes also a higher expression of predicted CAZY genes is observed. The observed higher expression of these regulators in the *clr4* mutant might derive from the higher expression of *clr2* in the *clr4* mutant. Contrary, in the publication of Liu et al. 2019a, it was observed that expression of *clr2* and *xyr1* are negatively influenced by the deletion of *clr4* assuming a positive regulatory function of Clr4 for cellulase expression in *T. thermophilus*.

The expression of predicted transcription factor genes (see Figures 5.74-5.84, Pages 166 ff.) is also, as seen for all described categories so far, very similar compared to the reference strain. Worth to mention at this point is that the observed expression trends for the upregulated genes in the reference strain fit to the thesis of Clr4 as a negative regulator of mainly *clr2* expression, since, e.g. the genes MYCTH_2121737 (sterol regulatory Cys6/Ara1) and MYCTH_2295908 (Pro1), which are upregulated in the reference strain (see previous chapters) but downregulated in the *clr2* mutant (see previous chapter) and therefore, presumably regulated via Clr2 are stronger expressed in the *clr4* mutant (MYCTH_2121737 counts after spike: 2600-6700/ MYCTH_2295908 counts after spike: 1500-4300).

In accordance with the already presented results within this thesis a negative regulatory function of Clr4 in *T. thermophilus* is postulated in contrast to Liu et al. 2019a: Clr4 could act as an inducer activated (due to the constitutive expression in the reference strain) repressor or as an activator of a so far unknown repressor for fine tuning of the lignocellulolytic response and especially Clr2 could be repressed via direct binding of Clr4. The ability of Clr4 to bind to the *clr2* promotor was already proven by Liu et al. 2019a.

In summary, the observed differences to the results of Liu et al. 2019a, could be either due to the different experimental set-up (shaking flasks) and media used or due to a so far undetected additional mutation that causes the observed phenotype. This could be investigated in future experiments by performing experiments analogous to Liu et al. 2019a or by performing a comparative genome analysis to identify additional mutations.

6.3.5 The effects of the deletion of cellulase regulator *vib1*

In *N. crassa* this transcription factor is known to be involved in CCR via repressing *creA/1*, *creB*, *creD*, and *bglR*/*col26* expression. Therefore, a deletion of *vib1* in *N. crassa* causes a significantly lower expression of lignocellulolytic enzymes (Hynes and Kelly 1977; Kelly and Hynes 1977; Lockington and Kelly 2002; Boase and Kelly 2004; Nitta et al. 2012; Xiong et al. 2014). The expression of *vib1* is also known to be regulated via Clr1/2 and Xlr1 in *N. crassa* (Craig et al. 2015). In the reference strain and the *clr1*, *clr2* and *clr4* mutants a very low constitutive expression of *vib1* is observed (reference strain: counts before spike: ~350, counts after spike: ~350-450), suggesting a minor role in the regulation of cellulase expression. Nevertheless, regarding the plate growth assay (see Figure 5.2, Page 66), a growth defect on every carbon source could be observed. This growth defect was unusually strong for cellulose, cellobiose, pectin, polygalacturonic acid, and galacturonic acid. For growth on starch and its degradation products as carbon sources only minor growth defects could be observed. In contrast to these observations, the *vib1* mutant was able to use cellulose as a carbon source during chemostat

cultivation (see Figure 5.12, Page 77). The physiological response was very similar to the reference strain. Yet, more base (~1/3 more) was consumed during the cultivation compared to all other strains. Further experiments like protein analysis and transcriptomic analysis need to be carried out in future and can, therefore, not be discussed within this thesis. Taken together, the results of the protein analysis and the transcriptomic analysis are very important to clarify the exact role of Vib1 in *T. thermophilus*, since the results of the plate growth assay suggest a similar essential function compared to *N. crassa* and the physiology of the bioreactor runs suggests no essential function of Vib1 for cellulase expression. Together with evaluating the results of the transcriptomic data regarding hydrolase and transcription factor gene expression, the so far unexplainable higher acidification in BS9.3 and therefore a possible role of Vib1 in pH regulation should be investigated as well. Besides a secretion of acids, an enhanced ammonium uptake could be a possible reason for the observations, which could be identified in further experiments via analysis of the supernatant.

6.3.6 The effects of the deletion of cellulase regulator *vib2*

The gene *vib2* as well as its orthologs have not been described in literature, yet. Regarding the plate growth assay (see Figure 5.2, Page 66), results were identical compared to the reference strain. During chemostat cultivation, the physiological response to cellulose was identical to the reference strain (see Figure 5.8, Page 73). The results of the GO term enrichment (see Tables 5.6 and 5.7, Page 90) and protein analysis (see Figures 5.15 and 5.16, Pages 80-81) were identical compared to the reference strain. The similarity of transcriptome of the *vib2* mutant and the reference strain during growth on glucose and after spiking cellulose is clarified by the PCA plot results (see Figure 5.17, Page 86), where a clustering is visible.

The expression of predicted protease genes in the *vib2* mutant (see Figures 5.23 and 5.24, Pages 95-96) is very similar compared to the reference strain, meaning the same conclusions as for the reference strain can be made.

The expression of predicted CAZY genes in the *vib2* mutant (see Figures 5.29-5.36, Pages 105 ff.) is also very similar compared to the reference strain with the exception that a higher expression of the especially strongest expressed predicted CAZY genes (cellulase, hemicellulase, pectinase, and esterase genes) can be observed 0.5 h to 2 h after the cellulose spike. Nevertheless, after 4 h expression is for most of those predicted CAZY genes higher in the reference strain.

The expression of genes encoding for orthologs of known regulators and already investigated regulators of plant biomass degradation in the *vib2* mutant (see Figure 5.44, Page 120) is also very similar to the reference strain with the same exceptions as observed for the expression of predicted CAZY genes. In the reference strain *vib2* seems to be only slightly upregulated in response to cellulose 1 h to 2 h after the spike (counts ~1100-1300). Furthermore, the deletion of *clr1*, *clr2* or *clr4* does not severely influence expression levels of *vib2*.

For the expression of predicted transcription factor genes (see Figures 5.40-5.43, Pages 117 ff.) the same observations as for the known regulators can be made.

Taken together, besides the observed stronger expression of especially the highest expressed genes in the *vib2* mutant 0.5 h to 2 h, but not 4 h after spiking cellulose, no other differences to the reference

strain could be observed. It is likely that Vib2 plays a role in the early response to cellulose (in this case 0.5 h to 2 h after the spike). By deletion of *vib2* a putative regulatory repressing element could have been deleted, which now enables a stronger earlier expression of the required genes. Due to the low expression values of *vib2* in the reference strain and an only weak upregulation in response to cellulose, Vib2 function could be inducer activated. Therefore, Vib2 might play a role in CCR. Similar observations, as described above, were already made for protein kinase A in *A. nidulans*. Here, deletion mutants growing on cellulose are expressing lignocellulolytic enzymes at higher levels earlier in response to cellulose and it was suggested that the cells are glucose-blind and therefore, allow the production of cellulases under conditions in which CCR is active in wild type cells (De Assis et al. 2015). Although the mechanism could be possibly similar, the function of Vib2 in *T. thermophilus* is presumably independent from protein kinase A, since for the respective *T. thermophilus* ortholog, no significant differential expression in comparison to the reference strain could be observed in the *vib2* mutant as well as no upregulation in response to cellulose in the reference strain (data not shown). The higher expression of predicted CAZY genes 0.5 h to 2 h after cellulose addition in the *vib2* mutant could not be confirmed via Bradford assay and SDS PAGE. Possible reasons were already described for the *clr4* mutant. Nevertheless, for a final prove of the possible involvement of Vib2 in the lignocellulolytic response an overexpression experiment is recommended. This could verify the putative repressing function of Vib2 in the early response to cellulose.

6.3.7 Hydrolase and transcription factor expression in the reference strain in response to xylose

In the following, observed transcriptomic as well as physiological differences between cultivation on glucose and xylose are discussed. The results of the plate growth assay (see Figure 5.2, Page 66) were able to show that *T. thermophilus* can utilize a broad spectrum of carbon sources. However, xylose is a less preferred carbon source compared to, e.g. glucose or cellobiose. These results fit with the data of the bioreactor cultivation (see Figure 5.5, Page 70). In the batch phase heavy sporulation was observed. A possible reason could be the need for adaptation. Spores that were used for inoculation were harvested from plates with glucose as a carbon source. After starting the chemostat program the morphology was similar to cultivations using glucose apart from a lower biomass value (compare Figure 5.4, Page 69 and Figure 5.5, Page 70). During the wash out, a phase to determine the maximum growth rate a surprising observation was made: the morphology changed to hyphae swellings and thickenings and fragmented mycelium (see Figure 5.5, Page 70). This phenotype looked very similar to the propagules of the industrial isolates of *T. thermophilus* (internal communication). In literature it is known that environmental stress or changing carbon sources are able to trigger morphological changes in filamentous fungi, especially plant pathogens, as an important mechanism to adapt to the new situation (Jia et al. 2009; Francisco et al. 2018). Note that this phenotype was not observable during the exponential phase, which is surprising because the conditions were very similar, except that the fungus was not yet adapted to xylose in exponential phase. Via analysis of the transcriptome data and the planning of experiments to reproduce this phenotype the exact genetic mechanisms/requirements could

be identified, which are so far unknown due to the fact that this phenotype was achieved via undirected UV mutagenesis (Visser et al. 2011).

Regarding the results of the Bradford assay (see Figure 5.15, Page 80), the same amount of proteins was secreted in steady state and exponential state compared to the cultivation on glucose, although the enzymes secreted were very different compared to the cultivation on glucose, which can be seen in the results of the SDS PAGEs (see Figure 5.16, Page 81). As seen in the PCA plot (see Figure 5.17, Page 86), an altered expression of genes in comparison to growth on glucose is observable. This fits to the results of the SDS PAGEs.

The GO term enrichment analysis (see Tables 5.27-5.32, Pages 179 ff.) reveals that in comparison to cultivation on glucose mainly genes that are presumably required for xylose metabolism like oxidoreductases (Laluce et al. 2012) and transporters as well as balancing of energy metabolism like gluconeogenesis (Hector et al. 2011) are upregulated in exponential state, whereas also xylan catabolic processes including xylanase activity are upregulated. Xylose presumably acts as an inducer for the expression of those enzymes (Prathumpai et al. 2004; Mach-Aigner et al. 2010). Among the downregulated genes in exponential state in comparison to glucose are also mainly carbohydrate related processes including presumably not required transporters and oxidoreductases. In steady state very similar observations could be made, with the exception that oxidoreductases were not among the up- or downregulated genes as well as genes related to gluconeogenesis. Furthermore, cellulose binding genes were enriched among the upregulated genes.

To further investigate the metabolism on xylose, the expression profile of different genes of the XR-XDH pathway that is commonly used in fungi was analyzed. In this pathway, xylose is first converted into xylitol via xylose reductase (XR). Afterwards, xylitol is metabolized into xylulose via xylose dehydrogenase (XDH). Finally, xylulose is converted into xylulose-5-phosphate via xylulokinase (XK). The resulting xylulose-5-phosphate enters the pentose phosphate pathway where it is directly converted to glycerinaldehyde-3-phosphate via transketolase or indirectly via transaldolase. Glycerinaldehyde-3-phosphate itself is converted to pyruvate, which can enter the citrate cycle to generate reducing equivalents for the respiratory chain (Laluce et al. 2012). The expression of the genes encoding for the respective orthologs in *T. thermophilus* during growth on xylose in comparison to growth on glucose is visualized in Table 6.4.

Table 6.4: Expression of genes encoding for the orthologs of the main genes that are required for the metabolization of xylose in filamentous fungi. Shown are the normalized raw count values in exponential state (ex1/2= 3/4 h after starting the wash out) and steady state (SS) during growth on xylose (xyl) and glucose (glc).

		mean normalized raw counts				
gene	ortholog	**SS glc**	**SS xyl**	**ex glc**	**ex1 xyl**	**ex2 xyl**
MYCTH_43671	xylose reductase	451	4612	573	17814	17580
MYCTH_2293953	xylitol dehydrogenase	190	1204	111	1367	1259
MYCTH_67060	xylulokinase	311	1674	312	2325	2305
MYCTH_2300643	transketolase	4185	14434	8189	32078	35767
MYCTH_2297820	transaldolase	5937	8377	9914	21923	22152

As seen in Table 6.4, genes encoding for orthologs that are required for xylose metabolism are indeed upregulated. The higher expression values during exponential growth are likely related to a higher

growth rate (0.1 1/h vs 0.366 1/h). This might explain why genes required for gluconeogenesis were only enriched during exponential growth. Here, a higher demand for NADPH, which is required by the xylose reductase to convert xylose to xylitol might be provided by gluconeogenesis. This requirement was already demonstrated and proven in *S. cerevisiae* strains that were engineered for xylose metabolism (Hector et al. 2011).

Regarding the expression of predicted protease genes (see Figures 5.89-5.93, Pages 185 ff.), only genes are notably stronger expressed compared to cultivation of glucose. Among them are the genes MYCTH_2312662 (steady state and exponential state), MYCTH_2303011 (steady state only), MYCTH_2299023 (exponential state only), and MYCTH_2112181 (exponential state only). Together with the already made recommendations for the deletion of predicted protease genes in *T. thermophilus* when grown on glucose, the deletion of the predicted proteases genes listed in Table 6.5 is recommended when xylose is used as a carbon source to possibly enhance yield and purity of secreted proteins. The expression of especially the highly expressed predicted protease genes is not very different compared to growth on glucose, therefore, almost the same recommendations for deletion can be made.

Table 6.5: Predicted proteases, recommended for deletion to possibly enhance yield and purity of secreted proteins in *T. thermophilus*. Shown are the respective genes encoding for the predicted proteases with (+SP) and without (-SP) a signal peptide as well as the normalized raw count values in steady state (SS) and exponential state (ex 1/2= 3/4 h after starting the wash out) while using glucose (glc) or xylose (xyl) as carbon source. Proteases where the effect of deletion on protein expression in *T. thermophilus* was already examined are written in bold letters.

Gene		mean normalized raw counts				
-SP	+SP	**SS glc**	**SS xyl**	**ex glc**	**ex1 xyl**	**ex2 xyl**
MYCTH_2299023		1031	1502	1242	6245	5216
MYCTH_2312662		3797	10746	2861	11574	11610
	MYCTH_2295287	2108	1832	1087	1674	2045
	MYCTH_2304704	4282	3612	3709	2091	2270
	MYCTH_2297779	6493	6923	4569	2521	2688
	MYCTH_2303011	5596	7974	66	710	652
	MYCTH_2308737	9578	4711	7928	2424	2759
	MYCTH_2112181	1299	1639	2267	4350	4745

Upregulated predicted CAZY genes in response to xylose in comparison to growth on glucose (see Figures 5.97-5.104, Pages 194 ff.) can be mainly found in the hemicellulase category with the predicted xylanase/xylosidase genes MYCTH_89603, MYCTH_116553, and MYCTH_80104, the predicted arabinase/arabinofuranosidase genes MYCTH_39555 and MYCTH_2301869, the predicted galactosidase gene MYCTH_40111, and the predicted acetyl esterase gene MYCTH_40885. The expression of all of the above-mentioned predicted hemicellulase genes is also upregulated in response to cellulose as seen in the cellulose spiking experiments, suggesting either a bifunctional role in cellulose/xylan degradation or the already discussed evolutionary coupling of hemicellulase and cellulase expression. In case the bifunctional role can be confirmed via further experiments, those enzymes are interesting candidates for enzyme cocktails used in lignocellulolytic biorefineries. The above made observations also fit to the expression of the predicted endoglucanase gene MYCTH_109444 and the LPMO gene MYCTH_85556. Both are highly expressed during growth on

xylose and cellulose. The LPMO MYCTH_85556 is the only LPMO that was so far described to be acting on xylan as well as cellulose. The observed upregulation of gene expression of this LPMO on xylose as well as cellulose fits to the observation by Frommhagen et al. 2015. Expression of the predicted endoglucanase gene MYCTH_109444 seems to be also triggered via xylose and cellulose, suggesting bifunctional xylanase/endoglucanase activities, which have already been observed in bacteria (Chang et al. 2011). All the above described enzymes have the potential to be applied in enzyme cocktails due to their potential activities on xylan and cellulose. The general potential of *T. thermophilus* to use cellulose and xylan as carbon source have in addition already been seen in the plate growth assay, where those were regarded as two of the best carbon sources for *T. thermophilus*. A few predicted CAZY genes have been found to be highly expressed in response to xylose only and were not described as hemicellulase genes or genes encoding for lignocellulolytic enzymes in previous publications, like Berka et al. 2011 and Karnaouri et al. 2014a. Those predicted CAZY gens include MYCTH_96141 and MYCTH_55568. MYCTH_96141 has here an extremely high expression value (counts in steady state: ~45000, counts in exponential state: ~10000). According to the CAZY database the GH families those proteins belong to have hemicellulose as their most common substrate, which fits to the observed results (CAZY database).

Among the upregulated predicted transcription factor genes that are no ortholog of known regulators of plant biomass degradation with the strongest differential expression in comparison to growth on glucose (see Figures 5.108-5.111, Pages 206 ff.) are the genes that are presented in the following Table 6.6.

Table 6.6: Strongest differentially expressed predicted transcription factor genes that are no ortholog of known regulators of plant biomass degradation during growth on xylose in comparison to growth on glucose. Shown are the descriptions of the best protein blast hits (lowest E-values), performed using the NCBI database as well as the normalized raw count values in exponential state (ex1/2= 3/4 h after starting the wash out) and steady state (SS) during growth on glucose (glc) or xylose (xyl). The color of the gene indicates the general expression trend (up- or downregulation) during growth on xylose compared to growth on glucose in steady state (color of MYCTH_) and exponential state (color of number); green= upregulated, red= downregulated, black= no differential expression.

		mean normalized raw counts				
gene	description best blast hits	**SS glc**	**SS xyl**	**ex glc**	**ex1 xyl**	**ex2 xyl**
MYCTH_2303596	sterol uptake control protein 2/ beauvericin cluster specific repressor	2157	2852	110	1397	1166
MYCTH_2310586	hypothetical	926	354	1385	130	125
MYCTH_2089183	Nit4/NirA	522	571	1592	317	302
MYCTH_73708	Moc3/control of sterol uptake	317	589	1116	301	227
MYCTH_2306627	transcriptional activator Dal81	83	844	83	198	179
MYCTH_2301577	hypothetical	997	148	20	24	23
MYCTH_92449	hypothetical	56	43	6293	4	6
MYCTH_2297012	all development altered	138	895	282	356	350
MYCTH_2304200	centromere binding protein	566	274	972	108	83
MYCTH_2311995	transcription initiation factor	175	601	233	877	994

None of these genes was observed to be extremely upregulated during growth on cellulose, therefore, up- or downregulation is presumably connected to the use of xylose as a carbon source only. Note that the general expression values are also very low for most of the genes here. Among those genes, no gene is known to be involved in lignocellulolytic enzyme expression or xylose metabolism. The genes that are up- or downregulated in steady state as well as in exponential state are the most promising

candidates for putative regulators of hemicellulase expression. The exact function of those regulators cannot be predicted via analyzing the transcriptomic data, since expression of hemicellulase genes could either be induced or repressed via xylose. The dual function of xylose as an inducer as well as a repressor of xylanase expression, e.g. is already described for *A. nidulans* and *T. reesei* (Prathumpai et al. 2004; Mach-Aigner et al. 2010). Therefore, to finally investigate the exact function of the predicted transcription factors regarding hemicellulase expression, further deletion or overexpression experiments with those genes must be carried out. In the following the functions of the putative orthologs of the above-mentioned predicted transcription factors are shortly discussed. This will not be done for orthologs, whose description can be clearly assigned to a function as it is the case for, e.g. the description “centromere binding protein”. The involvement of sterol regulatory pathway proteins in lignocellulolytic enzyme expression has already been discussed in a previous chapter (Reilly et al. 2015), although so far none of those protein were shown to be involved in hemicellulase expression. Nit4/NirA and Dal81 are both regulators known to be involved in nitrogen metabolism (Bricmont et al. 1991; Yuan et al. 1991). Genes involved in nitrogen metabolism are known to be involved in affecting the expression of (hemi)-cellulases in *A. nidulans* and *N. crassa* (Arst and Cove 1973; Fu and Marzluf 1990; Berger et al. 2008; Gonçalves et al. 2011; Amore et al. 2013). The genes *moc3* and the all development altered genes like, e.g. *ada6*, are known to be involved in sexual development and stress response (Goldar et al. 2005; Sun et al. 2019). As already observed during growth on cellulose, developmental regulatory genes could presumably be involved in the regulation of the lignocellulolytic response.

