

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Transcriptional Regulators of Cellulase Gene Expression in *Neurospora crassa*

Permalink

<https://escholarship.org/uc/item/1n10873x>

Author

Coradetti, Samuel Thomas

Publication Date

2014

Peer reviewed|Thesis/dissertation

**Transcriptional Regulators of Cellulase Gene Expression in
*Neurospora crassa***

By

Samuel Thomas Coradetti

A dissertation submitted in partial satisfaction of the requirements for
the degree of
Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor N. Louise Glass, Chair

Professor Chris Somerville

Professor Jasper Rine

Fall 2014

Copyright © 2014 Samuel Coradetti
All Rights Reserved

Abstract

Transcriptional Regulators of Cellulase Gene Expression in *Neurospora crassa*

by
Samuel Thomas Coradetti

Doctor of Philosophy in Microbiology
University of California, Berkeley
Professor N. Louise Glass, Chair

A substantive transition from an unsustainable fossil material and energy economy to a robust and sustainable biofuels and biomaterials economy will require a fundamental change in the economics of cellulosic biomass deconstruction to sugars with secreted fungal enzymes. To supply these enzymes on the scale required by global biorefinery industry will likely necessitate rational engineering of fungal production strains to produce a wider diversity of enzymes at higher yield and on more diverse substrates. This will, in turn require more complete knowledge of how expression of fungal enzymes, particularly cellulases, are regulated. To further elucidate general features of fungal enzyme regulation and identify new regulatory proteins affecting enzyme expression, I carried out a research program on transcriptional regulation of cellulases in the model filamentous fungus *Neurospora crassa* with comparative experiments in the evolutionarily divergent species *Aspergillus nidulans*.

A screen of predicted transcription factor mutants from the *N. crassa* deletion collection identified several genes with previously unknown effects on cellulase expression on cellulose. When *N. crassa* is transferred to inducing substrates such as cellulose, approximately 200 genes are transcriptionally induced, including several dozen genes that encode known carbohydrate active enzymes (CAZy genes) with expression increased as much as several thousand fold over non-inducing starvation conditions. This transcriptional induction was compromised in several transcription factor deletion mutants. Two deletion mutants, for the genes that I named cellulose degradation regulator 1 and cellulose degradation regulator 2 (CLR-1 and CLR-2), and entirely lost their transcriptional response to cellulose.

Functional CLR-1 was required for efficient cellobiose utilization and cellulase secretion. Though CLR-1 transcript abundance increases on cellulose, constitutive expression of CLR-1 at or above levels observed in inducing conditions was insufficient to stimulate cellulase secretion in non-inducing conditions, suggesting a requirement for post-translational activation of CLR-1. Functional CLR-2 is required for cellulose utilization but not cellobiose utilization. CLR-2 expression also increases dramatically on cellulose and constitutive expression of CLR-2 at comparably high levels was sufficient for cellulase secretion in non-inducing conditions. However, not all β -glucosidases and cellobiose transporters were expressed in CLR-2 constitutive expression strains. CLR-2 expression was

CLR-1 dependent. I proposed a regulatory model in which a cellobiose sensing activates CLR-1, which in turn induces gene expression of CLR-2, the primary direct regulator of cellulase expression.

Analysis of these transcription factors in *A. nidulans* revealed conserved and divergent aspects, with a drastically reduced role for the CLR-1 ortholog ClrA, which regulates only a single major cellulase. The CLR-2 ortholog ClrB was not dependent on ClrA and was not only required for cellulose utilization but for cellobiose sensing and utilization as well. Constitutive expression of ClrB was not sufficient for constitutive cellulase secretion. Cellulases were strongly induced when WT *A. nidulans* was exposed to cellobiose, but not when it was exposed to mannobiose. When ClrB constitutive/over-expression strains were exposed to mannobiose, major cellulases were strongly induced.

Global transcriptional comparisons of WT and mutant strains of *N. crassa* and *A. nidulans* also provided new insights into shared and divergent strategies for cellulose degradation, particularly by the lytic polysaccharide monooxygenase (LPMO) enzyme class. Differing expression profiles between LPMO subclasses between *N. crassa* and *A. nidulans* may be reflective of still unappreciated differences in enzymatic mechanism or preferred substrate between LPMO classes.

To address emerging difficulties in correlating sequence similarity and functional conservation for fungal transcription factors, both in the above studies and contemporary studies on other regulators of plant cell wall degrading enzymes, I initiated a search for additional components of the cellobiose sensing system. I constructed an antibiotic resistance construct under regulation of CLR-1 and CLR-2, then selected for mutants with antibiotic resistance in carefully selected non-inducing, non-repressing conditions, isolating several mutants of interest. Bulk segregant analysis on the first of these mutants identified a new repressor of CLR-1 activation (NCU05846). Further characterization of NCU05846 and additional components of the cellobiose-sensing network may further enhance our ability to manipulate cellulase gene expression in *N. crassa*. That knowledge may also enrich our understanding of fungal nutrient sensing, and enable more nuanced and predictive analysis of how such systems are conserved across fungal species in general and highly productive industrial strains in particular.

Table of Contents

Abstract.....	1
List of Tables	i
List of Figures.....	ii
Acknowledgements.....	iv
Funding	v
Chapter 1 : Introduction.....	1
Chapter 2 : Carbon Source, Growth and Regulated Transcription of Cellulases in <i>Neurospora crassa</i>	14
Chapter 3 : Screening the <i>N. crassa</i> Deletion Collection for Novel Transcription Factors Regulating Cellulose Utilization.....	27
Chapter 4 Cellulose Degradation Regulators CLR-1 and CLR-2	37
Chapter 5 Conservation of CLR-1 and CLR-2 Function in <i>Aspergillus nidulans</i>	53
Chapter 6 A Selective Screen for Genes Involved in Cellobiose Sensing.....	80
Chapter 7 NCU05846: a Repressor of CLR-1 Activation.....	90
Chapter 8 Discussion and Conclusions.....	103
References	107
Appendix A: The <i>N. crassa</i> Avicel Regulon	121
Appendix B: RT-PCR primers.....	125
Appendix C: Plasmid Sequences	126

List of Tables

Table 1. Regulation of plant cell wall degrading enzymes by XlnR orthologs.....	11
Table 2. Phenotypes of transcription factor deletion strains selected for secondary screening.....	30
Table 3. Conserved orthologous genes in the CLR-2 and ClrB regulons.....	67
Table 4. Bulk segregation of mutations identified in strain M13.....	86
Table 5: CLR-2 Responsive genes involved in protein secretion	99