In the following only the expression of genes encoding for orthologs of known regulators and already investigated regulators of plant biomass degradation that are very important in hemicellulose degradation and are notably upregulated or downregulated in both steady state and exponential state (see Figure 5.112, Page 210) will be discussed in detail. Note that expression levels here are very low and differential expression is much weaker compared to the response to cellulose. Among the upregulated genes the orthologs of *creB*, *nmrA/1*, *stk12*, and *vib1* can be found. An upregulation of the ortholog of *creB* was already observed in combination with a weaker *cre1* expression during lignocellulolytic response in the *clr4* deletion strain. In contrast, *cre1* expression is only slightly upregulated in exponential state during growth on xylose in comparison to growth on glucose (counts: ~1350 vs. ~950). The reason for the upregulation of *cre1* in exponential state could be that as observed during the GO enrichment analysis, gluconeogenesis is used to gain NADPH. The produced glucose could be the reason for the higher *cre1* expression (Orejas et al. 1999; Orejas et al. 2001; Tamayo et al. 2008; Sun and Glass 2011). On the other hand CreA/1 is known to be involved in regulating gluconeogenesis via CCR in filamentous fungi (Beattie et al. 2017). The reasons for the enhanced expression of the ortholog of *creB* could be similar as already described for the *crl4* mutant. The observed higher expression of *amyR*/*bglR* during exponential state only (counts: ~2750 vs. ~800) in comparison to growth on glucose could have the same reasons as the higher expression of *cre1*, since also *amyR/bglR* expression is known to be influenced by glucose (Benocci et al. 2017). The higher expression of the ortholog of *nmrA/1* (counts in exponential state: ~1100 vs. ~500, counts in steady state: ~1050 vs. ~600) stresses the potential involvement of regulators of nitrogen metabolism in the regulation of lignocellulolytic enzyme expression (Macios et al. 2012). The ortholog of *stk12* is the gene

with the strongest upregulation in response to xylose in comparison to growth on glucose (counts in exponential state: ~1100 vs. ~100, counts in steady state: ~2450 vs. ~150). Besides upregulation on xylose, also a high expression in the *clr1* and *clr4* deletion strain were already shown, which stresses the already suggested importance of the gene encoding for the ortholog of Stk12. In which way this gene exactly influences (hemi)-cellulase expression is not clear according to the observed data and needs to be clarified via further experiments. The expression of *vib1* on xylose is also slightly elevated in comparison to glucose (counts in exponential state: ~450 vs. ~100, counts in steady state: ~700 vs. ~350), which underlines the putative importance in the regulation of plant biomass degradation. The expression of *xyr1* (counts in exponential state: ~500 vs. ~100, counts in steady state: ~700 vs. ~500) is only at exponential state on xylose significantly higher in comparison to growth on glucose but expression levels are very low. The observed higher expression of some predicted hemicellulase genes during growth on xylose is, therefore, probably not related to Xyr1. Nevertheless, Dos Santos Gomes et al. 2019 observed a requirement of *xyr1* for growth on xylose and xylan in *T. thermophilus*. Taken this observation into account, only minor amounts of Xyr1 seem to be required for growth on xylose. In which way *xyr1* is necessary for growth on xylose is not clear and should be investigated in future experiments. However, Xyr1 does not seem to be the only main regulator of hemicellulase expression in *T. thermophilus*.

Among the notably downregulated genes the orthologs of the regulators Hap2, Hap3, Hap5, McmA/1, and Mhr1 can be found. Growth on xylose or the expression of hemicellulases could be controlled via the heterochromatin status due to the observed downregulation of genes encoding for the orthologs of the Hap complex during growth on xylose. The enhancement of transcription of xylanases, e.g. has been demonstrated via mutating the CCAAT sequences, where the Hap complex binds (Zeilinger et al. 1998; Würleitner et al. 2003). The ortholog of *mcmA/1* is also slightly downregulated during growth on xylose (counts in exponential state: ~2900 vs. ~5600, counts in steady state: ~3400 vs. ~4600). Together with the already made observations in the *clr1* and *clr4* deletion strains, an inducer activated regulatory role of the gene encoding for the ortholog of McmA/1 exclusively during cellulase expression is possible. Here, xylose could also have a repressing function. Finally, expression of *mhr1*, whose gene product was characterized as a repressor of hemicellulase as well as cellulase activities in *T. thermophilus* is also downregulated (counts in exponential state: ~80 vs. ~450, counts in steady state: ~180 vs. ~300) during growth on xylose (Wang et al. 2018). Note that expression levels are here also very low. The observations made fit to the observations of Wang et al. 2018, suggesting a repressing role of Mhr1 during hemicellulase expression.

Taken together, the requirement of specific transcription factors during growth on xylose is not clear according to the transcriptomic data. Furthermore, no single transcription factor that is presumably involved in hemicellulase expression could be clearly determined. To investigate this, bioreactor cultivations similar to those in this work with a shift to xylan as a carbon source are recommended. In combination with the data of the xylose cultivation much more precise statements about possible regulation processes for hemicellulase expression as well as the connection of hemicellulase and cellulase expression can be made.

6.3.8 Response of the reference strain to a pH shift

The pH shift to 5 results in a lower growth rate (see Figure 5.13, Page 78) compared to growth at pH 6.7. This fits to the described pH optimum in literature. Here it is described that *T. thermophilus* is able to grow best at neutral or alkaline pH (Singh 2014). The optimum pH range of the industrial isolate C1 is observed to be between pH 5 and pH 8 (Vitikainen 2018). This means that the acidic pH of 5 marks the lowest optimum pH value. Besides the lower growth rate, no differences compared to growth at pH 6.7 can be observed (morphology, physiology). Therefore, the results of the protein and transcriptome analysis that are not available yet and thus cannot be discussed in this thesis are important to understand the underlying processes. Within this analysis it is expected that orthologs of genes that are known to be involved in the pH regulation in filamentous fungi are differentially expressed after shifting the pH to 5. These genes include *pacC*, *palA*, *palB*, *palC*, *palF*, *palH*, and *palI* which were first identified in *A. nidulans*. Here, *pacC* encodes a zinc finger transcription factor which is mainly responsible for the transcriptional activation (e.g. of alkaline phosphatase PalD) or repression (e.g. of acid phosphatase PacA) of genes required for ambient pH regulation and the *pal* genes encode for an ambient pH signal transduction pathway, which is required to transform PacA from an inactive to an activate form (Denison 2000; Peñalva and Arst 2002). The identification of genes that are involved in ambient pH regulation could be also beneficial for the identification of genes that are involved in lignocellulolytic enzyme expression, since it is known that for example, PacC regulates (hemi)-cellulase expression in *A. tubingensis*, *T. reesei*, and *A. nidulans* (De Graaff et al. 1994; He et al. 2014; Kunitake et al. 2016). For *T. thermophilus* an upregulation of the ortholog of *pacA* at the early points in time after the cellulose spike could be observed in all strains as well, indicating, that PacC could possibly play a role in (hemi)-cellulase expression in *T. thermophilus*. Further advantages of additional transcriptomic data sets, like the dataset obtained from the pH shift experiment, is the resulting expansion of the transcriptomic network of *T. thermophilus*, which allows the prediction of processes in which so far uncharacterized genes are putatively involved via co-expression analysis as described in Schäpe et al. 2019.

7. Conclusions and Outlook

In the presented work, the successful establishment of experimental methods for working with *T. thermophilus* like chemostat bioreactor cultivation or general molecular biological methods was described and the respective protocols were provided. With these protocols, studies on *T. thermophilus* will be as easy as in already established organisms like, e.g. *A. niger* or *N. crassa*. Furthermore, a deep insight into the regulation and expression of predicted hydrolase genes, especially in response to cellulose, could be gained. As a result, so far uncharacterized highly expressed predicted protease genes (e.g. MYCTH_2304704 and MYCTH_2297779) were identified. Their deletion will presumably be a great benefit to reduce proteolytic degradation and therefore, increase the yield of secreted proteins. Using the CRISPR/Cas12a system, which was successfully used within this work, will allow multiple deletions simultaneously and presumably enhance the effects of protease deletion. The lignocellulolytic network analysis revealed a tightly coupled expression of predicted cellulase, hemicellulase, pectinase, and esterase genes in response to cellulose and therefore the enormous potential of *T. thermophilus* or its enzymes for applications in lignocellulolytic biorefineries, which should also be investigated in future experiments. Here, especially heterologous expression of CAZYs like the LPMO MYCTH_85556 and the predicted endoglucanase MYCTH_109444, which are presumably acting on both hemicellulose and cellulose, is of interest. An analysis of the degradation potential of those enzymes on hemicellulose and cellulose can clarify the potential of those enzymes for applications in biorefineries. This could be achieved via either expression of the enzymes in an already optimized expression host that does not secrete many side products (e.g. bacterium or yeast), with following isolation, purification, and subsequent analysis or via overexpression of the respective enzymes in *T. thermophilus* with following growth assays on lignocellulose. Figure 7.1 shows a summary of the regulation dependencies between the highest expressed and therefore presumably most important orthologs to known or so far uncharacterized putative regulators of plant biomass degradation in *T. thermophilus* based on the transcriptomic data obtained in this study together with a comparison of the function of those regulators in *T. reesei* and *N. crassa.* Most of the similarities can be found between *T. thermophilus* and *N. crassa*, e.g. the essential function of Clr1 and Clr2. But also, similarities to *T. reesei*, e.g. the putative function of the ortholog of VeA can be found. Additionally, regulation mechanisms that are independent from the so far described ones were identified. An example is the possible function of the orthologs of Stk12 and McmA/1 as well as the possible involvement of the orthologs of pectinase regulators GaaR and GaaX in cellulase expression. Surprisingly, Clr4 was identified as a repressor in contrast to the descriptions in literature. In addition, a potential new regulator (Vib2) could be discovered whose deletion enabled an earlier overall lignocellulolytic response. Since the transcriptomic response for the *vib1* deletion strain and the pH experiment was not analyzed within this study, the same procedure as presented in this work is recommended. Furthermore, putative transcription factors that have so far not been described as responsible for regulating lignocellulolytic enzyme expression were identified. Examples are MYCTH_57777, MYCTH_2121737, and MYCTH_63778. The deletion or overexpression of all above-mentioned transcription factors is highly recommended to characterize or confirm the exact function of those regulators, especially in strains used by industry. This can be supported by using the CRISPR/Cas12a system, especially for the simultaneous modification of multiple genes to investigate

combinatorial effects. For a controlled overexpression further work on the Tet-On/Off system is required. A stable integration of the Tet-On/Off system without the possibility of homologous recombination after integration is recommended. To investigate the effects of deletion or overexpression of putative regulators on hydrolase expression and secretion, an experimental set up analogously to the one presented in this work is recommended. With the availability of more transcriptomic data (further strains and conditions), the transcriptional network of *T. thermophilus* can be expanded, allowing to predict the processes in which so far uncharacterized genes are putatively involved via co-expression analysis. The prediction of those processes is with regards to the percentage of hypothetical genes in *T. thermophilus* (~8400 of 9100 genes) extremely useful to uncover possible gene functions. In addition, the transcriptomic data offer further possibilities. Here, for example, inducible or non-inducible promotors (according to the cultivation conditions) for heterologous protein expression in industrial expression hosts can be identified as well as strongly expressed loci for the integration of the respective expression cassettes. Finally, all discoveries made within this work contribute to the characterization and optimization of *T. thermophilus* as a fungal production host.

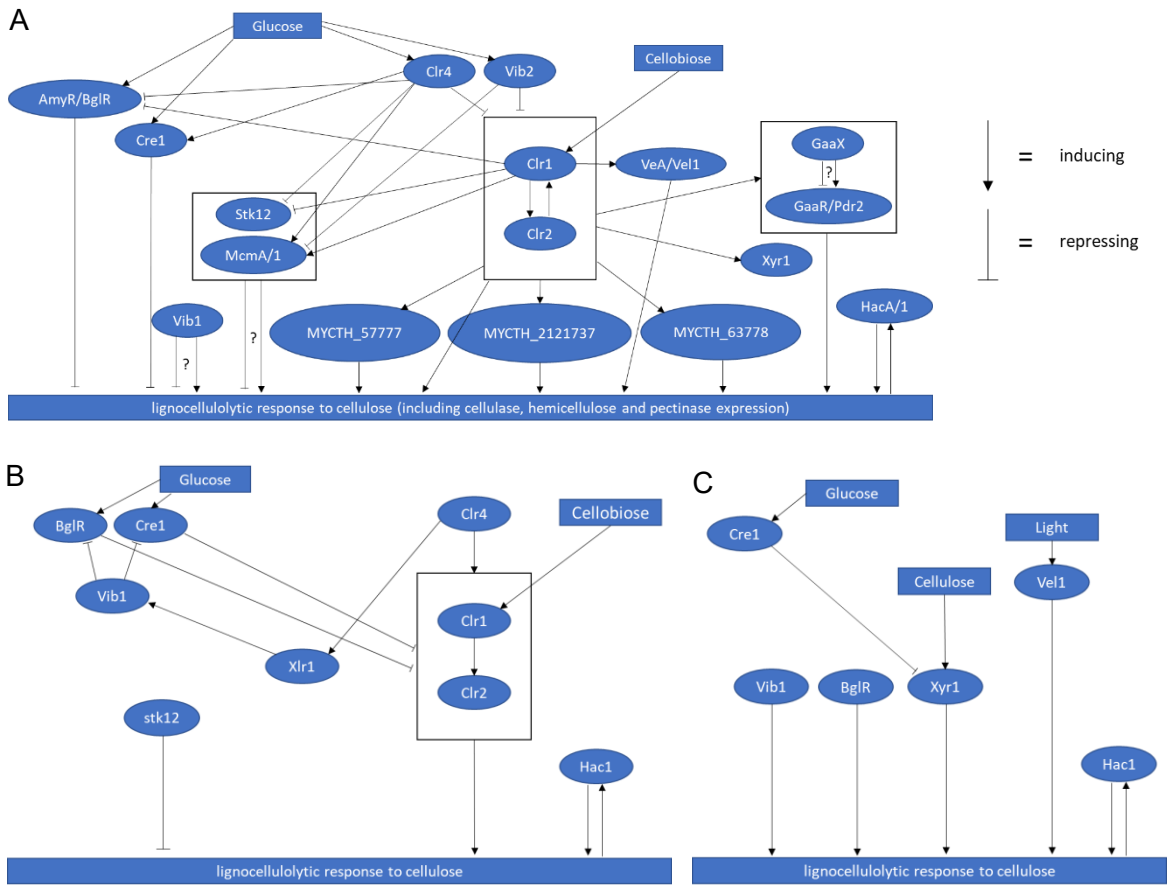

Figure 7.1: Predicted regulatory network of lignocellulolytic enzyme expression in *T. thermophilus* based on the transcriptomic fingerprint obtained in this study and the known function of the respective regulators in *T. reesei* and *N. crassa*. Shown are the respective networks for (A) *T. thermophilus*, (B) *N. crassa*, and (C) *T. reesei*. The regulators mentioned in (A) and not mentioned in (B) and (C) do not have a role in regulating the lignocellulolytic response to cellulose in the respective fungi or have not been studied, yet. Question marks indicate regulation processes for which contrary observations and therefore no clear prediction could be made within this work.

8. Appendix

8.1 Diagnostic PCR scheme

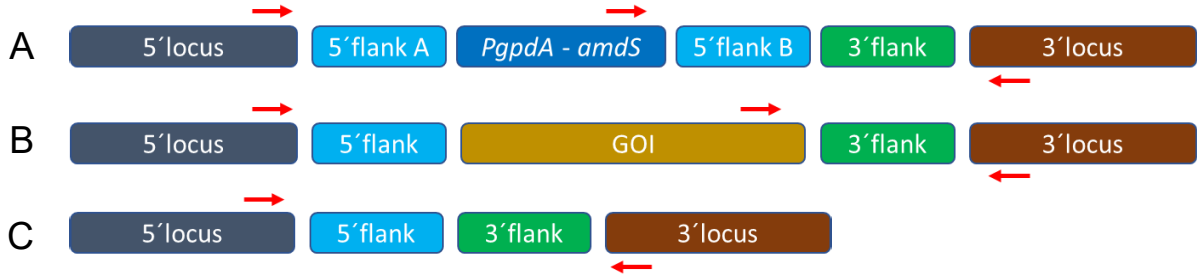

Figure 8.1: Schematic depiction of all possible genotypes (except multiple integration) after transformation. (A) Genotype after integration of the deletion cassette. (B) Genotype of the wildtype strain, whereas GOI can be *ku80*, *vib1*, *vib2*, *clr1*, *clr2*, or *clr4* (C) Genotype after successful marker loop out in (A). Locations of binding sites of the used primers listed in Table 4.5 are marked with a red arrow.

8.2 Cloning experiments

Table 8.1: Information about the templates and primers used to amplify the fragments and fused fragments for cloning of the *clr4* deletion plasmid via CPEC.

fragment	template used to amplify fragment	primer used to amplify fragment	fused fragments	primer used to fuse fragments
BB	MT134	2132+2089	BB+5´flank A	2132+2136
5´flank A	gDNA	2128+2081		
P An *gpdA*–An *amdS*-T An *amdS*	pMJK19.7	2129+2083		
5´flank B	gDNA	2130+2085	5´flank B+3´flank	2130+2139
3´flank	gDNA	2131+2087		

Table 8.2: Information about the templates and primers used to amplify the fragments for cloning of the *vib1* deletion plasmid via Gibson cloning.

fragment	template used to amplify fragment	primer used to amplify fragment
BB	pMJK18.1	2297+2298
5´flank A	MT1402	2291+2292
P An *gpdA*–An *amdS*-T An *amdS*	pMJK18.1	2293+2294
5´flank B, 3´flank	MT1402	2295+2296

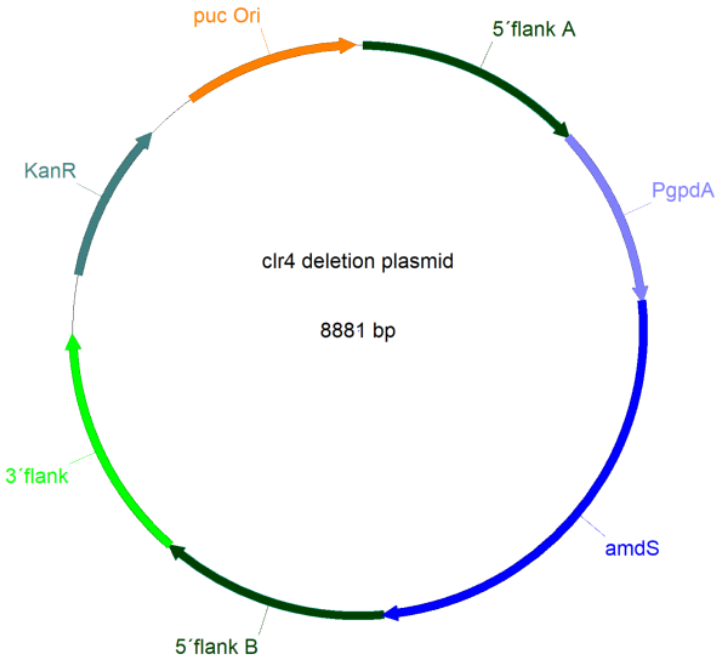

Figure 8.2: Plasmid map of the *clr4* deletion plasmid. The plasmid is carrying a 5´ flank (A) and a 3´flank, surrounding the *clr4* ORF for homologous integration as well as via *PgpdA* constitutively expressed *amdS* (including terminator) as selection marker and a second 5´flank (B) for marker removal. KanR and puc Ori are required for selection and amplification in *E. coli*.

8.3 Southern data

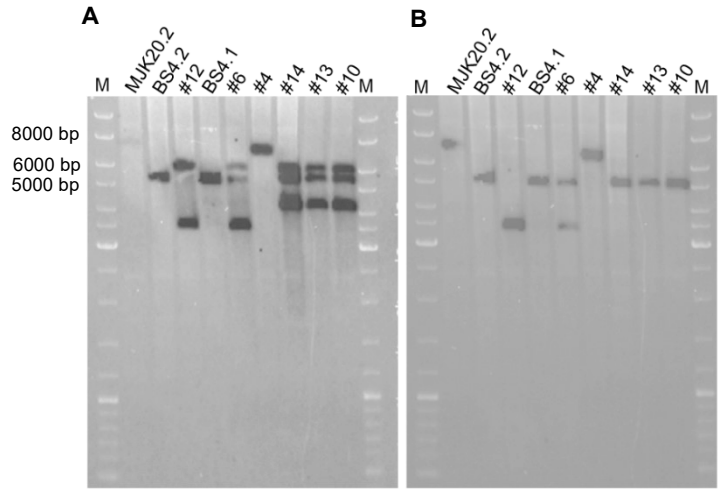

Figure 8.3: Results of the Southern analysis after deletion of *vib2* in MJK20.2. Shown are the results of the Southern analysis after marker recycling using the 5´probe (A) and 3´probe (B). Enzymatic digestion was performed with XmnI. The marker (M) was used as a reference for fragment sizes. The expected fragment sizes are listed in Table 8.3. MJK20.2= parental strain (Δ*ku80*). Clones with the prefix BS were stored at -80 °C. All other clones (prefix #) were discarded. The numbering of the clones refers to the entries in the lab journal and the in-house database.

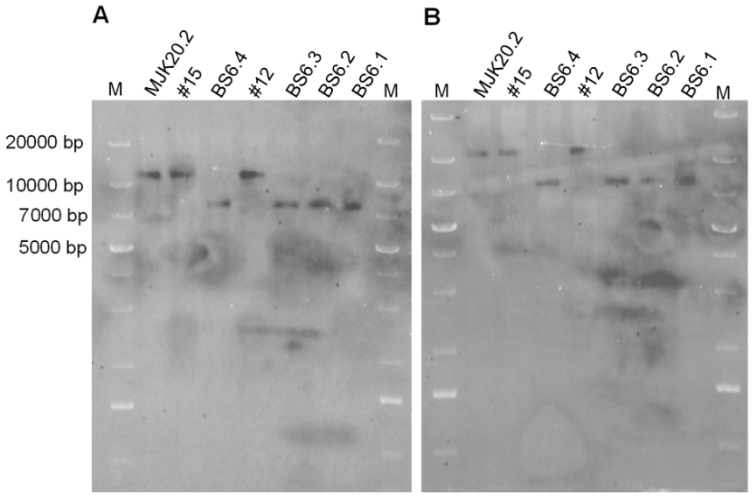

Figure 8.4: Results of the Southern analysis after deletion of *clr2* in MJK20.2. Shown are the results of the Southern analysis after marker recycling using the 5´probe (A) and 3´probe (B). Enzymatic digestion was performed with XbaI. The marker (M) was used as a reference for fragment sizes. The expected fragment sizes are listed in Table 8.3. MJK20.2= parental strain (Δ*ku80*). Clones with the prefix BS were stored at -80 °C. All other clones (prefix #) were discarded. The numbering of the clones refers to the entries in the lab journal and the in-house database.

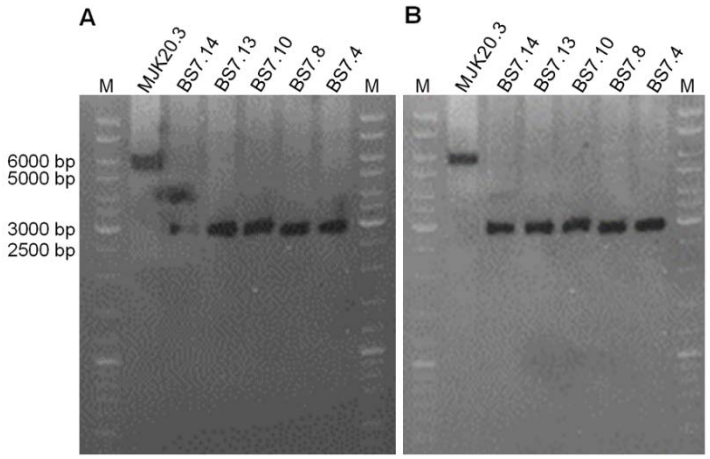

Figure 8.5: Results of the Southern analysis after deletion of *clr1* in MJK20.3. Shown are the results of the Southern analysis after marker recycling using the 5´probe (A) and 3´probe (B). Enzymatic digestion was performed with XmnI. The marker (M) was used as a reference for fragment sizes. The expected fragment sizes are listed in Table 8.3. MJK20.3= parental strain (Δ*ku80*). Clones with the prefix BS were stored at -80 °C. The numbering of the clones refers to the entries in the lab journal and the in-house database.

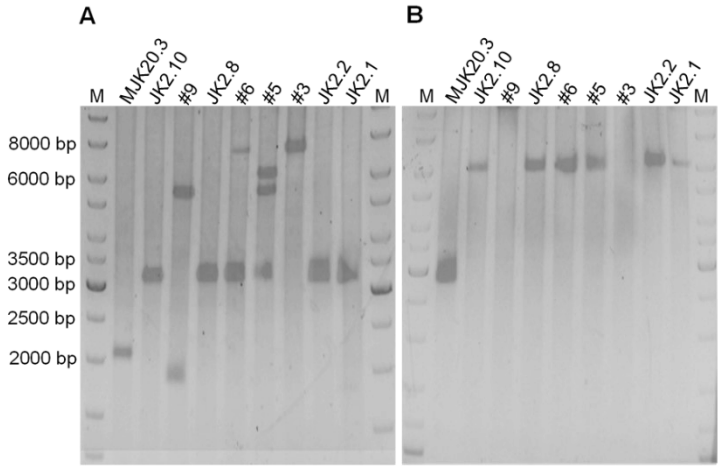

Figure 8.6: Results of the Southern analysis after deletion of *clr4* in MJK20.3. Shown are the results of the Southern analysis after marker recycling using a 5´probe. Enzymatic digestion was performed with SalI (A) and KpnI (B). The marker (M) was used as a reference for fragment sizes. The expected fragment sizes are listed in Table 8.3. MJK20.3= parental strain (Δ*ku80*). Clones with the prefix JK were stored at -80 °C. All other clones (prefix #) were discarded. The numbering of the clones refers to the entries in the lab journal and the in-house database.

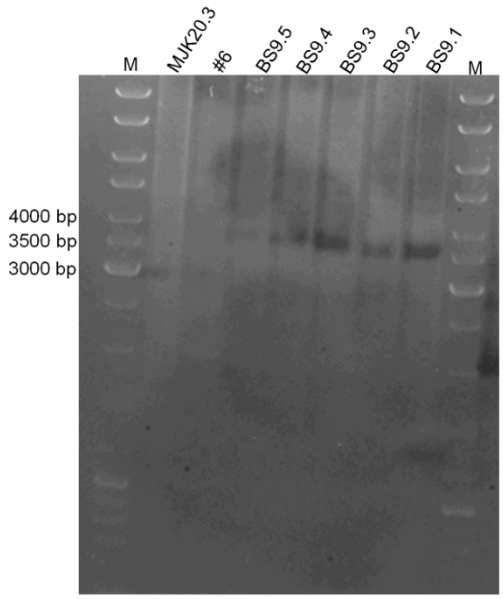

Figure 8.7: Results of the Southern analysis after deletion of *vib1* in MJK20.3. Shown are the results of the Southern analysis after marker recycling using a 5´probe. Enzymatic digestion was performed with SacI. The marker (M) was used as a reference for fragment sizes. The expected fragment sizes are listed in Table 8.3. MJK20.3= parental strain (Δ*ku80*). Clones with the prefix BS were stored at -80 °C. All other clones (prefix #) were discarded. The numbering of the clones refers to the entries in the lab journal and the in-house database.

Table 8.3: Expected fragment sizes of the Southern analysis of the regulator deletion clones in relation to the enzyme used for digestion and the resulting genotype. The parental strain (Δ*ku80*) was used as a negative control.

deleted gene	enzyme	genotype	fragment size [bp]
clr2	XbaI	Δ*ku80*	11879
		Δ*ku80*, Δ*clr2*	8180
vib2	XmnI	Δ*ku80*	7590
		Δ*ku80*, Δ*vib2*	5559
clr1	XmnI	Δ*ku80*	5660
		Δ*ku80*, Δ*clr1*	2940
clr4	SalI	Δ*ku80*	2150
		Δ*ku80*, Δ*clr4*	3274
	KpnI	Δ*ku80*	2919
		Δ*ku80*, Δ*clr4*	6054
vib1	SacI	Δ*ku80*	3000
		Δ*ku80*, Δ*vib*1	3726

8.4 SDS PAGEs

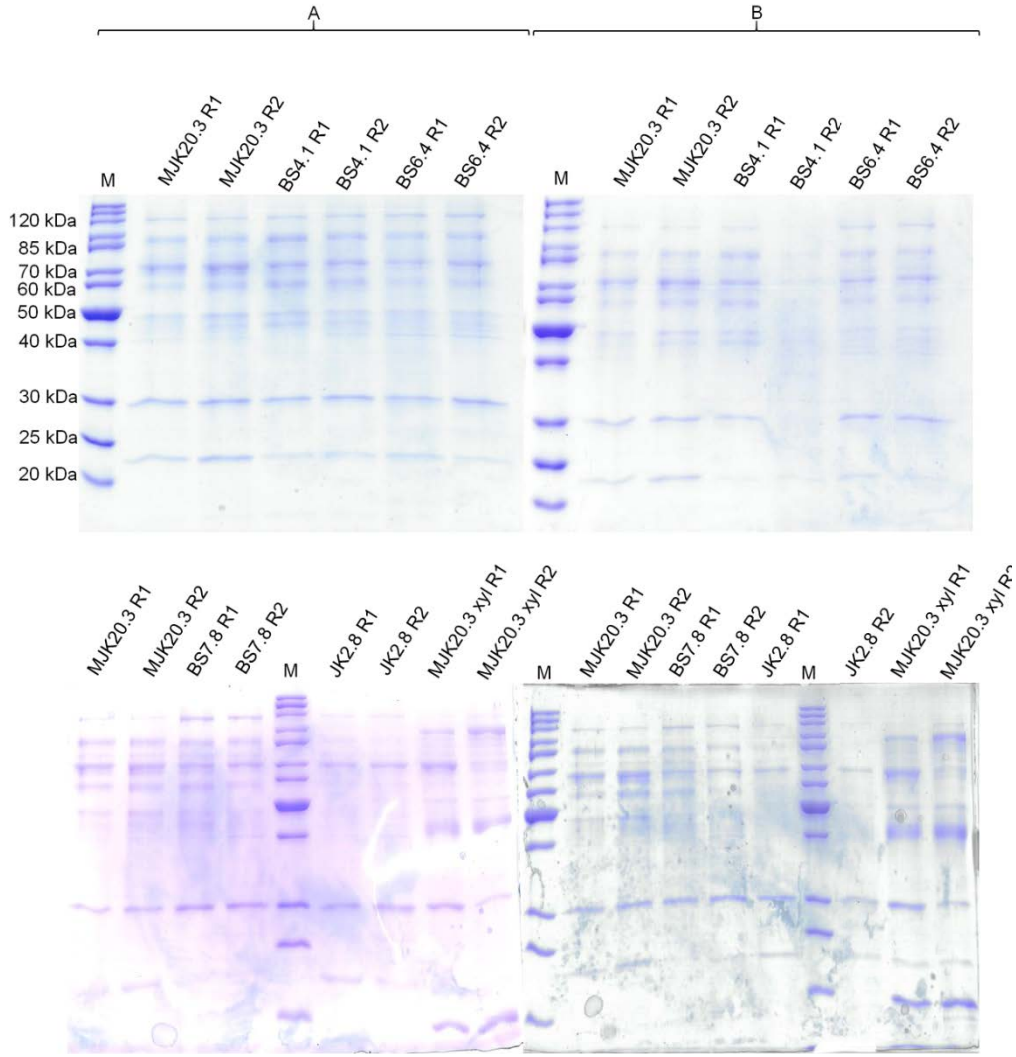

Figure 8.8: Protein analysis of culture supernatants at steady state via SDS-PAGE. Shown are the results of the SDS PAGEs with (A) 2 µg protein (determined via Bradford assay) and (B) 20 µL supernatant for the strains/conditions MJK20.3, BS4.1, BS6.4, BS7.8, JK2.8, and MJK20.3 xylose at steady state with two replicates (R1, R2). As a marker for protein size determination [kDa], PageRuler Unstained Protein Ladder (M) was used. If not mentioned separately, glucose was used as carbon source in the cultivations.

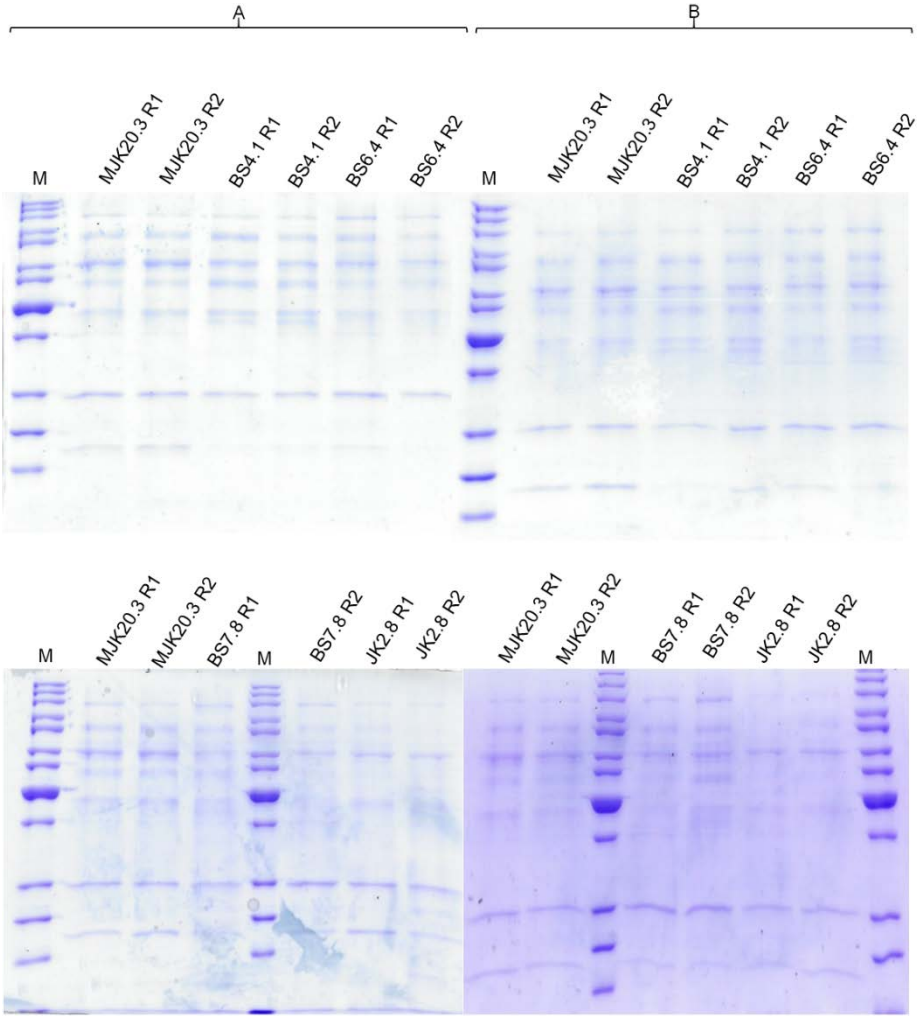

Figure 8.9: Protein analysis of culture supernatants 0.5 h after spiking cellulose via SDS-PAGE. Shown are the results of the SDS PAGEs with (A) 2 µg protein (determined via Bradford assay) and (B) 20 µL supernatant for the strains/conditions MJK20.3, BS4.1, BS6.4, BS7.8, JK2.8, and MJK20.3 xylose 0.5 h after spiking cellulose (t1) with two replicates (R1, R2). As a marker for protein size determination [kDa], PageRuler Unstained Protein Ladder (M) was used. If not mentioned separately, glucose was used as carbon source in the cultivations.

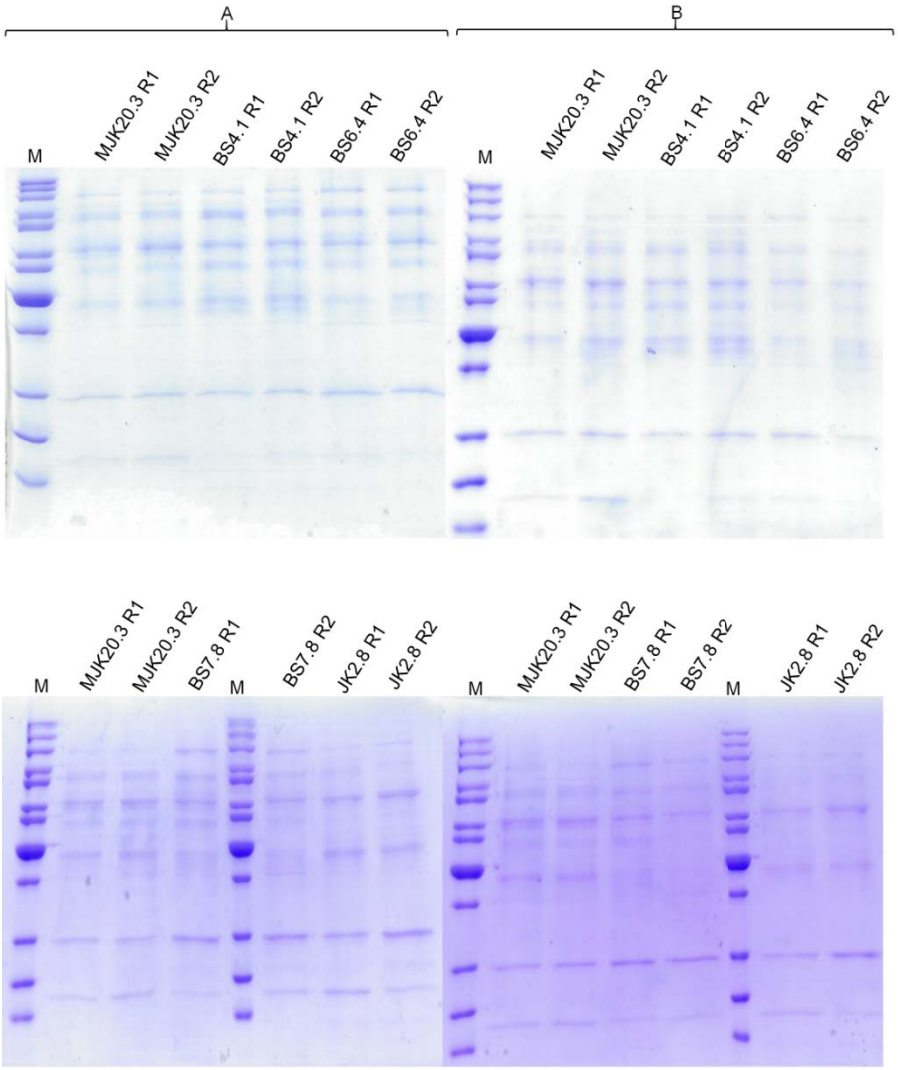

Figure 8.10: Protein analysis of culture supernatants 1 h after spiking cellulose via SDS-PAGE. Shown are the results of the SDS PAGEs with (A) 2 µg protein (determined via Bradford assay) and (B) 20 µL supernatant for the strains/conditions MJK20.3, BS4.1, BS6.4, BS7.8, JK2.8, and MJK20.3 xylose 1 h after spiking cellulose (t2) with two replicates (R1, R2). As a marker for protein size determination [kDa], PageRuler Unstained Protein Ladder (M) was used. If not mentioned separately, glucose was used as carbon source in the cultivations.