List of Figures

Figure 1. Major plant cell wall polysaccharides and their constituent sugars	5
Figure 2. Phylogenetic tree of ascomycete fungi.....	7
Figure 3. Zinc Binuclear Cluster Transcription Factor Domain Architecture.....	10
Figure 4. Endoglucanase Activity in <i>N. crassa</i> supernatants after shift to Avicel plus repressing carbon source.	17
Figure 5. Growth curves on cellulase repressing carbon sources.	18
Figure 6. Transcriptional induction by cellulosic substrates in <i>Neurospora crassa</i> as measured by RNA-Seq.....	19
Figure 7. Differentially expressed <i>N. crassa</i> genes after shift to either Avicel versus sucrose cultures or Avicel versus no carbon cultures.....	21
Figure 8. Heat map of expression levels in wild type on sucrose, no carbon and Avicel for 333 genes differentially expressed on Avicel and no carbon 4 hrs after shift.....	22
Figure 9. Visual inspection of transcription factor deletion mutants growing on Avicel, sucrose and xylan.	29
Figure 10. Heat map of hierarchical clusters of genes in the <i>N. crassa</i> Avicel regulon from microarray transcriptional profiles of transcription factor deletion strains switched to Avicel.	32
Figure 11. Growth and activity phenotype of <i>clr</i> deletion strains.	38
Figure 12. Secreted proteins in $\Delta clr-1$ and $\Delta clr-2$ culture supernatants.....	39
Figure 13. Cellulase transcription as a function of <i>clr-1</i> and <i>clr-2</i> transcription.....	40
Figure 14. Enzyme activity and protein secretion in the <i>Pccg-1:clr-2</i> mutant	41
Figure 15. Transcript abundance for major cellulases and hemicellulases in <i>clr-1</i> and <i>clr-2</i> deletion mutants	42
Figure 16. The <i>clr</i> regulon.	43
Figure 17. Transcript abundance for cellobiose and cellobionic acid utilization genes	45
Figure 18. Cellulase induction by cellobiose in $\Delta clr-1$ and $\Delta clr-2$ mutants.....	46
Figure 19. Transcript abundance of <i>clr-1</i> and <i>clr-2</i> after a shift to Avicel or no carbon.	47
Figure 20. Model for regulation of cellulose and xylan degradation in <i>N. crassa</i>	49
Figure 21. Phylogenetic tree of CLR-1 homologs in filamentous fungi.	55
Figure 22. Phylogenetic tree of CLR-2 homologs in filamentous fungi.	56
Figure 23. Phenotype of <i>Aspergillus nidulans</i> <i>clr</i> deletion strains.	57
Figure 24. The Avicel regulons of <i>N. crassa</i> and <i>A. nidulans</i>	59
Figure 25. Time course of enzyme activity from <i>N. crassa</i> and <i>A. nidulans</i> culture supernatants.....	62
Figure 26. Hierarchical clustering of orthologous and paralogous genes induced by exposure to Avicel in <i>N. crassa</i> and <i>A. nidulans</i>	63
Figure 27. Dependence of Avicel-induced genes on functional ClrB in <i>A. nidulans</i>	66
Figure 28. Differential expression in <i>clrA</i> and <i>clrB</i> deletion strains.....	69

Figure 29. Growth phenotypes of <i>A. nidulans</i> <i>clrB</i> and <i>clr-2</i> misexpression mutants.	70
Figure 30. Cellulase induction in <i>A. nidulans</i> <i>clrB</i> and <i>clr-2</i> misexpression mutants.	71
Figure 31. Timecourse of cellulase induction in <i>A. nidulans</i> FGSC A4 shifted to 1mM Cellobiose or 1mM Mannobiose.	72
Figure 32. Induction of <i>A. nidulans</i> cellulases in response to 1,4- β -D-cellobiose and 1,4- β -D-mannobiose.	73
Figure 33. Expression profiles of candidate genes for selectable markers	81
Figure 34. Selectable construct for cellulase induction.	82
Figure 35. Growth of the cellulose inducible selection strain on CMC/xylan hygromycin media	83
Figure 36. Hygromycin resistant colonies from NTG exposed conidia growing on 0.5% xylan with 175 μ g/mL hygromycin.	84
Figure 37. Cellulase activity of xylan cultures from mutant strains. Endoglucanase activity in xylan grown culture supernatants of NTG mutant hygromycin resistant strains from parent SC3G4.	84
Figure 38. Segregation of hygromycin resistance and constitutive cellulase activity in M13 cross progeny.	85
Figure 39. The M13 mutation is a non-synonymous mutation in a conserved domain of unknown function.	91
Figure 40. Phylogenetic Tree of NCU05846 and Homologs.	92
Figure 41. Endoglucanase activity in supernatants from NCU05846 deletion mutants 48 hours after a shift to inducing and noninducing conditions.	93
Figure 42. Endoglucanase activity on xylan for NCU02600 deletion mutants.	94
Figure 43. Constitutive cellulase activity in the Δ NCU05846 mutant requires <i>clr-1</i>	95
Figure 44. Induction of CLR-2 independent genes in the Δ NCU05846 mutant.	96
Figure 45. Expression of cellulase and cellobiose utilization genes in the Δ NCU05846 mutant.	97
Figure 46. Transcription of Avicel induced genes in the Δ NCU05846 mutant on no carbon media.	98
Figure 47. Alternative models for the <i>N. crassa</i> cellobiose-sensing network.	100
Figure 48. Relative expression and phylogenetic distribution of LPMOs (AA9 CAZy class) by subclass.	105

Acknowledgements

N. Louise Glass has made this body of work possible through her excellent mentorship and consistent support. Louise provided me with all the tools I required to conduct my experiments, secured the lion's share of my funding and gave me the respect and freedom to follow my ideas where they led. Perhaps most importantly, she created a vibrant, collaborative, laboratory filled with skilled and generous colleagues. In this way she helped mold an extraordinarily pleasant and productive environment that has greatly enhanced my scientific progress and general quality of life during my tenure at Berkeley.

The many direct experimental contributions of my labmates to this work are highlighted at the end of each chapter, but I would be remiss not to take this opportunity to offer special thanks to the senior lab members who patiently coached my education in everything *Neurospora*. **Abigail Leeder**, **Anna Simonin** and **Betsy Hutchison** taught me basic techniques to culture and manipulate *N. crassa* when I joined the lab, and later provided sage advice and counsel as I developed my project and advanced to candidacy. **Jianping Sun** and **Trevor Starr** introduced me to many fundamental aspects of plant cell wall degradation by fungi in general and taught me the laboratory's accumulated techniques of cell culture and analysis specific to the cellulose deconstruction project. **Joanna Bueche** and **Trevor Starr** coached me on techniques for protein extraction and western blots.

The Energy Biosciences Institute has been an extraordinarily productive place to conduct my research. In particular, a synergistic relationship with the laboratory of Professor **Jamie Cate** has contextualized my work in ways that would not be possible without our close interaction and communication. Work by **Will Beeson**, **Chris Phillips**, **John Galazka** and **Xin Li** in characterizing the function of several proteins involved in cellulose utilization have given me an unusually detailed perspective and context from which to interpret my own results.

On a practical level, the tireless support of the staff at the EBI has freed all of us to concentrate on our experiments instead of logistics. **Mara Bryan** maintained an impressive array of equipment and solved any problem I threw her way. **Crystal Chan** kept the reagents and supplies flowing like a river. **Marta Ortega** and **Ofelia Gainza** held back a tide of dirty glassware from over a dozen labs. **Adam Cohen** kept the electrons flowing. **Zack Phillips** kept the lights on, the water in the pipes, the ceiling above us, and the researchers not on fire.

Finally, as my bay-mates, **Abigail Leeder**, **Javier Palma-Guerrero**, **James Craig**, **Christy Roche**, **Morgann Reilly** and **Lori Huberman** have been my fellow travelers and captive audience during my time at Berkeley. They have offered helpful advice, celebrated my successes, lamented my failures, and most of all, endured my endless rants and bizarre ideas for 5 long years.

"These rare souls; they are the salt of the earth."
--Henry James.

Funding

This work was supported by:

The Energy Biosciences Institute

Through the project grant

" Systems biology analyses of plant cell wall deconstruction by the model filamentous fungus, *Neurospora crassa*"
to N. Louise Glass

and through the general infrastructure at the Energy Biosciences Institute

and by

The National Institutes of Health

Through an NRSA Trainee Grant #2 T32 GM 7127-36 A1
to Samuel Coradetti

and through NIH S10 Instrumentation Grants S10RR029668 and S10RR027303
to the Vincent J. Coates Genomics Sequencing Laboratory
at UC Berkeley
which processed many key data sets for this work

and through the NIH Program Project Grant P01GM068087,
"Functional analysis of a model filamentous fungus"
to the Fungal Genetics Stock Center
which maintains the *Neurospora crassa* deletion collection.

and by

The People of California

Through the general facilities of the University of California
whom I urge to continue their support for education, science and culture
to enrich our state, our country, and our world.

"Don't *tell* me about your values. Show me your budget."
-- Joe Biden

Chapter 1 : Introduction

Biofuels, biorefinery and a carbon-neutral material economy

Billions of barrel-equivalents of petroleum products are consumed annually (seven billion in the United States alone) mostly as liquid fuels, but also as feedstocks for chemical applications and synthetic polymers [1]. The deleterious effects of burning all those hydrocarbons on the global biosphere, as well as their human and economic costs are well known [2]. However, the political will to alter personal lifestyles and the global economy to mitigate carbon emissions has not materialized. In fact, as of this writing, global petroleum production — particularly North American production — continues to rise and the prices of petroleum products continue to fall. Today the adage that "the stone age didn't end for lack of stones and the oil age will end, but not for lack of oil" [3] seems more appropriate than ever. The only way that anthropogenic climate change will be blunted is through the development of sustainable energy sources that can compete in the open marketplace.

Cellulosic "second generation" biofuels have been long been touted as an attractive alternative to fossil fuels, with abundant potential resources for relatively cheap and low impact biomass feedstocks for conversion to fuels and other products [4]. However, cellulosic biofuels remain a vanishingly small component of the global energy economy. As production of "first generation" biofuels made from high starch feedstocks such as corn continues to grow (constituting approximately 6% of US transportation fuels in 2012 [5]), it is clear that neither the demand for biofuels, nor the infrastructure for fermentation or downstream processing are limiting factors for cellulosic ethanol. A significant impediment to wider adoption is the challenge of processing plant cell wall polysaccharides, the largest reservoir of terrestrial carbohydrates. For use by biofuel production organisms, these materials must be depolymerized into their constituent sugars; that process is surprisingly challenging and expensive.

Saprotrophic fungal species are major agents of plant decay in the environment [6,7], with influences on the global carbon cycle [8], local nutrient flows [9] and the historic accumulation of fossil fuel deposits [10]. With that understanding, the application of fungal enzymes in controlled processes was a natural approach for plant polysaccharide depolymerization as development of modern biofuel production began in the mid-late twentieth century. However, despite decades of research and development, production of these enzymes, particularly cellulases, remains prohibitively expensive for cost-effective biofuel production from plant cell wall polysaccharides [11,12]. To bring those costs down will likely require rational engineering of fungal strains, for which we will need a deeper understanding of the fungal biology of protein secretion and of the regulation of enzyme expression at the transcriptional and translational levels. The transcriptional regulation of secreted enzymes, particularly cellulases, in the model filamentous fungus *Neurospora crassa* is the subject of this dissertation.

Plant cell wall polysaccharides and fungal enzymes

Plant cell walls are complex, structured assemblages of polysaccharides [13] with wide variation among plant species and even between specific tissues and growth stages within the same species [14,15]. Those complexities are beyond the scope of this work. I will make reference to a simplified model in which we consider plant cell walls as consisting of 5 major polysaccharides: cellulose, xylan, xyloglucan, mannan and pectin.

The basic structure and composition of each of these plant cell wall polysaccharides are summarized in Figure 1, along with the plant cell wall degrading enzymes that fungi use to degrade them. The enzymatic reactions are denoted by their carbohydrate active enzyme (CAZy) classes, a standardized and actively maintained catalog of enzymes with known or predicted activities on carbohydrates grouped by inferred homology and predicted structural folds [16]. CAZy classes are subdivided into glycoside hydrolases (GH), carbohydrate esterases (CE), glycosyl transferases (GT), polysaccharide lyases (PL) and auxiliary activities (AA). Many of these CAZy classes are reviewed more thoroughly in [17-21].