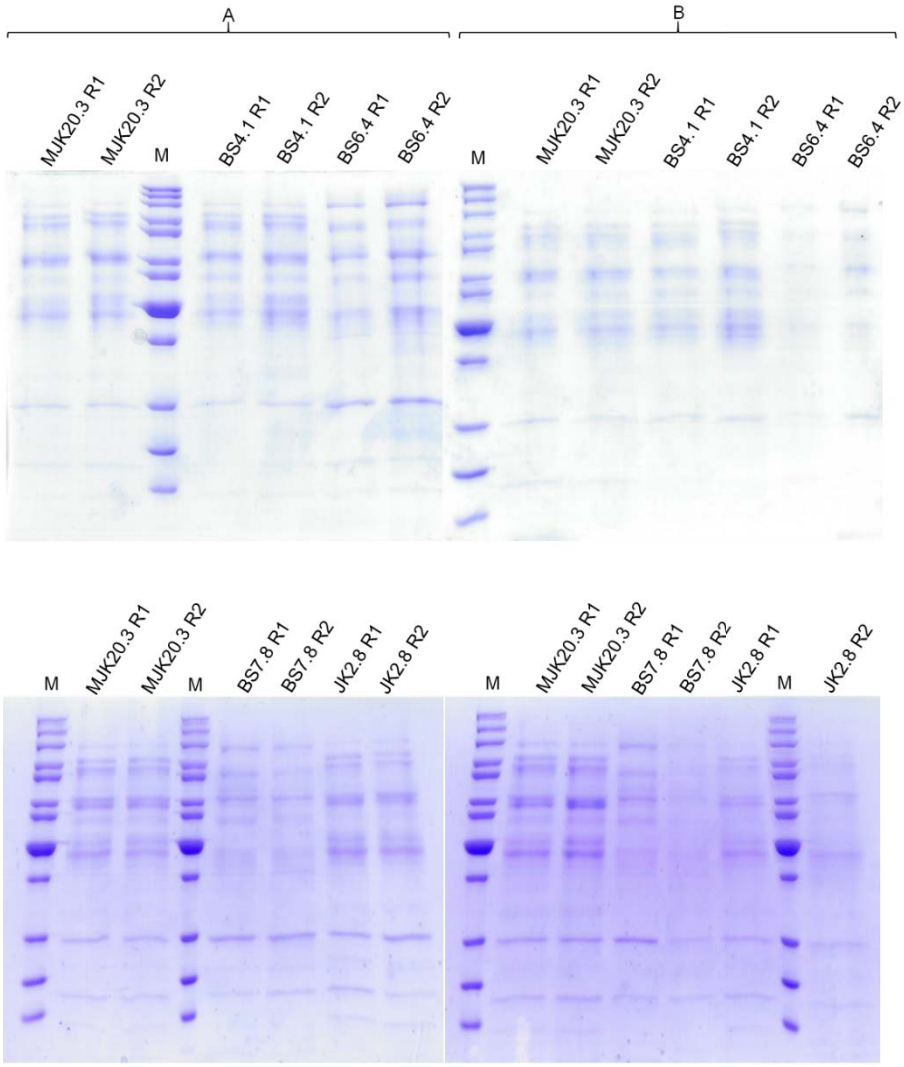

Figure 8.11: Protein analysis of culture supernatants 2 h after spiking cellulose via SDS-PAGE. Shown are the results of the SDS PAGEs with (A) 2 µg protein (determined via Bradford assay) and (B) 20 µL supernatant for the strains/conditions MJK20.3, BS4.1, BS6.4, BS7.8, JK2.8, and MJK20.3 xylose 2 h after spiking cellulose (t2) with two replicates (R1, R2). As a marker for protein size determination [kDa], PageRuler Unstained Protein Ladder (M) was used. If not mentioned separately, glucose was used as carbon source in the cultivations.

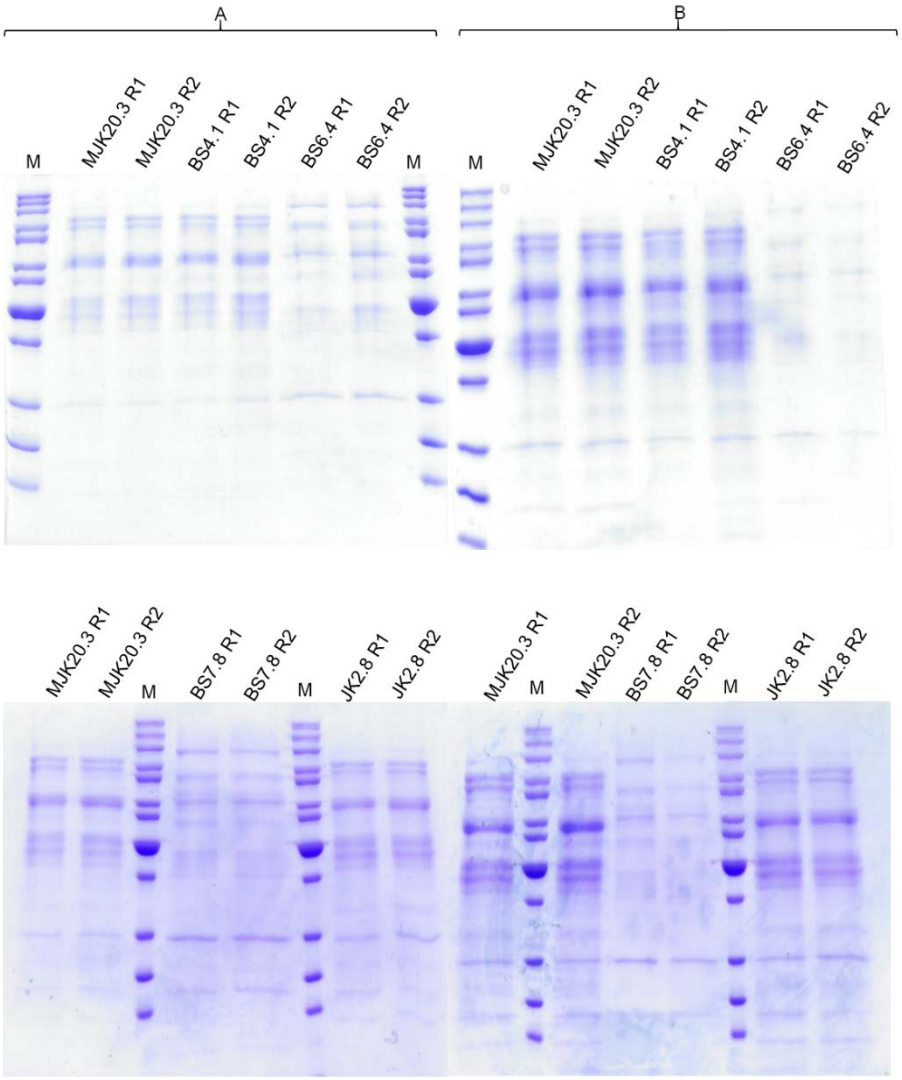

Figure 8.12: Protein analysis of culture supernatants 4 h after spiking cellulose via SDS-PAGE. Shown are the results of the SDS PAGEs with (A) 2 µg protein (determined via Bradford assay) and (B) 20 µL supernatant for the strains/conditions MJK20.3, BS4.1, BS6.4, BS7.8, JK2.8, and MJK20.3 xylose 4 h after spiking cellulose (t2) with two replicates (R1, R2). As a marker for protein size determination [kDa], PageRuler Unstained Protein Ladder (M) was used. If not mentioned separately, glucose was used as carbon source in the cultivations.

8.5 qPCR data

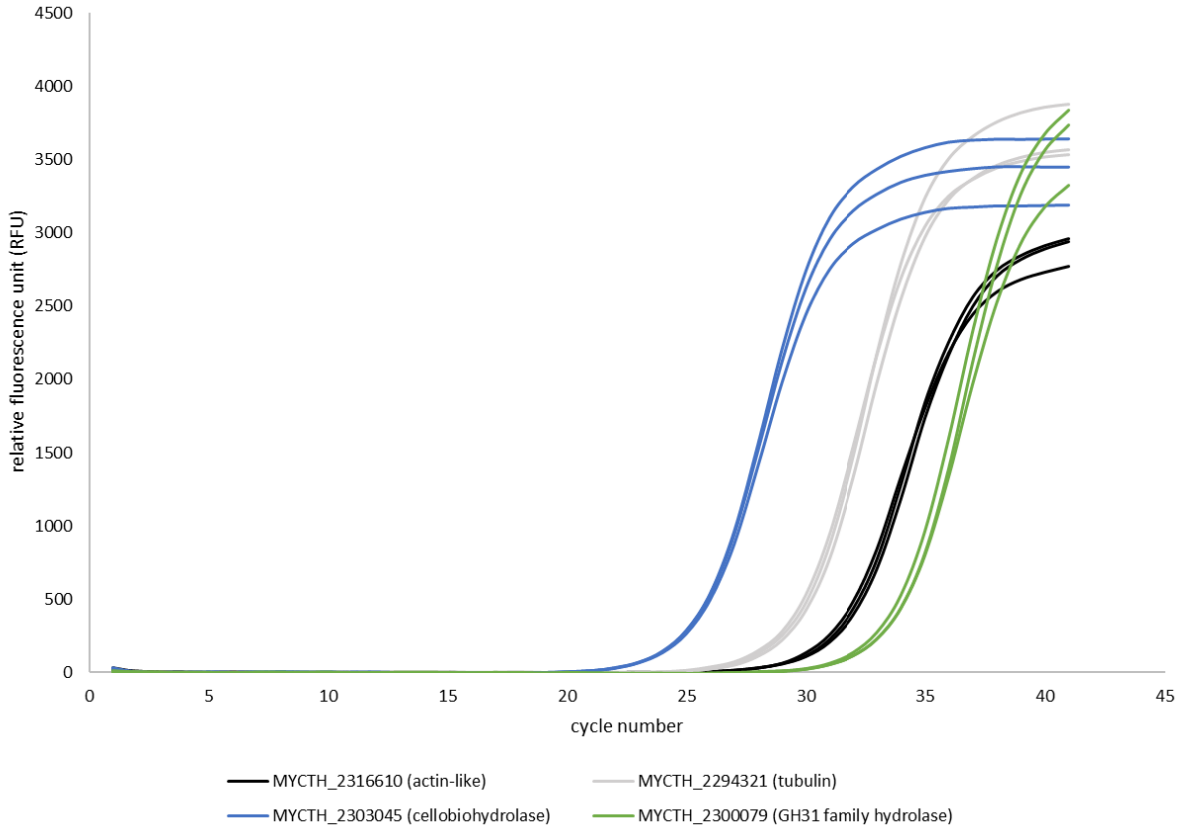

Figure 8.13: Quantitative real-time PCR to confirm the quality of the RNA seq. analysis. Shown are the relative fluorescence units (RFU) by cycle number. The genes that were examined include MYCTH_2316610 (actin-like), MYCTH_2294321 (tubulin), MYCTH_2303045 (cellobiohydrolase), and MYCTH_2300079 (GH31 family hydrolase).

Table 8.4: RNA seq. counts of the qPCR target genes in sample MJK20.3 R1, t2

target gene	counts in sample MJK20.3 R1, t2
MYCTH_2316610 (actin-like)	1107
MYCTH_2294321 (tubulin)	1829
MYCTH_2303045 (cellobiohydrolase)	48232
MYCTH_2300079 (GH 31 family hydrolase)	240

9. References

Agger JW, Isaksen T, Várnai A, Vidal-Melgosa S, Willats WGT, Ludwig R, Horn SJ, Eijsink VGH, Westereng B (2014) Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. Proc Natl Acad Sci U S A 111:6287–6292. https://doi.org/10.1073/pnas.1323629111

Alalwan HA, Alminshid AH, Aljaafari HAS (2019) Promising evolution of biofuel generations. Subject review. Renew Energy Focus 28:127–139. https://doi.org/10.1016/j.ref.2018.12.006

Alazi E, Niu J, Kowalczyk JE, Peng M, Aguilar Pontes MV, Van Kan JAL, Visser J, De Vries RP, Ram AFJ (2016) The transcriptional activator GaaR of *Aspergillus niger* is required for release and utilization of d-galacturonic acid from pectin. FEBS Lett 590:1804–1815. https://doi.org/10.1002/1873-3468.12211

Alonso DM, Bond JQ, Dumesic JA (2010) Catalytic conversion of biomass to biofuels. Green Chem 12:1493–1513. https://doi.org/10.1039/c004654j

Amore A, Giacobbe S, Faraco V (2013) Regulation of Cellulase and Hemicellulase Gene Expression in Fungi. Curr Genomics 14:230–249. https://doi.org/10.2174/1389202911314040002

Andrianopoulos A, Kourambas S, Sharp JA, Davis MA, Hynes MJ (1998) Characterization of the *Aspergillus nidulans nmrA* gene involved in nitrogen metabolite repression. J Bacteriol 180:1973–1977. https://doi.org/10.1128/jb.180.7.1973-1977.1998

Antoniêto ACC, Dos Santos Castro L, Silva-Rocha R, Persinoti GF, Silva RN (2014) Defining the genome-wide role of CRE1 during carbon catabolite repression in *Trichoderma reesei* using RNA-Seq analysis. Fungal Genet Biol 73:93–103. https://doi.org/10.1016/j.fgb.2014.10.009

Apinis AE (1963) Occurrence of thermophilous microfungi in certain alluvial soils near Nottingham. Nov Hedwigia 5:57–77

Arentshorst M, Lagendijk EL, Ram AF (2015) A new vector for efficient gene targeting to the *pyrG* locus in *Aspergillus niger*. Fungal Biol Biotechnol 2:2. https://doi.org/10.1186/s40694-015-0012-4

Arentshorst M, Ram AFJ, Meyer V (2012) Using non-homologous end-joining-deficient strains for functional gene analyses in filamentous fungi. In: Methods in Molecular Biology. Humana Press, pp 133–150

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

Aro N, Pakula T, Penttilä M (2005) Transcriptional regulation of plant cell wall degradation by filamentous fungi. FEMS Microbiol Rev 29:719–739. https://doi.org/10.1016/j.femsre.2004.11.006

Aro N, Saloheimo A, Ilmén M, Penttilä M (2001) ACEII, a Novel Transcriptional Activator Involved in Regulation of Cellulase and Xylanase Genes of *Trichoderma reesei*. J Biol Chem 276:24309–24314. https://doi.org/10.1074/jbc.M003624200

Arst HN, Cove DJ (1973) Nitrogen metabolite repression in *Aspergillus nidulans*. Mol Gen Genet 126:111–141. https://doi.org/10.1007/BF00330988

Bajracharya S, Vanbroekhoven K, Buisman CJN, Strik DPBTB, Pant D (2017) Bioelectrochemical conversion of CO_2 to chemicals: CO_2 as a next generation feedstock for electricity-driven bioproduction in batch and continuous modes. Faraday Discuss 202:433–449. https://doi.org/10.1039/c7fd00050b

Barrett L, Orlova M, Maziarz M, Kuchin S (2012) Protein kinase a contributes to the negative control of SNF1 protein kinase in *Saccharomyces cerevisiae*. Eukaryot Cell 11:119–128. https://doi.org/10.1128/EC.05061-11

Baruah J, Nath BK, Sharma R, Kumar S, Deka RC, Baruah DC, Kalita E (2018) Recent trends in the pretreatment of lignocellulosic biomass for value-added products. Front Energy Res 6:141. https://doi.org/10.3389/fenrg.2018.00141

Battaglia E, Hansen SF, Leendertse A, Madrid S, Mulder H, Nikolaev I, De Vries RP (2011) Regulation of pentose utilisation by AraR, but not XlnR, differs in *Aspergillus nidulans* and *Aspergillus niger*. Appl Microbiol Biotechnol 91:387–397. https://doi.org/10.1007/s00253-011-3242-2

Bayram Ö, Braus GH (2012) Coordination of secondary metabolism and development in fungi: The velvet family of regulatory proteins. FEMS Microbiol Rev 36:1–24. https://doi.org/10.1111/j.1574-6976.2011.00285.x

Bayram Ö, Krappmann S, Ni M, Jin WB, Helmstaedt K, Valerius O, Braus-Stromeyer S, Kwon NJ, Keller NP, Yu JH, Braus GH (2008) VelB/VeA/LaeA complex coordinates light signal with fungal development and secondary metabolism. Science 320:1504–1506. https://doi.org/10.1126/science.1155888

Beattie SR, Mark KMK, Thammahong A, Ries LNA, Dhingra S, Caffrey-Carr AK, Cheng C, Black CC, Bowyer P, Bromley MJ, Obar JJ, Goldman GH, Cramer RA (2017) Filamentous fungal carbon catabolite repression supports metabolic plasticity and stress responses essential for disease progression. PLOS Pathog 13:e1006340. https://doi.org/10.1371/journal.ppat.1006340

Bech L, Herbst F-A (2015) On-Site Enzyme Production by *Trichoderma asperellum* for the Degradation of Duckweed. Fungal Genomics Biol 5:126. https://doi.org/10.4172/2165-8056.1000126

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B 57:289–300. https://doi.org/10.2307/2346101

Benocci T, Aguilar-Pontes MV, Zhou M, Seiboth B, De Vries RP (2017) Regulators of plant biomass degradation in ascomycetous fungi. Biotechnol Biofuels 10:152. https://doi.org/10.1186/s13068-017-0841-x

Benz PJ, Chau BH, Zheng D, Bauer S, Glass NL, Somerville CR (2014) A comparative systems analysis of polysaccharide-elicited responses in *Neurospora crassa* reveals carbon source-specific cellular adaptations. Mol Microbiol 91:275–299. https://doi.org/10.1111/mmi.12459

Berezina O V., Herlet J, Rykov S V., Kornberger P, Zavyalov A, Kozlov D, Sakhibgaraeva L, Krestyanova I, Schwarz WH, Zverlov V V., Liebl W, Yarotsky S V. (2017) Thermostable multifunctional GH74 xyloglucanase from *Myceliophthora thermophila*: high-level expression in *Pichia pastoris* and characterization of the recombinant protein. Appl Microbiol Biotechnol 101:5653–5666. https://doi.org/10.1007/s00253-017-8297-2

Berger H, Basheer A, Böck S, Reyes-Dominguez Y, Dalik T, Altmann F, Strauss J (2008) Dissecting individual steps of nitrogen transcription factor cooperation in the *Aspergillus nidulans* nitrate cluster. Mol Microbiol 69:1385–1398. https://doi.org/10.1111/j.1365-2958.2008.06359.x

Berger H, Pachlinger R, Morozov I, Goller S, Narendja F, Caddick M, Strauss J (2006) The GATA factor AreA regulates localization and *in vivo* binding site occupancy of the nitrate activator NirA. Mol Microbiol 59:433–446. https://doi.org/10.1111/j.1365-2958.2005.04957.x

Berka RM, Grigoriev I V., Otillar R, Salamov A, Grimwood J, Reid I, Ishmael N, John T, Darmond C, Moisan MC, Henrissat B, Coutinho PM, Lombard V, Natvig DO, Lindquist E, Schmutz J, Lucas S, Harris P, Powlowski J, Bellemare A, Taylor D, Butler G, De Vries RP, Allijn IE, Van Den Brink J, Ushinsky S, Storms R, Powell AJ, Paulsen IT, Elbourne LDH, Baker SE, Magnuson J, Laboissiere S, Clutterbuck AJ, Martinez D, Wogulis M, De Leon AL, Rey MW, Tsang A (2011) Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. Nat Biotechnol 29:922–929. https://doi.org/10.1038/nbt.1976

Berka RM, Hayenga KJ, Lawlis VB, Ward M (2003) Aspartic proteinase deficient filamentous fungi. Patent number: US6509171B1.

Berka RM, Schneider P, Golightly EJ, Brown SH, Madden M, Brown KM, Halkier T, Mondorf K, Xu F (1997) Characterization of the gene encoding an extracellular laccase of *Myceliophthora thermophila* and analysis of the recombinant enzyme expressed in *Aspergillus oryzae*. Appl Environ Microbiol 63:3151–3157. https://doi.org/10.1128/aem.63.8.3151-3157.1997

Besson M, Gallezot P, Pinel C (2014) Conversion of biomass into chemicals over metal catalysts. Chem Rev 114:1827–1870. https://doi.org/10.1021/cr4002269

Bhatia SK, Kim SH, Yoon JJ, Yang YH (2017) Current status and strategies for second generation biofuel production using microbial systems. Energy Convers Manag 148:1142–1156. https://doi.org/10.1016/j.enconman.2017.06.073

Bi F, Barad S, Ment D, Luria N, Dubey A, Casado V, Glam N, Mínguez JD, Espeso EA, Fluhr R, Prusky D (2015) Carbon regulation of environmental pH by secreted small molecules that modulate pathogenicity in phytopathogenic fungi. Mol Plant Pathol 17:1178–1195. https://doi.org/10.1111/mpp.12355

Bibbins M, Crepin V, Cummings N, Mizote T, Baker K, Mellits K, Connerton I (2002) A regulator gene for acetate utilisation from *Neurospora crassa*. Mol Genet Genomics 267:498–505. https://doi.org/10.1007/s00438-002-0682-5

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

Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski J V., Conway JM, Adams MWW, Kelly RM (2014) Thermophilic lignocellulose deconstruction. FEMS Microbiol Rev 38:393–448. https://doi.org/10.1111/1574-6976.12044

Boase NA, Kelly JM (2004) A role for *creD*, a carbon catabolite repression gene from *Aspergillus nidulans*, in ubiquitination. Mol Microbiol 53:929–940. https://doi.org/10.1111/j.1365-2958.2004.04172.x

Borneman AR, Hynes MJ, Andrianopoulos A (2001) An STE12 homolog from the asexual, dimorphic fungus *Penicillium marneffei* complements the defect in sexual development of an *Aspergillus nidulans steA* mutant. Genetics 157:1003–1014.