Cellulose

Cellulose is the most abundant and most recalcitrant plant cell wall polysaccharide. Cellulose consists solely of β -1,4-D-glucose homopolymers arranged in microfibrils of parallel polymers stabilized with extensive hydrogen bonding. This hydrogen bonding creates dense structures, which can be either crystalline or amorphous [22], that are particularly resistant to depolymerization. This simplicity of composition allows for depolymerization with just a few classes of enzymes. Endoglucanases (GH5.5) and lytic polysaccharide monooxygenases (LPMO, class AA9) create breaks in the cellulose chains. Cellobiohydrolases then processively cleave cellobiose units from the reducing end (GH7) or non-reducing end (GH6) of the cleaved chains. Finally, β -glucosidases hydrolyse cellobiose to glucose. For the purposes of this dissertation, the term cellulases generally refers to LPMO, GH6, GH7 and GH5.5 enzymes.

The inherent stability of tightly packed, extensively hydrogen bonded structures renders cellulose the most recalcitrant plant polysaccharides by far. This recalcitrance is reflected by the complete dominance of cellulases measured in quantitative proteomic studies of fungal secreted enzymes on complex plant cell wall material [23,24]. As the limiting enzymes in industrial enzyme mixtures, cellulases have thus received the most attention of all plant cell wall degrading enzymes in terms of protein engineering [25], heterologous expression [26] and regulation [27]. Cellulase regulation is the main focus of this work as well. However, pure cellulose never exists alone in nature, but is always arranged in complex, cross-linked structures with other major polysaccharides. Thus, efficient cellulose depolymerization requires simultaneous depolymerization of many other polysaccharides. The consistent association of these polysaccharides implies that detection of cellulose may serve as a reliable proxy for detection for other plant cell wall components or vice versa.

Mannans

Mannans, together with xyloglucan and xylan are collectively termed hemicellulose: a class distinguished by backbones with β -1,4 linkages, but of more complex composition than cellulose. Mannan is a polymer backbone of β -1,4-D-mannose monomers. Mannans can also exist with a lower proportion of β -1,4-D-glucose substituting for mannose in a non-repeating pattern (glucomannans). Both mannan and glucomannan are often found with α -1,6-galactose side chains (galactomannan) and are often acetylated. Mannans are a relatively low abundance polymer in most potential biorefinery feedstocks [28], but close association with cellulose and the structural similarity of mannobiose to cellobiose may have implications for fungal discrimination of cellulose and mannan, as described in Chapter 5. Endomannanases show high similarity to endoglucanases and are included in the same CAZy family, GH5, subclass 7.

Xyloglucan

Xyloglucans are relatively abundant hemicellulose components [28] with a backbone of β -1,4-D-glucose and variable side chains including α -1,6-D-xylose, β -1,2-D-galactose, and α -1,2-D-fucose [29]. Xyloglucan is often crosslinked to cellulose microfibrils [29] and pectin [30]. By breaking up the crosslinked complexes of polysaccharides, xyloglucanases (GH74 and GH12) may synergistically enhance cellulase activity on complex plant cell wall substrates [31].

Xylan

Xylans are the most abundant hemicellulose components in grasses [28]. As such, deconstruction of this polymer will be critical for effective utilization of proposed energy crops such as switchgrass and *Miscanthus* [32]. Xylans are comprised of a backbone of β -1,4-D-xylose with acetylated and feruloylated side chains of β -1,2-D-xylose and L-arabinose. The feruloyl groups in turn link the xylan into the lignin network [33]. Xylans require several enzyme classes to fully degrade, but endoxylanases (GH10 and GH11) perform the critical hydrolysis of the polymer backbone and receive the most attention in this dissertation.

Pectin

Pectin is a broad classification for a diverse group of complex polymers. The common element in pectins is a backbone of α -1,4-galacturonic acid (galacturonans), or in the case of rhamnogalacturonans, alternating α -1,4-D-galacturonic acid with α -1,2-L-rhamnose [34]. Galacturonans and rhamnogalacturonans can have diverse side chains and like many hemicellulose polymers, they are also extensively acetylated and methylated. Different types of pectin, each with its own complex structure, can alternate in large blocks on a single contiguous polymer backbone. Shown in Figure 1 are some of the most common pectins [34], homogalacturonan and rhamnogalacturonan I, which has side chains of β -1,4-D-galactose and β -1,5-L-arabinose. The complexity of pectin structure and composition is reflected in the large number of enzymes required to completely deconstruct it.

Lignin

Lignin is not comprised of polysaccharides and not directly addressed in this work, but due to its importance in plant cell walls, it merits some discussion here. Lignin is a complex network of covalently linked aromatic molecules deeply imbedded in and crosslinked to the polysaccharide network in plant cell walls, mostly associated with hemicellulose in secondary cell walls [35]. Due to its composition, the chemistry of lignin is complex and several processes for its conversion into fuels or other chemicals are under investigation [36]. For any process including the fermentation of released sugars from plant polysaccharides, lignin is a major complication. Lignin reduces the efficiency of cellulases [37] and its breakdown products are toxic to microbial fuel production strains [38]. Efficient separation and processing of lignin will be critical for the commercialization of cellulosic biofuels.

The fungi studied in this work are not thought to degrade significant amounts of lignin. The regulation of the few putative lignin active enzymes they do express will be considered in the context of plant polysaccharide depolymerization. For example, ferolulyl esterases (CE1) are required to liberate crosslinked xylan from the lignin network.