Bricmont PA, Daugherty JR, Cooper TG (1991) The DAL81 gene product is required for induced expression of two differently regulated nitrogen catabolic genes in *Saccharomyces cerevisiae*. Mol Cell Biol 11:1161–1166. https://doi.org/10.1128/mcb.11.2.1161

Bridgwater A V. (1995) The technical and economic feasibility of biomass gasification for power generation. Fuel 74:631–653. https://doi.org/10.1016/0016-2361(95)00001-L

Brotman Y, Briff E, Viterbo A, Chet I (2008) Role of swollenin, an expansin-like protein from *Trichoderma*, in plant root colonization. Plant Physiol 147:779–789. https://doi.org/10.1104/pp.108.116293

Brown NA, De Gouvea PF, Krohn NG, Savoldi M, Goldman GH (2013) Functional characterisation of the non-essential protein kinases and phosphatases regulating *Aspergillus nidulans* hydrolytic enzyme production. Biotechnol Biofuels 6:91. https://doi.org/10.1186/1754-6834-6-91

Brown NA, Ries LNA, Goldman GH (2014) How nutritional status signalling coordinates metabolism and lignocellulolytic enzyme secretion. Fungal Genet Biol 72:48–63. https://doi.org/10.1016/j.fgb.2014.06.012

Brunner K, Lichtenauer AM, Kratochwill K, Delic M, Mach RL (2007) Xyr1 regulates xylanase but not cellulase formation in the head blight fungus *Fusarium graminearum*. Curr Genet 52:213–220. https://doi.org/10.1007/s00294-007-0154-x

Bulter T, Alcalde M, Sieber V, Meinhold P, Schlachtbauer C, Arnold FH (2003) Functional expression of a fungal laccase in *Saccharomyces cerevisiae* by directed evolution. Appl Environ Microbiol 69:987–995. https://doi.org/10.1128/AEM.69.2.987-995.2003

Caddick MX (1994) Nitrogen metabolite repression. Prog Ind Microbiol 29:323–353.

Caffall KH, Mohnen D (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydr Res 344:1879–1900. https://doi.org/10.1016/j.carres.2009.05.021

Calero-Nieto F, Di Pietro A, Roncero MIG, Hera C (2007) Role of the transcriptional activator XlnR of *Fusarium oxysporum* in regulation of xylanase genes and virulence. Mol Plant-Microbe Interact 20:977–985. https://doi.org/10.1094/MPMI-20-8-0977

Cao Y, Zheng F, Wang L, Zhao G, Chen G, Zhang W, Liu W (2017) Rce1, a novel transcriptional repressor, regulates cellulase gene expression by antagonizing the transactivator Xyr1 in *Trichoderma reesei*. Mol Microbiol 105:65–83. https://doi.org/10.1111/mmi.13685

Castellanos F, Schmoll M, Martínez P, Tisch D, Kubicek CP, Herrera-Estrella A, Esquivel-Naranjo EU (2010) Crucial factors of the light perception machinery and their impact on growth and cellulase gene transcription in *Trichoderma reesei*. Fungal Genet Biol 47:468–476. https://doi.org/10.1016/j.fgb.2010.02.001

CAZY database. Overview. http://www.cazy.org/Welcome-to-the-Carbohydrate-Active.html. Accessed 13 May 2020a

CAZY database. Glycoside hydrolases. http://www.cazy.org/Glycoside-Hydrolases.html. Accessed 13 May 2020b

CAZY database. Glycosyl transferases. http://www.cazy.org/GlycosylTransferases.html. Accessed 13 May 2020c

CAZY database. Carbohydrate esterases. http://www.cazy.org/Carbohydrate-Esterases.html. Accessed 13 May 2020d

CAZY database. Polysaccharide lyases. http://www.cazy.org/Polysaccharide-Lyases.html. Accessed 13 May 2020e

CAZY database. Auxiliary activities. http://www.cazy.org/Auxiliary-Activities.html. Accessed 13 May 2020f

CAZY database. Carbohydrate binding module. www.cazy.org/Carbohydrate-Binding-Modules.html. Accessed 13 May 2020g

CAZY database. AA9/GH61. http://www.cazy.org/AA9.html. Accessed 13 May 2020h

CAZY database. *Neurospora crassa* OR74A CAZYs. http://www.cazy.org/eN.html. Accessed 15 May 2020i

CAZY database. *Aspergillus niger* CBS 513.88 CAZYs. http://www.cazy.org/e525.html. Accessed 15 May 2020j

Cepeda-García C, Domínguez-Santos R, García-Rico RO, García-Estrada C, Cajiao A, Fierro F, Martín JF (2014) Direct involvement of the CreA transcription factor in penicillin biosynthesis and expression of the *pcbAB* gene in *Penicillium chrysogenum*. Appl Microbiol Biotechnol 98:7113–7124. https://doi.org/10.1007/s00253-014-5760-1

Chakraborti S, Chakraborti T, Dhalla NS (2017) Proteases in human diseases. Springer Singapore

Chang L, Ding M, Bao L, Chen Y, Zhou J, Lu H (2011) Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. Appl Microbiol Biotechnol 90:1933–1942. https://doi.org/10.1007/s00253-011-3182-x

Chen H, Wang L (2017) Enzymatic Hydrolysis of Pretreated Biomass. In: Technologies for Biochemical Conversion of Biomass. Elsevier, pp 65–99

Chen L, Zou G, Wang J, Wang J, Liu R, Jiang Y, Zhao G, Zhou Z (2016) Characterization of the Ca^{2+}-responsive signaling pathway in regulating the expression and secretion of cellulases in *Trichoderma reesei* Rut-C30. Mol Microbiol 100:560–575. https://doi.org/10.1111/mmi.13334

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

Cherubini F (2010) The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. Energy Convers Manag 51:1412–1421. https://doi.org/10.1016/j.enconman.2010.01.015

Cherubini F, Jungmeier G, Wellisch M, Willke T, Skiadas I, Van Ree R, De Jong E (2009) Toward a common classification approach for biorefinery systems. Biofuels, Bioprod Biorefining 3:534–546. https://doi.org/10.1002/bbb.172

Chiang YM, Szewczyk E, Davidson AD, Keller N, Oakley BR, Wang CCC (2009) A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in *Aspergillus nidulans*. J Am Chem Soc 131:2965–2970. https://doi.org/10.1021/ja8088185

Cho DW, Kwon EE, Song H (2016) Use of carbon dioxide as a reaction medium in the thermo-chemical process for the enhanced generation of syngas and tuning adsorption ability of biochar. Energy Convers Manag 117:106–114. https://doi.org/10.1016/j.enconman.2016.03.027

Christensen U, Gruben BS, Madrid S, Mulder H, Nikolaev I, De Vries RP (2011) Unique regulatory mechanism for D-galactose utilization in *Aspergillus nidulans*. Appl Environ Microbiol 77:7084–7087. https://doi.org/10.1128/AEM.05290-11

Chylenski P, Bissaro B, Sørlie M, Røhr ÅK, Várnai A, Horn SJ, Eijsink VGH (2019) Lytic Polysaccharide Monooxygenases in Enzymatic Processing of Lignocellulosic Biomass. ACS Catal 9:4970–4991. https://doi.org/10.1021/acscatal.9b00246

Chylenski P, Forsberg Z, Ståhlberg J, Várnai A, Lersch M, Bengtsson O, Sæbø S, Horn SJ, Eijsink VGH (2017) Development of minimal enzyme cocktails for hydrolysis of sulfite-pulped lignocellulosic biomass. J Biotechnol 246:16–23. https://doi.org/10.1016/j.jbiotec.2017.02.009

Climent MJ, Corma A, Iborra S (2014) Conversion of biomass platform molecules into fuel additives and liquid hydrocarbon fuels. Green Chem 16:516–547. https://doi.org/10.1039/c3gc41492b

Conesa A, Punt PJ, Van Den Hondel CAMJJ (2002) Fungal peroxidases: Molecular aspects and applications. J Biotechnol 93:143–158. https://doi.org/10.1016/S0168-1656(01)00394-7

Conlon H, Zadra I, Haas H, Arst HN, Jones MG, Caddick MX (2001) The *Aspergillus nidulans* GATA transcription factor gene *areB* encodes at least three proteins and features three classes of mutation. Mol Microbiol 40:361–375. https://doi.org/10.1046/j.1365-2958.2001.02399.x

Bio based Industries Consortium. Mapping European Biorefineries. https://biconsortium.eu/news/mapping-european-biorefineries. Accessed 21 Sep 2020

Coradetti ST, Craig JP, Xiong Y, Shock T, Tian C, Glass NL (2012) Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. Proc Natl Acad Sci U S A 109:7397–7402. https://doi.org/10.1073/pnas.1200785109

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

Cosgrove DJ (2000) Loosening of plant cell walls by expansins. Nature 407:321–326. https://doi.org/10.1038/35030000

Cosgrove DJ, Li LC, Cho HT, Hoffmann-Benning S, Moore RC, Blecker D (2002) The growing world of expansins. Plant Cell Physiol 43:1436–1444. https://doi.org/10.1093/pcp/pcf180

Coutinho EC, Corrêa LC (1999) The induction of sporulation in the aquatic fungus *Blastocladiella emersonii* is dependent on extracellular calcium. FEMS Microbiol Lett 179:353–359. https://doi.org/10.1111/j.1574-6968.1999.tb08749.x

Couturier M, Ladevèze S, Sulzenbacher G, Ciano L, Fanuel M, Moreau C, Villares A, Cathala B, Chaspoul F, Frandsen KE, Labourel A, Herpoël-Gimbert I, Grisel S, Haon M, Lenfant N, Rogniaux H, Ropartz D, Davies GJ, Rosso MN, Walton PH, Henrissat B, Berrin JG (2018) Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. Nat Chem Biol 14:306–310. https://doi.org/10.1038/nchembio.2558

Craig JP, Coradetti ST, Starr TL, Louise Glass N (2015) Direct target network of the *Neurospora crassa* plant cell wall deconstruction regulators Clr-1, Clr-2, andXlr-1. MBio 6:e01452. https://doi.org/10.1128/mBio.01452-15

Czernik S, Bridgwater A V. (2004) Overview of applications of biomass fast pyrolysis oil. Energy and Fuels 18:590–598. https://doi.org/10.1021/ef034067u

Cziferszky A, Mach RL, Kubicek CP (2002) Phosphorylation positively regulates DNA binding of the carbon catabolite repressor Cre1 of *Hypocrea jecorina* (*Trichoderma reesei*). J Biol Chem 277:14688–14694. https://doi.org/10.1074/jbc.M200744200

Cziferszky A, Seiboth B, Kubicek CP (2003) The Snf1 kinase of the filamentous fungus *Hypocrea jecorina* phosphorylates regulation-relevant serine residues in the yeast carbon catabolite repressor Mig1 but not in the filamentous fungal counterpart Cre1. Fungal Genet Biol 40:166–175. https://doi.org/10.1016/S1087-1845(03)00082-3

De Assis LJ, Ries LNA, Savoldi M, Dos Reis TF, Brown NA, Goldman GH (2015) *Aspergillus nidulans* protein kinase A plays an important role in cellulase production. Biotechnol Biofuels 8:213. https://doi.org/10.1186/s13068-015-0401-1

De Graaff LK, Van den Broeck HC, Van Ooijen AJJ, Visser J (1994) Regulation of the xylanase-encoding *xlnA* gene of *Aspergillus tubigensis*. Mol Microbiol 12:479–490. https://doi.org/10.1111/j.1365-2958.1994.tb01036.x

De Souza WR, De Gouvea PF, Savoldi M, Malavazi I, De Souza Bernardes LA, Goldman MHS, De Vries RP, De Castro Oliveira J V., Goldman GH (2011) Transcriptome analysis of *Aspergillus niger* grown on sugarcane bagasse. Biotechnol Biofuels 4:40. https://doi.org/10.1186/1754-6834-4-40

De Vries RP, Jansen J, Aguilar G, Pařenicová L, Joosten V, Wülfert F, Benen JAE, Visser J (2002) Expression profiling of pectinolytic genes from *Aspergillus niger*. FEBS Lett 530:41–47. https://doi.org/10.1016/S0014-5793(02)03391-4

De Vries RP, Visser J (2001) *Aspergillus* Enzymes Involved in Degradation of Plant Cell Wall Polysaccharides. Microbiol Mol Biol Rev 65:497–522. https://doi.org/10.1128/mmbr.65.4.497-522.2001

Denison SH (2000) pH regulation of gene expression in fungi. Fungal Genet Biol 29:61–71. https://doi.org/10.1006/fgbi.2000.1188

Denton JA, Kelly JM (2011) Disruption of *Trichoderma reesei cre2*, encoding an ubiquitin C-terminal hydrolase, results in increased cellulase activity. BMC Biotechnol 11:103. https://doi.org/10.1186/1472-6750-11-103

Derntl C, Gudynaite-Savitch L, Calixte S, White T, MacH RL, MacH-Aigner AR (2013) Mutation of the Xylanase regulator 1 causes a glucose blind hydrolase expressing phenotype in industrially used *Trichoderma* strains. Biotechnol Biofuels 6:62. https://doi.org/10.1186/1754-6834-6-62

Derntl C, Kluger B, Bueschl C, Schuhmacher R, Mach RL, Mach-Aigner AR (2017) Transcription factor Xpp1 is a switch between primary and secondary fungal metabolism. Proc Natl Acad Sci U S A 114:E560–E569. https://doi.org/10.1073/pnas.1609348114

Derntl C, Rassinger A, Srebotnik E, Mach RL, Mach-Aigner AR (2015) Xpp1 regulates the expression of xylanases, but not of cellulases in *Trichoderma reesei*. Biotechnol Biofuels 8:112. https://doi.org/10.1186/s13068-015-0298-8

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Dos Santos Gomes AC, Falkoski D, Battaglia E, Peng M, Nicolau de Almeida M, Coconi Linares N, Meijnen J-P, Visser J, De Vries RP (2019) *Myceliophthora thermophila* Xyr1 is predominantly involved in xylan degradation and xylose catabolism. Biotechnol Biofuels 12:220. https://doi.org/10.1186/s13068-019-1556-y

Drag M, Salvesen GS (2010) Emerging principles in protease-based drug discovery. Nat Rev Drug Discov 9:690–701. https://doi.org/10.1038/nrd3053

Dyadic International website. C1 expression system. https://www.dyadic.com/c1-technology/c1-expression-system. Accessed 27 Apr 2020a

Dyadic International website. C1 cellulase enzyme. https://www.dyadic.com/dyadic-international-completes-fda-gras-notification-process-for-c1-derived-cellulase-enzyme. Accessed 27 Apr 2020b

Emalfarb MA, Ben-Bassat A, Burlingame RP, Mikhaylovich Chernoglazov V, Nicolaevich O (1998) Cellulase compositions and methods of use. Patent number: US5811381A.

Empel J, Sitkiewicz I, Andrukiewicz A, Lasocki K, Borsuk P, Weglenski P (2001) *arcA*, the regulatory gene for the arginine catabolic pathway in *Aspergillus nidulans*. Mol Genet Genomics 266:591–597. https://doi.org/10.1007/s004380100575

ExPASy Bioinformatics Resource Portal. ENZYME – The Enzyme Data Bank. https://enzyme.expasy.org/cgi-bin/enzyme/enzyme-search-cl?3. Accessed 12 May 2020

Fan F, Ma G, Li J, Liu Q, Benz JP, Tian C, Ma Y (2015) Genome-wide analysis of the endoplasmic reticulum stress response during lignocellulase production in *Neurospora crassa*. Biotechnol Biofuels 8:66. https://doi.org/10.1186/s13068-015-0248-5

Fekete E, Karaffa L, Karimi Aghcheh R, Németh Z, Fekete É, Orosz A, Paholcsek M, Stágel A, Kubicek CP (2014) The transcriptome of *lae1* mutants of trichoderma reesei cultivated at constant growth rates reveals new targets of LAE1 function. BMC Genomics 15:447. https://doi.org/10.1186/1471-2164-15-447

Filiatrault-Chastel C, Navarro D, Haon M, Grisel S, Herpoël-Gimbert I, Chevret D, Fanuel M, Henrissat B, Heiss-Blanquet S, Margeot A, Berrin JG (2019) AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes. Biotechnol Biofuels 12:55. https://doi.org/10.1186/s13068-019-1394-y

Fincher GB (2016) Chemistry and Physicochemistry of Nonstarchy Polysaccharides. In: Reference Module in Food Science. Elsevier

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 (2003) Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. J Biol Chem 278:31988–31997. https://doi.org/10.1074/jbc.M304750200

Forsberg Z, Vaaje-kolstad G, Westereng B, Bunsæ AC, Stenstrøm Y, Mackenzie A, Sørlie M, Horn SJ, Eijsink VGH (2011) Cleavage of cellulose by a Cbm33 protein. Protein Sci 20:1479–1483. https://doi.org/10.1002/pro.689

Francisco CS, Ma X, Zwyssig MM, McDonald B, Palma-Guerrero J (2018) Coping with stress: morphological changes in response to environmental stimuli in a fungal plant pathogen. bioRxiv 372078. https://doi.org/10.1101/372078

Frommhagen M, Sforza S, Westphal AH, Visser J, Hinz SWA, Koetsier MJ, Van Berkel WJH, Gruppen H, Kabel MA (2015) Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase. Biotechnol Biofuels 8:101. https://doi.org/10.1186/s13068-015-0284-1

Fu YH, Marzluf GA (1990) *nit-2*, the major nitrogen regulatory gene of *Neurospora crassa*, encodes a protein with a putative zinc finger DNA-binding domain. Mol Cell Biol 10:1056–1065. https://doi.org/10.1128/mcb.10.3.1056

Fu YH, Marzluf GA (1987) Characterization of *nit-2*, the major nitrogen regulatory gene of *Neurospora crassa*. Mol Cell Biol 7:1691–1696. https://doi.org/10.1128/mcb.7.5.1691

Fujii T, Inoue H, Ishikawa K (2013) Enhancing cellulase and hemicellulase production by genetic modification of the carbon catabolite repressor gene, *creA*, in *Acremonium cellulolyticus*. AMB Express 3:73. https://doi.org/10.1186/2191-0855-3-73

Fujii T, Inoue H, Ishikawa K (2015) Decreased Cellulase and Xylanase Production in the Fungus *Talaromyces cellulolyticus* by Disruption of *tacA* and *tctA* Genes, Encoding Putative Zinc Finger Transcriptional Factors. Appl Biochem Biotechnol 175:3218–3229. https://doi.org/10.1007/s12010-015-1497-2

Fujinaga M, Cherney MM, Oyama H, Oda K, James MNG (2004) The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*. Proc Natl Acad Sci U S A 101:3364–3369. https://doi.org/10.1073/pnas.0400246101

Galluzzi L, Vitale I, Aaronson SA, Abrams JM, Adam D, Agostinis P, Alnemri ES, Altucci L, Amelio I, Andrews DW, Annicchiarico-Petruzzelli M, Antonov A V., Arama E, Baehrecke EH, Barlev NA, Bazan NG, Bernassola F, Bertrand MJM, Bianchi K, Blagosklonny M V., Blomgren K, Borner C, Boya P, Brenner C, Campanella M, Candi E, Carmona-Gutierrez D, Cecconi F, Chan FKM, Chandel NS, Cheng EH, Chipuk JE, Cidlowski JA, Ciechanover A, Cohen GM, Conrad M, Cubillos-Ruiz JR, Czabotar PE, D'Angiolella V, Dawson TM, Dawson VL, De Laurenzi V, De Maria R, Debatin KM, Deberardinis RJ, Deshmukh M, Di Daniele N, Di Virgilio F, Dixit VM, Dixon SJ, Duckett CS, Dynlacht BD, El-Deiry WS, Elrod JW, Fimia GM, Fulda S, García-Sáez AJ, Garg AD, Garrido C, Gavathiotis E, Golstein P, Gottlieb E, Green DR, Greene LA, Gronemeyer H, Gross A, Hajnoczky G, Hardwick JM, Harris IS, Hengartner MO, Hetz C, Ichijo H, Jäättelä M, Joseph B, Jost PJ, Juin PP, Kaiser WJ, Karin M, Kaufmann T, Kepp O, Kimchi A, Kitsis RN, Klionsky DJ, Knight RA, Kumar S, Lee SW, Lemasters JJ, Levine B, Linkermann A, Lipton SA, Lockshin RA, López-Otín C, Lowe SW, Luedde T, Lugli E, MacFarlane M, Madeo

F, Malewicz M, Malorni W, Manic G, Marine JC, Martin SJ, Martinou JC, Medema JP, Mehlen P, Meier P, Melino S, Miao EA, Molkentin JD, Moll UM, Muñoz-Pinedo C, Nagata S, Nuñez G, Oberst A, Oren M, Overholtzer M, Pagano M, Panaretakis T, Pasparakis M, Penninger JM, Pereira DM, Pervaiz S, Peter ME, Piacentini M, Pinton P, Prehn JHM, Puthalakath H, Rabinovich GA, Rehm M, Rizzuto R, Rodrigues CMP, Rubinsztein DC, Rudel T, Ryan KM, Sayan E, Scorrano L, Shao F, Shi Y, Silke J, Simon HU, Sistigu A, Stockwell BR, Strasser A, Szabadkai G, Tait SWG, Tang D, Tavernarakis N, Thorburn A, Tsujimoto Y, Turk B, Vanden Berghe T, Vandenabeele P, Vander Heiden MG, Villunger A, Virgin HW, Vousden KH, Vucic D, Wagner EF, Walczak H, Wallach D, Wang Y, Wells JA, Wood W, Yuan J, Zakeri Z, Zhivotovsky B, Zitvogel L, Melino G, Kroemer G (2018) Molecular mechanisms of cell death: Recommendations of the Nomenclature Committee on Cell Death 2018. Cell Death Differ 25:486–541. https://doi.org/10.1038/s41418-017-0012-4

Gambelli D, Alberti F, Solfanelli F, Vairo D, Zanoli R (2017) Third generation algae biofuels in Italy by 2030: A scenario analysis using Bayesian networks. Energy Policy 103:165–178. https://doi.org/10.1016/j.enpol.2017.01.013

Garcia-Ochoa F, Gomez E, Santos VE, Merchuk JC (2010) Oxygen uptake rate in microbial processes: An overview. Biochem Eng J 49:289–307. https://doi.org/10.1016/j.bej.2010.01.011

García I, Gonzalez R, Gómez D, Scazzocchio C (2004) Chromatin Rearrangements in the *prnD-prnB* Bidirectional Promoter: Dependence on Transcription Factors. Eukaryot Cell 3:144–156. https://doi.org/10.1128/EC.3.1.144-156.2004

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345. https://doi.org/10.1038/nmeth.1318

Goldar MM, Jeong HT, Tanaka K, Matsuda H, Kawamukai M (2005) Moc3, a novel Zn finger type protein involved in sexual development, ascus formation, and stress response of *Schizosaccharomyces pombe*. Curr Genet 48:345–355. https://doi.org/10.1007/s00294-005-0028-z

Gomi K, Akeno T, Minetoki T, Ozeki K, Kumagai C, Okazaki N, Iimura Y (2000) Molecular cloning and characterization of a transcriptional activator gene, *amyR*, involved in the amylolytic gene expression in *Aspergillus oryzae*. Biosci Biotechnol Biochem 64:816–827. https://doi.org/10.1271/bbb.64.816

Gonçalves RD, Cupertino FB, Freitas FZ, Luchessi AD, Bertolini MC (2011) A genome-wide screen for *Neurospora crassa* transcription factors regulating glycogen metabolism. Mol Cell Proteomics 10:M111.007963. https://doi.org/10.1074/mcp.M111.007963

Gruben BS, Zhou M, Wiebenga A, Ballering J, Overkamp KM, Punt PJ, De Vries RP (2014) *Aspergillus niger* RhaR, a regulator involved in L-rhamnose release and catabolism. Appl Microbiol Biotechnol 98:5531–5540. https://doi.org/10.1007/s00253-014-5607-9

Guadix A;, Guadix EM, ; Páez-Dueñas MP;, González-Tello PY, Camacho F (2000) Technological processes and methods of control in the hydrolysis of proteins. Ars Pharm 41:79–89.

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

Hadley G, Harrold CE (1958) The sporulation of *Penicillium notatum* westling in submerged liquid culture: I. The effect of calcium and nutrients on sporulation intensity. J Exp Bot 9:408–417. https://doi.org/10.1093/jxb/9.3.408

Haefner S, Thywissen A, Hartmann H, Boehmer N (2017a) Method of producing proteins in filamentous fungi with decreased Clr1 activity. Patent number: WO2017093451A1.

Haefner S, Thywissen A, Hartmann H, Boehmer N (2017b) Method of producing proteins in filamentous fungi with decreased CLR2 activity. Patent number:WO2017093450A1.

Häkkinen M, Valkonen MJ, Westerholm-Parvinen A, Aro N, Arvas M, Vitikainen M, Penttilä M, Saloheimo M, Pakula TM (2014) Screening of candidate regulators for cellulase and hemicellulase production in *Trichoderma reesei* and identification of a factor essential for cellulase production. Biotechnol Biofuels 7:14. https://doi.org/10.1186/1754-6834-7-14

Hartley BS (1960) Proteolytic enzymes. Annu Rev Biochem 29:45–72. https://doi.org/10.1146/annurev.bi.29.070160.000401

Hasegawa S, Takizawa M, Suyama H, Shintani T, Gomi K (2010) Characterization and expression analysis of a maltose-utilizing (MAL) cluster in *Aspergillus oryzae*. Fungal Genet Biol 47:1–9. https://doi.org/10.1016/j.fgb.2009.10.005

Hasper AA, Visser J, De Graaff LH (2000) The *Aspergillus niger* transcriptional activator XlnR, which is involved in the degradation of the polysaccharides xylan and cellulose, also regulates D-xylose reductase gene expression. Mol Microbiol 36:193–200. https://doi.org/10.1046/j.1365-2958.2000.01843.x

Hasunuma T, Okazaki F, Okai N, Hara KY, Ishii J, Kondo A (2013) A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. Bioresour Technol 135:513–522. https://doi.org/10.1016/j.biortech.2012.10.047

He R, Ma L, Li C, Jia W, Li D, Zhang D, Chen S (2014) Trpac1, a pH response transcription regulator, is involved in cellulase gene expression in *Trichoderma reesei*. Enzyme Microb Technol 67:17–26. https://doi.org/10.1016/j.enzmictec.2014.08.013

Hector RE, Mertens JA, Bowman MJ, Nichols NN, Cotta MA, Hughes SR (2011) *Saccharomyces cerevisiae* engineered for xylose metabolism requires gluconeogenesis and the oxidative branch of the pentose phosphate pathway for aerobic xylose assimilation. Yeast 28:645–660. https://doi.org/10.1002/yea.1893

Hemsworth GR, Johnston EM, Davies GJ, Walton PH (2015) Lytic Polysaccharide Monooxygenases in Biomass Conversion. Trends Biotechnol 33:747–761. https://doi.org/10.1016/j.tibtech.2015.09.006

Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VGH (2012) Novel enzymes for the degradation of cellulose. Biotechnol Biofuels 5:45. https://doi.org/10.1186/1754-6834-5-45

Hu J, Arantes V, Pribowo A, Saddler JN (2013) The synergistic action of accessory enzymes enhances the hydrolytic potential of a “cellulase mixture” but is highly substrate specific. Biotechnol Biofuels 6:112. https://doi.org/10.1186/1754-6834-6-112

Hu J, Chandra R, Arantes V, Gourlay K, Susan van Dyk J, Saddler JN (2015) The addition of accessory enzymes enhances the hydrolytic performance of cellulase enzymes at high solid loadings. Bioresour Technol 186:149–153. https://doi.org/10.1016/j.biortech.2015.03.055

Huang DW, Sherman BT, Lempicki RA (2009a) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1–13. https://doi.org/10.1093/nar/gkn923

Huang DW, Sherman BT, Lempicki RA (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57. https://doi.org/10.1038/nprot.2008.211

Huber GW, Iborra S, Corma A (2006) Synthesis of transportation fuels from biomass: Chemistry, catalysts, and engineering. Chem Rev 106:4044–4098. https://doi.org/10.1021/cr068360d

Huberman LB, Liu J, Qin L, Glass NL (2016) Regulation of the lignocellulolytic response in filamentous fungi. Fungal Biol Rev 30:101–111. https://doi.org/10.1016/j.fbr.2016.06.001

Humbird D, Davis R, Tao L, Kinchin C, Hsu D, Aden A, Schoen P, Lukas J, Olthof B, Worley M, Sexton D, Dudgeon D (2011) Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol: Dilute-Acid Pretreatment and Enzymatic Hydrolysis of Corn Stover. Technical Report for National Renewable Energy Laboratory

Hunter AJ, Morris TA, Jin B, Saint CP, Kelly JM (2013) Deletion of *creB* in *Aspergillus oryzae* increases secreted hydrolytic enzyme activity. Appl Environ Microbiol 79:5480–5487. https://doi.org/10.1128/AEM.01406-13

Hurley J, Loros JJ, Dunlap JC (2015) Dissecting the Mechanisms of the Clock in *Neurospora*. Methods Enzymol 551:29–52. https://doi.org/10.1016/bs.mie.2014.10.009

Huuskonen A (2020) Development of the filamentous fungus *Myceliophthora thermophila* C1 into a next-generation therapeutic protein production system. ECFG Rome

Hynes MJ, Kelly JM (1977) Pleiotropic mutants of *Aspergillus nidulans* altered in carbon metabolism. Mol Gen Genet 150:193–204. https://doi.org/10.1007/BF00695399

Idnurm A, Heitman J (2010) Ferrochelatase is a conserved downstream target of the blue light-sensing White collar complex in fungi. Microbiology 156:2393–2407. https://doi.org/10.1099/mic.0.039222-0

IEA Bioenergy. https://www.iea-bioenergy.task42-biorefineries.com/en/ieabiorefinery/Partners.htm. Accessed 21 Sep 2020

Igarashi K, Wada M, Samejima M (2007) Activation of crystalline cellulose to cellulose IIII results in efficient hydrolysis by cellobiohydrolase. FEBS J 274:1785–1792. https://doi.org/10.1111/j.1742-4658.2007.05727.x

Isaksen T, Westereng B, Aachmann FL, Agger JW, Kracher D, Kittl R, Ludwig R, Haltrich D, Eijsink VGH, Horn SJ (2014) A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides. J Biol Chem 289:2632–2642. https://doi.org/10.1074/jbc.M113.530196

Jeon JR, Chang YS (2013) Laccase-mediated oxidation of small organics: Bifunctional roles for versatile applications. Trends Biotechnol 31:335–341. https://doi.org/10.1016/j.tibtech.2013.04.002

Jia Z, Zhang X, Cao X (2009) Effects of carbon sources on fungal morphology and lovastatin biosynthesis by submerged cultivation of *Aspergillus terreus*. Asia-Pacific J Chem Eng 4:672–677. https://doi.org/10.1002/apj.316

Johansen KS (2016) Discovery and industrial applications of lytic polysaccharide mono-oxygenases. Biochem Soc Trans 44:143–149. https://doi.org/10.1042/BST20150204

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 (2013) Functional analyses of *Trichoderma reesei* LAE1 reveal conserved and contrasting roles of this regulator. G3 Genes, Genomes, Genet 3:369–378. https://doi.org/10.1534/g3.112.005140

Karimi-Aghcheh R, Németh Z, Atanasova L, Fekete E, Paholcsek M, Sándor E, Aquino B, Druzhinina IS, Karaffa L, Kubicek CP (2014) The VELVET a orthologue VEL1 of *Trichoderma reesei* regulates fungal development and is essential for cellulase gene expression. PLoS One 9:e112799. https://doi.org/10.1371/journal.pone.0112799

Karnaouri AC, Topakas E, Antonopoulou I, Christakopoulos P (2014a) Genomic insights into the fungal lignocellulolytic system of *Myceliophthora thermophila*. Front Microbiol 5:281. https://doi.org/10.3389/fmicb.2014.00281

Karnaouri AC, Topakas E, Christakopoulos P (2014b) Cloning, expression, and characterization of a thermostable GH7 endoglucanase from *Myceliophthora thermophila* capable of high-consistency enzymatic liquefaction. Appl Microbiol Biotechnol 98:231–242. https://doi.org/10.1007/s00253-013-4895-9

Kaufman RJ (1999) Stress signaling from the lumen of the endoplasmic reticulum: Coordination of gene transcriptional and translational controls. Genes Dev 13:1211–1233. https://doi.org/10.1101/gad.13.10.1211

Kelly JM, Hynes MJ (1985) Transformation of *Aspergillus niger* by the *amdS* gene of *Aspergillus nidulans*. EMBO J 4:475–479. https://doi.org/10.1002/j.1460-2075.1985.tb03653.x

Kelly JM, Hynes MJ (1977) Increased and decreased sensitivity to carbon catabolite repression of enzymes of acetate metabolism in mutants of *Aspergillus nidulans*. Mol Gen Genet 156:87–92. https://doi.org/10.1007/BF00272256

Kern M, McGeehan JE, Streeter SD, Martin RNA, Besser K, Elias L, Eborall W, Malyon GP, Payne CM, Himmel ME, Schnorr K, Beckham GT, Cragg SM, Bruce NC, McQueen-Mason SJ (2013) Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. Proc Natl Acad Sci U S A 110:10189–10194. https://doi.org/10.1073/pnas.1301502110

Kim HS, Han KY, Kim KJ, Han DM, Jahng KY, Chae KS (2002) The *veA* gene activates sexual development in *Aspergillus nidulans*. Fungal Genet Biol 37:72–80. https://doi.org/10.1016/S1087-1845(02)00029-4

Klaubauf S, Zhou M, Lebrun MH, De Vries RP, Battaglia E (2016) A novel l-arabinose-responsive regulator discovered in the rice-blast fungus *Pyricularia oryzae* (*Magnaporthe oryzae*). FEBS Lett 590:550–558. https://doi.org/10.1002/1873-3468.12070

Kolbusz MA, Di Falco M, Ishmael N, Marqueteau S, Moisan MC, Baptista C da S, Powlowski J, Tsang A (2014) Transcriptome and exoproteome analysis of utilization of plant-derived biomass by *Myceliophthora thermophila*. Fungal Genet Biol 72:10–20. https://doi.org/10.1016/j.fgb.2014.05.006

Konwar LJ, Mikkola JP, Bordoloi N, Saikia R, Chutia RS, Kataki R (2018) Sidestreams from bioenergy and biorefinery complexes as a resource for circular bioeconomy. In: Waste Biorefinery: Potential and Perspectives. Elsevier, pp 85–125.

Kowalczyk JE, Benoit I, De Vries RP (2014) Regulation of Plant Biomass Utilization in *Aspergillus*. In: Advances in Applied Microbiology. Academic Press Inc., pp 31–56.

Kowalczyk JE, Gruben BS, Battaglia E, Wiebenga A, Majoor E, De Vries RP (2015) Genetic interaction of *Aspergillus nidulans galR*, *xlnR* and *araR* in regulating D-galactose and L-arabinose release and catabolism gene expression. PLoS One 10:e0143200. https://doi.org/10.1371/journal.pone.0143200

Kubicek CP, Messner R, Gruber F, Mach RL, Kubicek-Pranz EM (1993) The *Trichoderma* cellulase regulatory puzzle: From the interior life of a secretory fungus. Enzyme Microb Technol 15:90–99. https://doi.org/10.1016/0141-0229(93)90030-6

Kudla B, Caddick MX, Langdon T, Martinez-Rossi NM, Bennett CF, Sibley S, Davies RW, Arst HN (1990) The regulatory gene *areA* mediating nitrogen metabolite repression in *Aspergillus nidulans*. Mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger. EMBO J 9:1355–1364. https://doi.org/10.1002/j.1460-2075.1990.tb08250.x

Kunitake E, Hagiwara D, Miyamoto K, Kanamaru K, Kimura M, Kobayashi T (2016) Regulation of genes encoding cellulolytic enzymes by Pal-PacC signaling in *Aspergillus nidulans*. Appl Microbiol Biotechnol 100:3621–3635. https://doi.org/10.1007/s00253-016-7409-8

Kunitake E, Kawamura A, Tani S, Takenaka S, Ogasawara W, Sumitani JI, Kawaguchi T (2015) Effects of *clbR* overexpression on enzyme production in *Aspergillus aculeatus* vary depending on the cellulosic biomass-degrading enzyme species. Biosci Biotechnol Biochem 79:488–495. https://doi.org/10.1080/09168451.2014.982501

Kunitake E, Tani S, Sumitani JI, Kawaguchi T (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*. Appl Microbiol Biotechnol 97:2017–2028. https://doi.org/10.1007/s00253-012-4305-8

Kwon MJ, Schütze T, Spohner S, Haefner S, Meyer V (2019) Practical guidance for the implementation of the CRISPR genome editing tool in filamentous fungi. Fungal Biol Biotechnol 6:15. https://doi.org/10.1186/s40694-019-0079-4

Lai Y, Deng T, Liu G, Wang J (2017) The influence of homologous overexpression of *bglR* on β-glucosidase activities in *Myceliophthora thermophila*. China Biotechnol 37:64–71. https://doi.org/10.13523/J.CB.20170712

Laird DA, Brown RC, Amonette JE, Lehmann J (2009) Review of the pyrolysis platform for coproducing bio-oil and biochar. Biofuels, Bioprod Biorefining 3:547–562. https://doi.org/10.1002/bbb.169

Laluce C, Schenberg ACG, Gallardo JCM, Coradello LFC, Pombeiro-Sponchiado SR (2012) Advances and developments in strategies to improve strains of *Saccharomyces cerevisiae* and processes to obtain the lignocellulosic ethanol - A review. Appl Biochem Biotechnol 166:1908–1926. https://doi.org/10.1007/s12010-012-9619-6

Lamb HK, Ren J, Park A, Johnson C, Leslie K, Cocklin S, Thompson P, Mee C, Cooper A, Stammers DK, Hawkins AR (2004) Modulation of the ligand binding properties of the transcription repressor NmrA by GATA-containing DNA and site-directed mutagenesis. Protein Sci 13:3127–3138. https://doi.org/10.1110/ps.04958904

Landowski C, Huuskonen A, Westerholm-Parvinen A, Saloheimo M, Kanerva A, Hiltunen J (2013) Multiple Proteases Deficient Filamentous Fungal Cells and Methods of Use Thereof. Patent number: US20200087696A1.

Lange L (2017) Fungal Enzymes and Yeasts for Conversion of Plant Biomass to Bioenergy and High-Value Products. Microbiol Spectr 5. https://doi.org/10.1128/microbiolspec.funk-0007-2016

Lee J, Myong K, Kim JE, Kim HK, Yun SH, Lee YW (2012) FgVelB globally regulates sexual reproduction, mycotoxin production and pathogenicity in the cereal pathogen *Fusarium graminearum*. Microbiology 158:1723–1733. https://doi.org/10.1099/mic.0.059188-0

Lehmbeck J, Udagawa H (2008) Fungal PepC inhibitor. Patent number: US8633010B2.