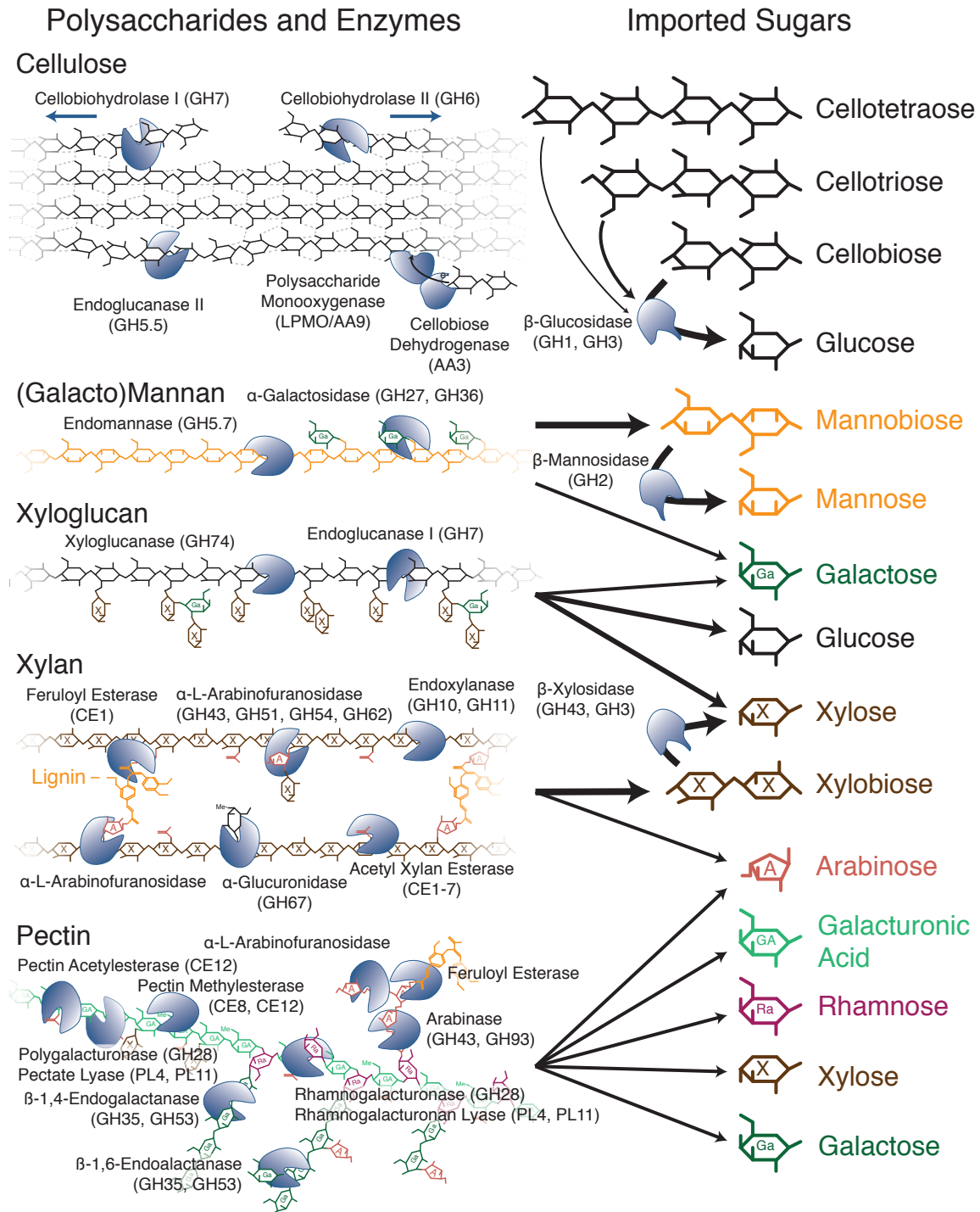


Figure 1. Major plant cell wall polysaccharides and their constituent sugars
 Major fungal carbohydrate active enzymes (CAZy) classes referenced in this work and the linkages they hydrolyze are shown for each polymer. Adapted from figures in [19,20,39,40].

***Neurospora crassa* and filamentous ascomycete fungi.**

N. crassa is the model filamentous fungus most prominently employed in this work. *N. crassa* is a saprophyte thought to primarily colonize freshly burned plant material [41,42] but which may also grow as a plant endophyte in some cases [43]. Before I began these experiments, recent studies of *N. crassa* had demonstrated robust growth on cellulosic material enabled by a diverse hydrolytic secretome [44,45]. An initial study on utilization of cellulose by *N. crassa* yielded promising candidate genes that were subsequently utilized in engineering strategies for improved cellulosic biofuel production in *Saccharomyces cerevisiae* [46] and contributed insights into a role for polysaccharide monooxygenases in cellulose deconstruction [47,48]. That progress continued in the years that followed with characterization of additional secreted enzymes [49], transporters [50-53] and other genes with novel function in cellulose utilization [54]. With these advances, *N. crassa*, already a well established model fungus [55], has become one of the best studied organisms for plant cell wall polysaccharide utilization and an excellent system in which to study cellulase gene regulation. However, *N. crassa* is not currently used in industrial enzyme production. For any findings in *N. crassa* to have significant impact, they must inform our understanding of other systems. Throughout this dissertation I make reference to other filamentous ascomycete species with established or proposed industrial applications or in which substantive work on cellulase regulation has been reported. The phylogenetic relationships among these species are shown in Figure 2.

Aspergillus nidulans

Aspergillus nidulans is another well-developed model filamentous fungus, with a rich body of literature and many established protocols for genetic manipulation [56]. *A. nidulans* and *N. crassa* are highly divergent, with an estimated 300 million years since their speciation [57], and thus make for excellent points of comparison for highly conserved genes. *A. nidulans* is a cosmopolitan species mostly found in soil and plant litter [58]. The *A. nidulans* genome encodes many CAZy genes suggesting substantive plant cell wall polysaccharide degradation capability, though exploration of that ability has been relatively limited until recently.

Trichoderma reesei

Trichoderma reesei, also referred to by the name of its teleomorph *Hypocrea jecornia* [59], has been the species most often employed for industrial cellulase production to date. From parent strains isolated in the South Pacific during the Second World War [60], several lines of mutant strains with elevated cellulase expression and secretion have been isolated through multiple rounds of random mutagenesis and laborious screening over an iterative, decades long process [61-63]. The increased enzyme secretion in those mutant lines has been impressive, but many of the mechanisms underlying those results remain unknown.