Lei Y, Liu G, Yao G, Li Z, Qin Y, Qu Y (2016) A novel bZIP transcription factor ClrC positively regulates multiple stress responses, conidiation and cellulase expression in *Penicillium oxalicum*. Res Microbiol 167:424–435. https://doi.org/10.1016/j.resmic.2016.03.001

Li H, Fang Z, Smith RL, Yang S (2016) Efficient valorization of biomass to biofuels with bifunctional solid catalytic materials. Prog Energy Combust Sci 55:98–194. https://doi.org/10.1016/j.pecs.2016.04.004

Li J, Lin L, Li H, Tian C, Ma Y (2014) Transcriptional comparison of the filamentous fungus *Neurospora crassa* growing on three major monosaccharides D-glucose, D-xylose and L-arabinose. Biotechnol Biofuels 7:31. https://doi.org/10.1186/1754-6834-7-31

Li J, Lin L, Sun T, Xu J, Ji J, Liu Q, Tian C (2019) Direct production of commodity chemicals from lignocellulose using *Myceliophthora thermophila*. Metab Eng. https://doi.org/10.1016/j.ymben.2019.05.007

Li J, Zhang Y, Li J, Sun T, Tian C (2020a) Metabolic engineering of the cellulolytic thermophilic fungus *Myceliophthora thermophila* to produce ethanol from cellobiose. Biotechnol Biofuels 13:23. https://doi.org/10.1186/s13068-020-1661-y

Li Q, Yi L, Marek P, Iverson BL (2013) Commercial proteases: Present and future. FEBS Lett 587:1155–1163. https://doi.org/10.1016/j.febslet.2012.12.019

Li X, Liu Q, Sun W, He Q, Tian C (2020b) Improving cellulases production by *Myceliophthora thermophila* through disruption of protease genes. Biotechnol Lett 42:219–229. https://doi.org/10.1007/s10529-019-02777-0

Li Z, Yao G, Wu R, Gao L, Kan Q, Liu M, Yang P, Liu G, Qin Y, Song X, Zhong Y, Fang X, Qu Y (2015) Synergistic and Dose-Controlled Regulation of Cellulase Gene Expression in *Penicillium oxalicum*. PLoS Genet 11:e1005509. https://doi.org/10.1371/journal.pgen.1005509

Lin L, Wang S, Li X, He Q, Philipp Benz J, Tian C (2019) STK-12 acts as a transcriptional brake to control the expression of cellulase-encoding genes in *Neurospora crassa*. PLoS Genet 15:e1008510. https://doi.org/10.1371/journal.pgen.1008510

Liu F, Xue Y, Liu J, Gan L, Long M (2018) ACE3 as a master transcriptional factor regulates cellulase and xylanase production in *Trichoderma orientalis* EU7-22. BioResources 13:6790–6801. https://doi.org/10.15376/biores.13.3.6790-6801

Liu Q, Gao R, Li J, Lin L, Zhao J, Sun W, Tian C (2017) Development of a genome-editing CRISPR/Cas9 system in thermophilic fungal *Myceliophthora* species and its application to hyper-cellulase production strain engineering. Biotechnol Biofuels 10:1. https://doi.org/10.1186/s13068-016-0693-9

Liu Q, Li J, Gao R, Li J, Ma G, Tian C (2019a) CLR-4, a novel conserved transcription factor for cellulase gene expression in ascomycete fungi. Mol Microbiol 111:373–394. https://doi.org/10.1111/mmi.14160

Liu Q, Zhang Y, Li F, Li J, Sun W, Tian C (2019b) Upgrading of efficient and scalable CRISPR-Cas-mediated technology for genetic engineering in thermophilic fungus *Myceliophthora thermophila*. Biotechnol Biofuels 12:293. https://doi.org/10.1186/s13068-019-1637-y

Lockington RA, Kelly JM (2002) The WD40-repeat protein CreC interacts with and stabilizes the deubiquitinating enzyme CreB *in vivo* in *Aspergillus nidulans*. Mol Microbiol 43:1173–1182. https://doi.org/10.1046/j.1365-2958.2002.02811.x

Lombraña M, Moralejo FJ, Pinto R, Martín JF (2004) Modulation of *Aspergillus awamori* thaumatin secretion by modification of *bipA* gene expression. Appl Environ Microbiol 70:5145–5152. https://doi.org/10.1128/AEM.70.9.5145-5152.2004

López-Otín C, Bond JS (2008) Proteases: Multifunctional enzymes in life and disease. J Biol Chem 283:30433–30437. https://doi.org/10.1074/jbc.R800035200

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8

Löwe J, Stock D, Jap B, Zwickl P, Baumeister W, Huber R (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. Science 268:533–539. https://doi.org/10.1126/science.7725097

Ma R, Xu Y, Zhang X (2015) Catalytic oxidation of biorefinery lignin to value-added chemicals to support sustainable biofuel production. ChemSusChem 8:24–51. https://doi.org/10.1002/cssc.201402503

Mach-Aigner AR, Pucher ME, Mach RL (2010) D-xylose as a repressor or inducer of xylanase expression in *Hypocrea jecorina* (*Trichoderma reesei*). Appl Environ Microbiol 76:1770–1776. https://doi.org/10.1128/AEM.02746-09

Mach-Aigner AR, Pucher ME, Steiger MG, Bauer GE, Preis SJ, Mach RL (2008) Transcriptional regulation of *xyr1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. Appl Environ Microbiol 74:6554–6562. https://doi.org/10.1128/AEM.01143-08

Macios M, Caddick MX, Weglenski P, Scazzocchio C, Dzikowska A (2012) The GATA factors AREA and AREB together with the co-repressor NMRA, negatively regulate arginine catabolism in *Aspergillus nidulans* in response to nitrogen and carbon source. Fungal Genet Biol 49:189–198. https://doi.org/10.1016/j.fgb.2012.01.004

Maheshwari R, Bharadwaj G, Bhat MK (2000) Thermophilic Fungi: Their Physiology and Enzymes. Microbiol Mol Biol Rev 64:461–488. https://doi.org/10.1128/mmbr.64.3.461-488.2000

Maity SK (2015) Opportunities, recent trends and challenges of integrated biorefinery: Part I. Renew Sustain Energy Rev 43:1427–1445. https://doi.org/10.1016/j.rser.2014.11.092

Mancuso F M (2010) Extracting biological meaning from large gene list with DAVID-Short Tutorial. https://de.slideshare.net/framancuso/david-5451863. Accessed 16 Oct 2020

Marin-Felix Y, Stchigel AM, Miller AN, Guarro J, Cano-Lira JF (2015) A re-evaluation of the genus *Myceliophthora* (Sordariales, Ascomycota): Its segregation into four genera and description of *Corynascus fumimontanus sp. nov*. Mycologia 107:619–632. https://doi.org/10.3852/14-228

Martínez-Medina GA, Barragán AP, Ruiz HA, Ilyina A, Hernández JLM, Rodríguez-Jasso RM, Hoyos-Concha JL, Aguilar-González CN (2019) Fungal proteases and production of bioactive peptides for the food industry. In: Enzymes in Food Biotechnology: Production, Applications, and Future Prospects. Elsevier, pp 221–246.

Marui J, Kitamoto N, Kato M, Kobayashi T, Tsukagoshi N (2002) Transcriptional activator, AoXlnR, mediates cellulose-inductive expression of the xylanolytic and cellulolytic genes in *Aspergillus oryzae*. FEBS Lett 528:279–282. https://doi.org/10.1016/S0014-5793(02)03328-8

Marzluf GA (1997) Genetic regulation of nitrogen metabolism in the fungi. Microbiol Mol Biol Rev 61:17–32. https://doi.org/10.1128/.61.1.17-32.1997

Mello-de-Sousa TM, Rassinger A, Derntl C, J. Poças-Fonseca M, L Mach R, R Mach-Aigner A (2016) The Relation Between Promoter Chromatin Status, Xyr1 and Cellulase Expression in *Trichoderma reesei*. Curr Genomics 17:145–152. https://doi.org/10.2174/1389202917666151116211812

Mello-de-Sousa TM, Rassinger A, Pucher ME, Dos Santos Castro L, Persinoti GF, Silva-Rocha R, Poças-Fonseca MJ, Mach RL, Nascimento Silva R, Mach-Aigner AR (2015) The impact of chromatin remodelling on cellulase expression in *Trichoderma reesei*. BMC Genomics 16:588. https://doi.org/10.1186/s12864-015-1807-7

Meyer V, Basenko EY, Benz JP, Braus GH, Caddick MX, Csukai M, De Vries RP, Endy D, Frisvad JC, Gunde-Cimerman N, Haarmann T, Hadar Y, Hansen K, Johnson RI, Keller NP, Kraševec N, Mortensen UH, Perez R, Ram AFJ, Record E, Ross P, Shapaval V, Steiniger C, Van Den Brink H, Van Munster J, Yarden O, Wösten HAB (2020) Growing a circular economy with fungal biotechnology: A white paper. Fungal Biol Biotechnol 7:5. https://doi.org/10.1186/s40694-020-00095-z

Mikkola JP, Sklavounos E, King AWT, Virtanen P (2015) CHAPTER 1: The biorefinery and green chemistry. In: RSC Green Chemistry. Royal Society of Chemistry, pp 1–37.

Moncada J, Tamayo JA, Cardona CA (2014) Integrating first, second, and third generation biorefineries: Incorporating microalgae into the sugarcane biorefinery. Chem Eng Sci 118:126–140. https://doi.org/10.1016/j.ces.2014.07.035

Montenegro-Montero A, Goity A, Larrondo LF (2015) The bZIP transcription factor HAC-1 is involved in the unfolded protein response and is necessary for growth on cellulose in *Neurospora crassa*. PLoS One 10:e0131415. https://doi.org/10.1371/journal.pone.0131415

Moralejo FJ, Cardoza RE, Gutierrez S, Lombraña M, Fierro F, Martíni JF (2002) Silencing of the aspergillopepsin B (*pepB*) gene of *Aspergillus awamori* by antisense RNA expression or protease removal by gene disruption results in a large increase in thaumatin production. Appl Environ Microbiol 68:3550–3559. https://doi.org/10.1128/AEM.68.7.3550-3559.2002

Mori K, Kawahara T, Yoshida H, Yanagi H, Yura T (1996) Signalling from endoplasmic reticulum to nucleus: Transcription factor with a basic-leucine zipper motif is required for the unfolded protein-response pathway. Genes Cells 1:803–817. https://doi.org/10.1046/j.1365-2443.1996.d01-274.x

Murakoshi Y, Makita T, Kato M, Kobayashi T (2012) Comparison and characterization of α-amylase inducers in *Aspergillus nidulans* based on nuclear localization of AmyR. Appl Microbiol Biotechnol 94:1629–1635. https://doi.org/10.1007/s00253-012-3874-x

Nakari-Setälä T, Paloheimo M, Kallio J, Vehmaanperä J, Penttilä M, Saloheimo M (2009) Genetic modification of carbon catabolite repression in *Trichoderma reesei* for improved protein production. Appl Environ Microbiol 75:4853–4860. https://doi.org/10.1128/AEM.00282-09

Narendja FM, Davis MA, Hynes MJ (1999) AnCF, the CCAAT Binding Complex of *Aspergillus nidulans*, Is Essential for the Formation of a DNase I-Hypersensitive Site in the 5′ Region of the *amdS* Gene. Mol Cell Biol 19:6523–6531. https://doi.org/10.1128/mcb.19.10.6523

Nguyen KA, Wikee S, Lumyong S (2018) Brief review: Lignocellulolytic enzymes from polypores for efficient utilization of biomass. Mycosphere 9:1073–1088. https://doi.org/10.5943/mycosphere/9/6/2

Nielsen ML, Albertsen L, Lettier G, Nielsen JB, Mortensen UH (2006) Efficient PCR-based gene targeting with a recyclable marker for *Aspergillus nidulans*. Fungal Genet Biol 43:54–64. https://doi.org/10.1016/j.fgb.2005.09.005

Nitta M, Furukawa T, Shida Y, Mori K, Kuhara S, Morikawa Y, Ogasawara W (2012) A new Zn(II) 2Cys 6-type transcription factor BglR regulates β-glucosidase expression in *Trichoderma reesei*. Fungal Genet Biol 49:388–397. https://doi.org/10.1016/j.fgb.2012.02.009

Niu J, Alazi E, Reid ID, Arentshorst M, Punt PJ, Visser J, Tsang A, Ram AFJ (2017) An evolutionarily conserved transcriptional activator-repressor module controls expression of genes for D-Galacturonic acid utilization in *Aspergillus niger*. Genetics 205:169–183. https://doi.org/10.1534/genetics.116.194050

Oda K (2012) New families of carboxyl peptidases: Serine-carboxyl peptidases and glutamic peptidases. J Biochem 151:13–25. https://doi.org/10.1093/jb/mvr129

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*. Biosci Biotechnol Biochem 77:426–429. https://doi.org/10.1271/bbb.120795

Opar A (2010) Excitement grows for potential revolution in hepatitis C virus treatment. Nat Rev Drug Discov 9:501–503. https://doi.org/10.1038/nrd3214

Orejas M, MacCabe AP, Pérez-González JA, Kumar S, Ramón D (2001) The wide-domain carbon catabolite repressor CreA indirectly controls expression of the *Aspergillus nidulans xlnB* gene, encoding the acidic endo-β-(1,4)-xylanase X24. J Bacteriol 183:1517–1523. https://doi.org/10.1128/JB.183.5.1517-1523.2001

Orejas M, MacCabe AP, Pérez González JA, Kumar S, Ramón D (1999) Carbon catabolite repression of the *Aspergillus nidulans xlnA* gene. Mol Microbiol 31:177–184. https://doi.org/10.1046/j.1365-2958.1999.01157.x

Paietta J V. (2008) DNA-binding specificity of the CYS3 transcription factor of *Neurospora crassa* defined by binding-site selection. Fungal Genet Biol 45:1166–1171. https://doi.org/10.1016/j.fgb.2008.05.001

Paschkowsky S, Hsiao JM, Young JC, Munter LM (2019) The discovery of proteases and intramembrane proteolysis. Biochem Cell Biol 97:265–269. https://doi.org/10.1139/bcb-2018-0186

Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, Ståhlberg J, Beckham GT (2015) Fungal cellulases. Chem Rev 115:1308–1448. https://doi.org/10.1021/cr500351c

Peñalva MA, Arst HN (2002) Regulation of Gene Expression by Ambient pH in Filamentous Fungi and Yeasts. Microbiol Mol Biol Rev 66:426–446. https://doi.org/10.1128/mmbr.66.3.426-446.2002

Petersen KL, Lehmbeck J, Christensen T (1999) A new transcriptional activator for amylase genes in *Aspergillus*. Mol Gen Genet 262:668–676. https://doi.org/10.1007/s004380051129

Phillips CM, Beeson WT, Cate JH, Marletta MA (2011) Cellobiose Dehydrogenase and a Copper-Dependent Polysaccharide Monooxygenase Potentiate Cellulose Degradation by *Neurospora crassa*. ACS Chem Biol 6:1399–1406. https://doi.org/10.1021/cb200351y

Pitt D, Ugalde UO (1984) Calcium in fungi. Plant, Cell Environ 7:467–475. https://doi.org/10.1111/j.1365-3040.1984.tb01437.x

Pollegioni L, Tonin F, Rosini E (2015) Lignin-degrading enzymes. FEBS J 282:1190–1213. https://doi.org/10.1111/febs.13224

Ponnusamy VK, Nguyen DD, Dharmaraja J, Shobana S, Banu JR, Saratale RG, Chang SW, Kumar G (2018) A review on lignin structure, pretreatments, fermentation reactions and biorefinery potential. Bioresour Technol 271:462–472. https://doi.org/10.1016/j.biortech.2018.09.070

Porc O, Hark N, Carus M, Dammer L, Carrez D (2020) European Bioeconomy in Figures 2008-2017. Bio based Industries Consortium.

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

Prathumpai W, McIntyre M, Nielsen J (2004) The effect of CreA in glucose and xylose catabolism in *Aspergillus nidulans*. Appl Microbiol Biotechnol 63:748–753. https://doi.org/10.1007/s00253-003-1409-1

Proietto M, Bianchi MM, Ballario P, Brenna A (2015) Epigenetic and posttranslational modifications in light signal transduction and the circadian clock in *Neurospora crassa*. Int J Mol Sci 16:15347–15383. https://doi.org/10.3390/ijms160715347

Purschwitz J, Müller S, Fischer R (2009) Mapping the interaction sites of *Aspergillus nidulans* phytochrome FphA with the global regulator VeA and the White Collar protein LreB. Mol Genet Genomics 281:35–42. https://doi.org/10.1007/s00438-008-0390-x

Purschwitz J, Müller S, Kastner C, Schöser M, Haas H, Espeso EA, Atoui A, Calvo AM, Fischer R (2008) Functional and Physical Interaction of Blue- and Red-Light Sensors in *Aspergillus nidulans*. Curr Biol 18:255–259. https://doi.org/10.1016/j.cub.2008.01.061

Qian Y, Zhong L, Sun Y, Sun N, Zhang L, Liu W, Qu Y, Zhong Y (2019) Enhancement of Cellulase Production in *Trichoderma reesei* via Disruption of Multiple Protease Genes Identified by Comparative Secretomics. Front Microbiol 10:2784. https://doi.org/10.3389/fmicb.2019.02784

Quan J, Tian J (2011) Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc 6:242–251. https://doi.org/10.1038/nprot.2010.181

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

Rauscher R, Würleitner E, Wacenovsky C, Aro N, Stricker AR, Zeilinger S, Kubicek CP, Penttilä M, Mach RL (2006) Transcriptional regulation of xyn1, encoding xylanase I, in *Hypocrea jecorina*. Eukaryot Cell 5:447–456. https://doi.org/10.1128/EC.5.3.447-456.2006

Rawlings ND, Barrett AJ, Bateman A (2011) Asparagine peptide lyases: A seventh catalytic type of proteolytic enzymes. J Biol Chem 286:38321–38328. https://doi.org/10.1074/jbc.M111.260026

Rawlings ND, Bateman A (2019) Origins of peptidases. Biochimie 166:4–18. https://doi.org/10.1016/j.biochi.2019.07.026

Reilly MC, Qin L, Craig JP, Starr TL, Glass NL (2015) Deletion of homologs of the SREPB pathway results in hyper-production of cellulases in *Neurospora crassa* and *Trichoderma reesei*. Biotechnol Biofuels 8:121. https://doi.org/10.1186/s13068-015-0297-9

Ries L, Belshaw NJ, Ilmén M, Penttilä ME, Alapuranen M, Archer DB (2014) The role of CRE1 in nucleosome positioning within the *cbh1* promoter and coding regions of *Trichoderma reesei*. Appl Microbiol Biotechnol 98:749–762. https://doi.org/10.1007/s00253-013-5354-3

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

Ries LNA, Beattie SR, Espeso EA, Cramer RA, Goldman GH (2016) Diverse regulation of the CreA carbon catabolite repressor in *Aspergillus nidulans*. Genetics 203:335–352. https://doi.org/10.1534/genetics.116.187872

Ronne H (1995) Glucose repression in fungi. Trends Genet 11:12–17. https://doi.org/10.1016/S0168-9525(00)88980-5

Rossi A, Cruz AHS, Santos RS, Silva PM, Silva EM, Mendes NS, Martinez-Rossi NM (2013) Ambient pH sensing in filamentous fungi: Pitfalls in elucidating regulatory hierarchical signaling networks. IUBMB Life 65:930–935. https://doi.org/10.1002/iub.1217

Rulli MC, Bellomi D, Cazzoli A, De Carolis G, D'Odorico P (2016) The water-land-food nexus of first-generation biofuels. Sci Rep 6:1–10. https://doi.org/10.1038/srep22521

Saloheimo A, Aro N, Ilmén M, Penttilä M (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*. J Biol Chem 275:5817–5825. https://doi.org/10.1074/jbc.275.8.5817

Santangelo GM (2006) Glucose Signaling in *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev 70:253–282. https://doi.org/10.1128/mmbr.70.1.253-282.2006

Schafmeier T, Diernfellner ACR (2011) Light input and processing in the circadian clock of *Neurospora*. FEBS Lett 585:1467–1473. https://doi.org/10.1016/j.febslet.2011.03.050

Schalén M, Anyaogu DC, Hoof JB, Workman M (2016) Effect of secretory pathway gene overexpression on secretion of a fluorescent reporter protein in *Aspergillus nidulans*. Fungal Biol Biotechnol 3:3. https://doi.org/10.1186/s40694-016-0021-y

Schäpe P, Kwon MJ, Baumann B, Gutschmann B, Jung S, Lenz S, Nitsche B, Paege N, Schütze T, Cairns TC, Meyer V (2019) Updating genome annotation for the microbial cell factory *Aspergillus niger* using gene co-expression networks. Nucleic Acids Res 47:559–569. https://doi.org/10.1093/nar/gky1183

Schmoll M, Tian C, Sun J, Tisch D, Glass NL (2012) Unravelling the molecular basis for light modulated cellulase gene expression - the role of photoreceptors in *Neurospora crassa*. BMC Genomics 13:127. https://doi.org/10.1186/1471-2164-13-127

Seh ZW, Kibsgaard J, Dickens CF, Chorkendorff I, Nørskov JK, Jaramillo TF (2017) Combining theory and experiment in electrocatalysis: Insights into materials design. Science 355:eaad4998. https://doi.org/10.1126/science.aad4998

Seiboth B, Karimi RA, Phatale PA, Linke R, Hartl L, Sauer DG, Smith KM, Baker SE, Freitag M, Kubicek CP (2012) The putative protein methyltransferase LAE1 controls cellulase gene expression in *Trichoderma reesei*. Mol Microbiol 84:1150–1164. https://doi.org/10.1111/j.1365-2958.2012.08083.x

Sharma A, Sharma A, Singh S, Kuhad RC, Nain L (2019) Thermophilic Fungi and Their Enzymes for Biorefineries. In: Fungi in Extreme Environments: Ecological Role and Biotechnological Significance. Springer International Publishing, pp 479–502.