Eurotiomycetes

Outgroups

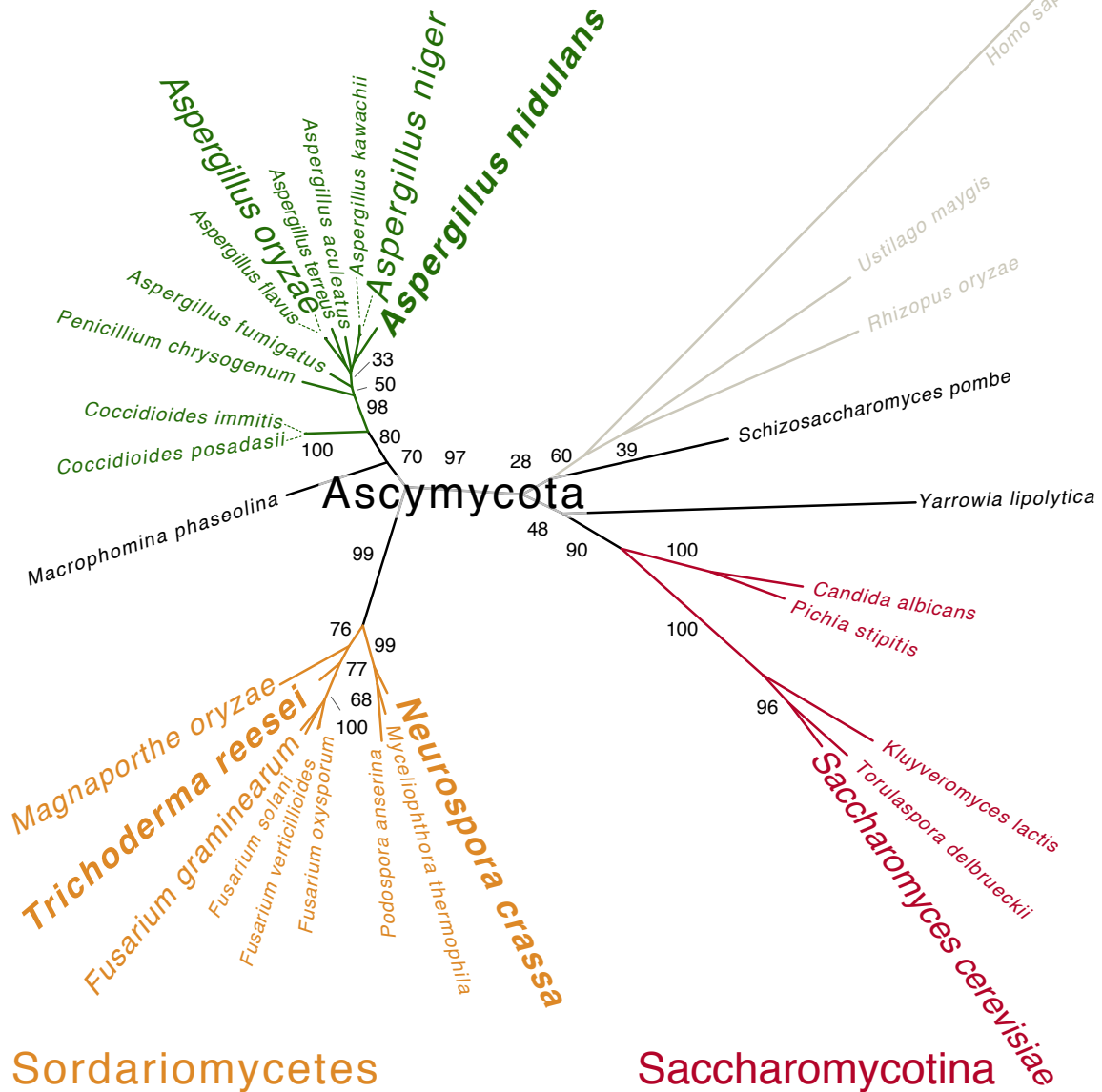


Figure 2. Phylogenetic tree of ascomycete fungi.

Species of particular note for this work in bold. Emphasis on the Sordariomycetes and Eurotiomycetes classes in yellow and green respectively. Several common model ascomycete yeasts in the subphylum Saccharomycotina shown in red. ML tree computed with DNA licensing factor *mcm-7*. Branches labeled with bootstrap support from 100 random samples.

Aspergillus niger

Although primarily used for citric acid production[64], *Aspergillus niger* has also been used for industrial enzyme production[65]. It has also been a significant species for study of regulatory factors for plant cell wall degradation, as detailed below.

Aspergillus oryzae

Aspergillus oryzae is used widely for fermentation of soy products. As an organism "generally recognized as safe" by the Food and Drug Administration [66] it is also an attractive species for future industrial applications.

Fusarium graminearum* and *Magnaporthea oryzae

Fusarium graminearum and *Magnaporthea oryzae* are major pathogens of wheat [67] and rice [68], respectively. Through destruction of critical food crops, these two filamentous ascomycetes represent a serious threat to human health and prosperity on a global scale [69]. Although my primary interest in enzyme regulation relates to engineering production strains, plant cell wall degrading enzymes also play a significant role in fungal pathogenesis of plants [70]. Any improved understanding of their regulation may have implications for improved crop protection and thus major economic and humanitarian implications.

Regulation of cellulases and other secreted enzymes in filamentous ascomycete fungi.

From the first experiments on fungal cellulases in *Trichoderma* species over 50 years ago, it became clear that cellulase expression was under strict regulation by environmental signals, particularly cellulosic materials but also some soluble sugars including cellobiose, lactose, galactose and sophorose [71-76]. It was later established with work in *T. reesei* that this regulation occurs primarily at the transcriptional level [77]. From just a few enzymes in early studies [73], an ever expanding list of enzymes was studied for regulated expression under different carbon sources [78] as the complimentary complexity of the plant cell wall and fungal secretomes were increasingly appreciated.

In 1998, the zinc binuclear cluster (ZnBNC) transcription factor XlnR was found to be essential for induction of xylanases and other hemicellulases in response to xylan and xylose in *A. niger* [79]. Soon after, XlnR was found to regulate cellulase gene expression as well, also in response to xylan and xylose [80,81] and to be modulated by the carbon catabolite repressor CreA [82]. In this period it was also found that orthologs of *S. cerevisiae* heme activator protein (HAP2/3/5 complex) bound *T. reesei* cellulase promoters [83] and that alteration of that binding compromised cellulase gene expression [84]. In the wake of these discoveries in 2001, ACE1 (a zinc finger transcription factor) ACE2 (another ZnBNC) were found to modulate cellulase gene expression in *T. reesei* [85,86]. When it was demonstrated that Xyr1, the *T. reesei* ortholog of XlnR was required for cellulase and xylanase induction by all known inducing molecules in 2006 [87] it seemed likely that

XlnR/Xyr1 was the universal master regulator of most plant cell wall polysaccharide degrading enzymes in ascomycete fungi, with other regulators playing a secondary role in modulating induction (ACE2) or repressing induction on preferred carbon sources (CreA). However, in 2007, it was demonstrated that Xyr1 was dispensable for cellulase induction, but still required for xylanase induction in *F. graminearum* [88]. My lab mate Jianping Sun then found the XlnR ortholog was also dispensable for cellulase induction in *N. crassa* [89].

These results together showed the criticality of transcriptional regulation for secreted cellulases and other plant cell wall polysaccharide degrading enzymes, but also that a number of unknown regulators likely existed. Given that XlnR/Xyr1 function was conserved between the relatively divergent *A. niger* and *T. reesei*, it was surprising that orthologs had less conserved function in *F. graminearum* and *N. crassa*, both of which are more closely related to *T. reesei* than *A. niger*. This discrepancy likely helped to motivate several investigators to embark on searches for additional regulators, myself and several members of my laboratory included. That search is the topic of this dissertation, and discussion thereof will begin in Chapter 2.