Shibuya N, Iwasaki T (1985) Structural features of rice bran hemicellulose. Phytochemistry 24:285–289. https://doi.org/10.1016/S0031-9422(00)83538-4

Shroff RA, O'Connor SM, Hynes MJ, Lockington RA, Kelly JM (1997) Null alleles of *creA*, the regulator of carbon catabolite repression in *Aspergillus nidulans*. Fungal Genet Biol 22:28–38. https://doi.org/10.1006/fgbi.1997.0989

Shrotri A, Kobayashi H, Fukuoka A (2017) Catalytic Conversion of Structural Carbohydrates and Lignin to Chemicals. In: Advances in Catalysis. Academic Press Inc., pp 59–123.

Sidrauski C, Walter P (1997) The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. Cell 90:1031–1039. https://doi.org/10.1016/S0092-8674(00)80369-4

Singh B (2014) *Myceliophthora thermophila* syn. *Sporotrichum thermophile*: A thermophilic mould of biotechnological potential. Crit Rev Biotechnol 36:59–69. https://doi.org/10.3109/07388551.2014.923985

Sinha R, Khare SK (2013) Thermostable Proteases. In: Thermophilic Microbes in Environmental and Industrial Biotechnology. Springer Netherlands, pp 859–880.

Smith L (2001) Plant cell division: building walls in the right places. Nat Rev Mol Cell Biol 2: 33–39. https://doi.org/10.1038/35048050

Steffens EK, Becker K, Krevet S, Teichert I, Kück U (2016) Transcription factor PRO1 targets genes encoding conserved components of fungal developmental signaling pathways. Mol Microbiol 102:792–809. https://doi.org/10.1111/mmi.13491

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

Sukharev SA, Pleshakova O V., Sadovnikov VB (1997) Role of proteases in activation of apoptosis. Cell Death Differ 4:457–462. https://doi.org/10.1038/sj.cdd.4400263

Sun J, Glass NL (2011) Identification of the CRE-1 cellulolytic regulon in *Neurospora crassa*. PLoS One 6:e25654. https://doi.org/10.1371/journal.pone.0025654

Sun J, Tian C, Diamond S, Louise Glassa N (2012) Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in *Neurospora crassa*. Eukaryot Cell 11:482–493. https://doi.org/10.1128/EC.05327-11

Sun X, Wang F, Lan N, Liu B, Hu C, Xue W, Zhang Z, Li S (2019) The Zn(II)2Cys6-type transcription factor ADA-6 regulates conidiation, sexual development, and oxidative stress response in *Neurospora crassa*. Front Microbiol 10:750 . https://doi.org/10.3389/fmicb.2019.00750

Sun Y, Cheng J (2002) Hydrolysis of lignocellulosic materials for ethanol production: A review. Bioresour Technol 83:1–11. https://doi.org/10.1016/S0960-8524(01)00212-7

Suzuki K, Tanaka M, Konno Y, Ichikawa T, Ichinose S, Hasegawa-Shiro S, Shintani T, Gomi K (2015) Distinct mechanism of activation of two transcription factors, AmyR and MalR, involved in amylolytic enzyme production in *Aspergillus oryzae*. Appl Microbiol Biotechnol 99:1805–1815. https://doi.org/10.1007/s00253-014-6264-8

Tamayo EN, Villanueva A, Hasper AA, Graaff LH d., Ramón D, Orejas M (2008) CreA mediates repression of the regulatory gene *xlnR* which controls the production of xylanolytic enzymes in *Aspergillus nidulans*. Fungal Genet Biol 45:984–993. https://doi.org/10.1016/j.fgb.2008.03.002

Tang Z, Lu Q, Zhang Y, Zhu X, Guo Q (2009) One step bio-oil upgrading through hydrotreatment, esterification, and cracking. Ind Eng Chem Res 48:6923–6929. https://doi.org/10.1021/ie900108d

Tani S, Kawaguchi T, Kobayashi T (2014a) Complex regulation of hydrolytic enzyme genes for cellulosic biomass degradation in filamentous fungi. Appl Microbiol Biotechnol 98:4829–4837. https://doi.org/10.1007/s00253-014-5707-6

Tani S, Kawamura A, Sumitani J, Kawaguchi T (2014b) Functional analysis of ClbR and ClbR2 controlling the cellulosic biomass degrading enzyme gene expression in *Aspergillus aculeatus*. Asperfest 11, Spain, Sevilla

Thieme N, Wu VW, Dietschmann A, Salamov AA, Wang M, Johnson J, Singan VR, Grigoriev I V., Glass NL, Somerville CR, Benz JP (2017) The transcription factor PDR-1 is a multi-functional regulator and key component of pectin deconstruction and catabolism in *Neurospora crassa*. Biotechnol Biofuels 10:149. https://doi.org/10.1186/s13068-017-0807-z

Thompson-Jaeger S, Francois J, Gaughran JP, Tatchell K (1991) Deletion of SNF1 affects the nutrient response of yeast and resembles mutations which activate the adenylate cyclase pathway. Genetics 129:697–706.

Tian C, Beeson WT, Iavarone AT, Sun J, Marletta MA, Cate JHD, Glass NL (2009) Systems analysis of plant cell wall degradation by the model filamentous fungus *Neurospora crassa*. Proc Natl Acad Sci U S A 106:22157–22162. https://doi.org/10.1073/pnas.0906810106

Tisch D, Schmoll M (2013) Targets of light signalling in *Trichoderma reesei*. BMC Genomics 14:657. https://doi.org/10.1186/1471-2164-14-657

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 15:425. https://doi.org/10.1186/1471-2164-15-425

Todd RB, Zhou M, Ohm RA, Leeggangers HACF, Visser L, de Vries RP (2014) Prevalence of transcription factors in ascomycete and basidiomycete fungi. BMC Genomics 15:1–12. https://doi.org/10.1186/1471-2164-15-214

Tomsett AB, Dunn-Coleman NS, Garrett RH (1981) The regulation of nitrate assimilation in *Neurospora crassa*: The isolation and genetic analysis of nmr-1 mutants. Mol Gen Genet 182:229–233. https://doi.org/10.1007/BF00269662

Topakas E, Moukouli M, Dimarogona M, Christakopoulos P (2012) Expression, characterization and structural modelling of a feruloyl esterase from the thermophilic fungus *Myceliophthora thermophila*. Appl Microbiol Biotechnol 94:399–411. https://doi.org/10.1007/s00253-011-3612-9

Toushik SH, Lee KT, Lee JS, Kim KS (2017) Functional Applications of Lignocellulolytic Enzymes in the Fruit and Vegetable Processing Industries. J Food Sci 82:585–593. https://doi.org/10.1111/1750-3841.13636

Tudzynski B (2014) Nitrogen regulation of fungal secondary metabolism in fungi. Front Microbiol 5:656. https://doi.org/10.3389/fmicb.2014.00656

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

Van Leeuwe TM, Arentshorst M, Ernst T, Alazi E, Punt PJ, Ram AFJ (2019) Efficient marker free CRISPR/Cas9 genome editing for functional analysis of gene families in filamentous fungi. Fungal Biol Biotechnol 6:13. https://doi.org/10.1186/s40694-019-0076-7

Van Oorschot C (1977) The genus *Myceliophthora*. Persoonia 9:401–408.

Van Peij NNME, Gielkens MMC, De Vries RP, Visser J, De Graaff LH (1998a) The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in *Aspergillus niger*. Appl Environ Microbiol 64:3615–3619. https://doi.org/10.1128/aem.64.10.3615-3619.1998

Van Peij NNME, Visser J, De Graaff LH (1998b) Isolation and analysis of *xlnR*, encoding a transcriptional activator co-ordinating xylanolytic expression in *Aspergillus niger*. Mol Microbiol 27:131–142. https://doi.org/10.1046/j.1365-2958.1998.00666.x

Van Ree R (2017) Biorefinery Approach in the EU and Beyond. Workshop on EU-AU R&I Partnership on Food and Nutrition Security and Sustainable Agriculture (FNSSA)-Brussels

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol 153:895–905. https://doi.org/10.1104/pp.110.155119

Varga J, Samson RA (2008) *Aspergillus* in the genomic era. Wageningen Academic Publishers

Verdoes JC, Punt PJ, Burlingame R, Bartels J, Van Dijk R, Slump E, Meens M, Joosten R, Emalfarb MA (2007) A dedicated vector for efficient library construction and high throughput screening in the hyphal fungus *Chrysosporium lucknowense*. Ind Biotechnol 3:48–57. https://doi.org/10.1089/ind.2007.3.048

Viikari L, Alapuranen M, Puranen T, Vehmaanperä J, Siika-Aho M (2007) Thermostable enzymes in lignocellulose hydrolysis. In: Advances in Biochemical Engineering/Biotechnology. Springer, Berlin, Heidelberg, pp 121–145.

Visser H, Joosten V, Punt PJ, Gusakov A V., Olson PT, Joosten R, Bartels J, Visser J, Sinitsyn AP, Emalfarb MA, Verdoes JC, Wery J (2011) Development of a mature fungal technology and production platform for industrial enzymes based on a *Myceliophthora thermophila* isolate, previously known as *Chrysosporium lucknowense* C1. Ind Biotechnol 7:214–223. https://doi.org/10.1089/ind.2011.7.214

Vitikainen M (2018) Development of *Myceliophthora thermophila* into a highly productive biologics production host. PEGS Europe

Volkov P V., Rozhkova AM, Gusakov A V., Zorov IN, Sinitsyn AP (2015) Glucoamylases from *Penicillium verruculosum* and *Myceliophthora thermophila*: Analysis of differences in activity against polymeric substrates based on 3D model structures of the intact enzymes. Biochimie 110:45–51. https://doi.org/10.1016/j.biochi.2014.12.010

Von Klopotek A (1974) Revision der thermophilen Sporotrichum-Arten: *Chrysosporium thermophilum* (Apinis) comb. nov. und *Chrysosporium fergusii* spec. nov. = status conidialis von *Corynascus thermophilus* (Fergus und Sinden) comb. nov. Arch Microbiol 98:365–369. https://doi.org/10.1007/BF00425296

Wang G, Zhang D, Chen S (2014) Effect of earlier unfolded protein response and efficient protein disposal system on cellulase production in Rut C30. World J Microbiol Biotechnol 30:2587–2595. https://doi.org/10.1007/s11274-014-1682-4

Wang H (2009) Filamentous fungi with inactivated protease genes for altered protein production. Patent number: CA2720827A1.

Wang J, Gong Y, Zhao S, Liu G (2018) A new regulator of cellulase and xylanase in the thermophilic fungus *Myceliophthora thermophila* strain ATCC 42464. 3 Biotech 8:160. https://doi.org/10.1007/s13205-017-1069-y

Wang J, Wu Y, Gong Y, Yu S, Liu G (2015) Enhancing xylanase production in the thermophilic fungus *Myceliophthora thermophila* by homologous overexpression of *Mtxyr1*. J Ind Microbiol Biotechnol 42:1233–1241. https://doi.org/10.1007/s10295-015-1628-3

Wanka F, Cairns T, Boecker S, Berens C, Happel A, Zheng X, Sun J, Krappmann S, Meyer V (2016) Tet-on, or Tet-off, that is the question: Advanced conditional gene expression in *Aspergillus*. Fungal Genet Biol 89:72–83. https://doi.org/10.1016/j.fgb.2015.11.003

Ward B (2015) Bacterial Energy Metabolism. In: Molecular Medical Microbiology: Second Edition. Elsevier Ltd, pp 201–233.

Wertz JL, Bédué O (2013) Lignocellulosic biorefineries. EPFL Press

Wu VW, Thieme N, Huberman LB, Dietschmann A, Kowbel DJ, Lee J, Calhoun S, Singan VR, Lipzen A, Xiong Y, Monti R, Blow MJ, O'Malley RC, Grigoriev I V., Benz JP, Glass NL (2020) The regulatory and transcriptional landscape associated with carbon utilization in a filamentous fungus. Proc Natl Acad Sci U S A 117:6003–6013. https://doi.org/10.1073/pnas.1915611117

Würleitner E, Pera L, Wacenovsky C, Cziferszky A, Zeilinger S, Kubicek CP, Mach RL (2003) Transcriptional regulation of *xyn2* in *Hypocrea jecorina*. Eukaryot Cell 2:150–158. https://doi.org/10.1128/EC.2.1.150-158.2003

Xiong Y, Sun J, Glass NL (2014) VIB1, a Link between Glucose Signaling and Carbon Catabolite Repression, Is Essential for Plant Cell Wall Degradation by *Neurospora crassa*. PLoS Genet 10:e1004500. https://doi.org/10.1371/journal.pgen.1004500

Xiong Y, Wu VW, Lubbe A, Qin L, Deng S, Kennedy M, Bauer D, Singan VR, Barry K, Northen TR, Grigoriev I V., Glass NL (2017) A fungal transcription factor essential for starch degradation affects integration of carbon and nitrogen metabolism. PLoS Genet 13:e1006737. https://doi.org/10.1371/journal.pgen.1006737

Xu G, Li J, Liu Q, Sun W, Jiang M, Tian C (2018) Transcriptional analysis of *Myceliophthora thermophila* on soluble starch and role of regulator AmyR on polysaccharide degradation. Bioresour Technol 265:558–562. https://doi.org/10.1016/j.biortech.2018.05.086

Xu J, Li J, Lin L, Liu Q, Sun W, Huang B, Tian C (2015) Development of genetic tools for *Myceliophthora thermophila*. BMC Biotechnol 15:35. https://doi.org/10.1186/s12896-015-0165-5

Yamakawa Y, Endo Y, Li N, Yoshizawa M, Aoyama M, Watanabe A, Kanamaru K, Kato M, Kobayashi T (2013) Regulation of cellulolytic genes by McmA, the SRF-MADS box protein in *Aspergillus nidulans*. Biochem Biophys Res Commun 431:777–782. https://doi.org/10.1016/j.bbrc.2013.01.031

Yang F, Gong Y, Liu G, Zhao S, Wang J (2015) Enhancing cellulase production in thermophilic fungus *Myceliophthora thermophila* ATCC42464 by RNA interference of *cre1* gene expression. J Microbiol Biotechnol 25:1101–1107. https://doi.org/10.4014/jmb.1501.01049

Yao G, Li Z, Gao L, Wu R, Kan Q, Liu G, Qu Y (2015) Redesigning the regulatory pathway to enhance cellulase production in *Penicillium oxalicum* David Wilson. Biotechnol Biofuels 8:71. https://doi.org/10.1186/s13068-015-0253-8

Yoon J, Maruyama JI, Kitamoto K (2011) Disruption of ten protease genes in the filamentous fungus *Aspergillus oryzae* highly improves production of heterologous proteins. Appl Microbiol Biotechnol 89:747–759. https://doi.org/10.1007/s00253-010-2937-0

Yuan GF, Fu YH, Marzluf GA (1991) *nit-4*, a pathway-specific regulatory gene of *Neurospora crassa*, encodes a protein with a putative binuclear zinc DNA-binding domain. Mol Cell Biol 11:5735–5745. https://doi.org/10.1128/mcb.11.11.5735

Yuan XL, Van Der Kaaij RM, Van Den Hondel CAMJJ, Punt PJ, Van Der Maarel MJEC, Dijkhuizen L, Ram AFJ (2008) Aspergillus niger genome-wide analysis reveals a large number of novel alpha-glucan acting enzymes with unexpected expression profiles. Mol Genet Genomics 279:545–561. https://doi.org/10.1007/s00438-008-0332-7

Zeilinger S, Mach RL, Kubicek CP (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. J Biol Chem 273:34463–34471. https://doi.org/10.1074/jbc.273.51.34463

Zeilinger S, Schmoll M, Pail M, Mach RL, Kubicek CP (2003) Nucleosome transactions on the *Hypocrea jecorina* (*Trichoderma reesei*) cellulase promoter *cbh2* associated with cellulase induction. Mol Genet Genomics 270:46–55. https://doi.org/10.1007/s00438-003-0895-2

Zhang F, Bunterngsook B, Li J-X, Zhao X-Q, Champreda V, Liu C-G, Bai F-W (2019) Regulation and production of lignocellulolytic enzymes from *Trichoderma reesei* for biofuels production. In: Advances in Bioenergy. Elsevier, pp 79–119.

Zhang Q, Chang J, Wang T, Xu Y (2007) Review of biomass pyrolysis oil properties and upgrading research. Energy Convers Manag 48:87–92. https://doi.org/10.1016/j.enconman.2006.05.010

Zhang X, Qu Y, Qin Y (2016) Expression and chromatin structures of cellulolytic enzyme gene regulated by heterochromatin protein 1. Biotechnol Biofuels 9:206. https://doi.org/10.1186/s13068-016-0624-9

Zheng Y, Zhao J, Xu F, Li Y (2014) Pretreatment of lignocellulosic biomass for enhanced biogas production. Prog Energy Combust Sci 42:35–53. https://doi.org/10.1016/j.pecs.2014.01.001

10. Acknowledgments

At first, I would like to thank my supervisor Prof. Dr. Vera Meyer and Dr. Stefan Haefner for making this dissertation in cooperation with the BASF SE possible. Here I would like to thank Dr. Stefan Haefner for the general and financial organization of the project within the BASF SE and Prof. Dr. Vera Meyer for the organization and possibility to work on this project at the TU Berlin. Furthermore, I would like to thank both for the valuable discussions and advices within the project and their participance in the doctoral committee. In this context I would also like to thank Prof. Dr. Peter Neubauer as the head of the doctoral committee.

I would also like to express my deep gratitude to Dr. Tabea Schütze for many discussions and essential advices within as well as besides this project. Her steady helpfulness, reliability and extensive knowledge helped me many times. Thank you very much!

Furthermore, I would like to thank all other colleagues from BASF (Dr. Andreas Thywissen, Dr. Sebastian Spohner, Dr. Nico Böhmer, Dr. Katja Gemperlein, Dr. Doug Hodgson, Dr. Paul Costea, Dr. Anna Schroer, Dr. Esther Dantas Costa) and TU Berlin (Dr. Min Jin Kwon), that were participating in this project and their valuable advices and support.

I would also like to thank Rita Waggad for our countless conversations that were a welcome change to the lab work. Furthermore, I want to stress her reliability and steady support during the last 3 years.

Many thanks also go to all my friends, especially Florian and Georg. I would like to thank Florian who is a friend of mine since more than 10 years and visited me in Berlin or hosted me in Heidelberg many times within the last 3 years and kept steady contact with me. I enjoyed our countless evenings within the LPMO squad very much. Many thanks also go to Georg for the countless boardgame evenings, our spontaneous midnight TM game and our friendship that developed within the last 3 years.

My deepest gratitude goes to my family, especially to my parents Donald and Rosie, my grandmother Ursula and my girlfriend Janice and her family. My parents for their constant emotional support, their countless visits in Berlin, which I enjoyed very much, and their support of my academic education. My grandmother Ursula for her emotional support and the numerous beautiful self-painted pictures she sent me within the last 3 years. Finally, I would like to thank my girlfriend Janice for her love and her constant support in all areas of life. I am unbelievably happy that we met each other!

Unfortunately, I cannot name everybody that contributed, directly or indirectly, to the success of this work. Therefore, I would like to thank all not mentioned people that were giving me academic or emotional support over the past 3 years.

Thank you very much to all of you!