First, by way of context, I will conclude this introduction with a brief summary of several other fungal transcriptional regulators of secreted plant cell wall degrading enzymes that have been discovered or further studied over the same period. The most notable feature of this collection of transcription factors is the commonality of the ZnBNC superfamily, typified by a conserved DNA binding domain and a conserved middle homology domain (Figure 3). This large and diverse superfamily (reviewed in [90]) was already known to include some regulators of alternative metabolism (most notably *S. cerevisiae* Gal4, activator of galactose utilization genes in response to galactose [91]), and to be dominated by regulators of stress responses or metabolism. These additional discoveries of ZnBNC regulators plant cell wall depolymerization fit well within that general theme, but would seem to indicate a preponderance of this superfamily in regulation of alternative sugar assimilation in fungi.

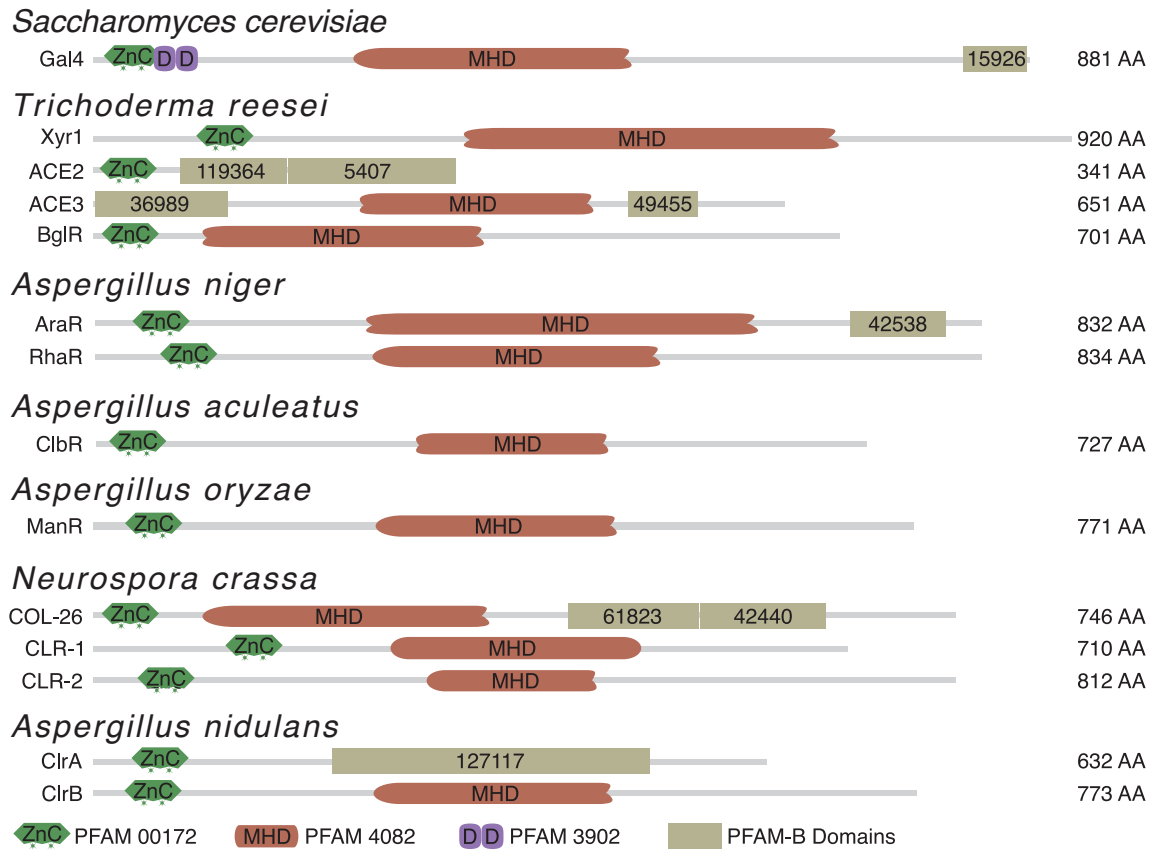


Figure 3. Zinc Binuclear Cluster Transcription Factor Domain Architecture

Fungal transcription factors with reported roles in the regulation of secreted fungal enzymes required for plant cell wall degradation. *S. cerevisiae* Gal4, the prototypical alternative sugar response regulator, is shown for reference, though neither it nor its orthologs have been shown to be critical for plant cell wall degradation.

ZnBNC transcription factors

XlnR/Xyr1/Xlr-1

This transcription factor has some conserved function across divergent species, but with polyphyletic variation in which genes are regulated and what inducing molecules are required. A brief summary of XlnR-dependent regulation of plant cell wall degrading enzymes across filamentous ascomycetes studied to date is shown in Table 1. In general regulation of β -xylosidases and pentose utilization were found to be regulated by XlnR across all species investigated, but regulation of plant cell wall degradation was variable. The polyphyletic pattern of these results makes their generalization to other species within the same evolutionary scope all but impossible without new experiments conducted on the species of interest.

Table 1. Regulation of plant cell wall degrading enzymes by XlnR orthologs

	CBH	β -GLUC	β -XYL	ABF	AGL	β -GAL
<i>N. crassa</i>	N	N	Y	Y	N	N
<i>M. oryzae</i>	N	N	Y	Y	Y	Y
<i>F. graminearum</i>	N	N	Y	N	Y	N
<i>T. reesei</i>	Y	Y	Y	N	N	Y
<i>A. niger</i>	Y	Y	Y	N	Y	Y
<i>A. nidulans</i>	N	N	Y	Y	N	Y
<i>A. aculeatus</i>	N	N	Y	ND	ND	ND

CBH = cellobiohydrolase, β -GLUC = β -glucosidase, β -XYL = β -xylosidase, ABF = arabinofuranosidase, AGL = α -galactosidase, β -GAL = β -galactosidase. Data summarized from [89,92-94]

Overexpression of XlnR increased gene expression of some enzymes in *A. niger* [80], *A. oryzae* [95], and *T. reesei* [96]. Such overexpression only slightly increased enzyme activity in *N. crassa* [89], and did not increase xylanase activity in *Fusarium oxysporum* [97]. In all these studies, constitutive or overexpression of XlnR is not sufficient for enzyme expression in non-inducing conditions, suggesting a post-translational modification or protein-protein interaction is required for XlnR mediated gene expression.

In *A. oryzae*, XlnR was reversibly phosphorylated in the presence of xylose [98]. The investigators took this to suggest that phosphorylation state dictates XlnR's competency to induce expression of target genes, but this causality is not yet proven. Given that phosphorylation of *S. cerevisiae* Gal4 has been shown to be a dispensable consequence of its activation as opposed to the activation mechanism [99], absent further evidence, the role of phosphorylation in the activation of XlnR remains unclear.

Mutational analysis of XlnR in *A. niger* produced a line of mutants with altered nuclear import [100] and nuclear accumulation Xyr1 correlated with cellulase induction in *T. reesei* [101]. However, nuclear accumulation of XlnR was not sufficient for enzyme expression in *A. oryzae* [95] and *N. crassa* [102].

A mutant strain of *T. reesei* with a constitutively active variant of Xyr1 was recently isolated [63]. The causal mutation is a single, conservative (alanine to valine) amino acid substitution near the C-terminus. How this mutation renders Xyr1 constitutively active is not known, but the homologous mutation rendered XLR-1 constitutively active in *N. crassa* [102] and a similarly activating point mutation was identified previously in the C-terminus of XlnR in *A. niger* [100]. The mutated region is not in any known protein domain or at the site of any known protein modification, but is likely involved in activation through an unknown mechanism. XlnR has been compared to the *S. cerevisiae* amino acid biosynthesis regulator Leu3 [100] as a model for an intramolecular conformation change in

response to a small molecule [103]. Leu3 was shown to functionally respond to α -isopropylmalate in heterologous systems [104] suggesting no additional proteins are required for Leu3 function. However, no such independent functionality has been demonstrated for XlnR.

ACE3

Uncharacterized transcriptional regulators with increased transcript abundance on cellulosic substrates were identified in *T. reesei* transcriptional profiles. Over expression strains for these regulators were then tested for altered cellulase activity in lactose grown cultures [105]. One of these regulators, named ACE3 was found to stimulate increased cellulase expression when over expressed and a deletion mutant for that gene lost induction of several cellulases and one xylan esterase.

BglR

Analysis of point mutations in the *T. reesei* hyper-secretion strain PC-7 identified a missense mutation in a ZnBNC transcription factor [106]. Deletion of that gene in the parental strain used to create PC-7 increased cellulase production on cellobiose [106]. The authors postulate that loss of BglR reduces expression of β -glucosidase, and thus reduces carbon catabolite repression in the mutant.

COL-26

To test if the *N. crassa* ortholog to BglR (NCU07788, COL-26) also influenced expression of β -glucosidase genes, I cultivated the deletion strain from the *N. crassa* deletion collection. Control growth experiments on various carbon sources strongly indicated altered carbon catabolite repression in this strain.

Later, my labmate, Yi Xiong, found *col-26* had altered transcriptional profiles in another transcription factor mutant ($\Delta vib-1$, NCU03725) and that it was co-expressed with the known carbon catabolite regulating transcription factor *cre-1*. Yi found the $\Delta col-26$ deletion strain to have reduced sensitivity to 2-deoxyglucose, a further indication that it is involved in carbon catabolite repression [107].

CibR

An insertional mutagenesis screen in *Aspergillus aculeatus* produced a strain with reduced cellulase activity [108]. The disrupted gene was found to be a ZnBNC transcription factor and the deletion mutant had reduced induction of genes previously found to exhibit XlnR-independent induction on cellulose.

AraR

Transcriptional profiling of uncharacterized transcription factors with high sequence similarity to XlrR identified a gene that was induced by arabinose and arabitol [109]. A deletion mutant for AraR grew poorly on arabinose and had impaired activity for α -L-arabinofuranosidase activity and several enzymes involved in arabinose utilization.

RhaR

A co-regulation analysis of *A. niger* transcription factors and enzymes with known activity on pectin or pectin associated sugars identified a candidate regulator of L-rhamnose catabolism and pectin degradation [110]. A deletion mutant for RhaR had drastically reduced growth on pectin and reduced expression of several plant cell wall degrading enzymes, mostly active on rhamnogalacturonan, and of genes involved in rhamnose utilization. Based on the remaining growth on pectin and transcriptional response to pectin by the Δ RhaR mutant, the authors postulate the existence of an independently acting regulator of galacturonic acid utilization.

Follow-up studies found that in *A. nidulans* and *N. crassa* deletions of the RhaR orthologs compromised growth on rhamnose, and reduced expression of secreted α -L-rhamnosidases [111].

ManR

An insertional mutagenesis screen in *A. oryzae* produced a strain with deficient utilization of mannan [112]. The disrupted gene was found to be a ZnBNC transcription factor, which had been incorrectly annotated in the genome due to a gene model error. The authors reported loss of mannanase induction in the deletion mutant and increased mannanase expression in a ManR overexpression mutant on mannan and mannobiose, with no phenotype on cellobiose. Furthermore, their microarray data (GEO accession GSE40590) show that ManR also regulated several cellulases and at least one xylanase. The authors did not realize at the time of publication that ManR was the ortholog of *N. crassa* CLR-2, a subject for further discussion in Chapter 5.

CLR-1/ClrA and CLR-2/ClrB

The major focus of the present work. Identified from transcriptional profiles of *N. crassa* culture on ground *Miscanthus* and in a screen of transcription factor mutants in the *N. crassa* deletion collection, respectively. CLR-1 and CLR-2 regulate transcription of a substantial regulon in response to cellobiose.

Chapter 2 :

Carbon Source, Growth and Regulated Transcription of Cellulases in *Neurospora crassa*

Abstract

Cellulase transcription in filamentous fungi is regulated by a balance of an active induction in response to environmental sugars and repression by preferred carbon sources, known as carbon catabolite repression (CCR). When studying the regulation of those enzymes and the their global transcriptional context, it is therefore critical to carefully establish the relevant conditions for repression, induction, and neutral regulation. I established those conditions with a series of cultures tracking growth and enzyme secretion, and characterized the *Neurospora crassa* transcriptional response to cellulose with RNA sequencing. Glycerol, historically consider a neutral carbon source, strongly suppressed cellulase secretion. Xylan was the least repressing non-inducing carbon source. Transcriptional profiles of *N. crassa* cultures shifted to cellulose (Avicel) more closely resemble carbon starvation than a reference sucrose condition. In comparison to carbon-starved cultures, Avicel-induced cultures experienced a strong induction of approximately 200 genes, including all major cellulases. This Avicel regulon will be the reference set for later exploration of cellulase regulators.

Introduction

Cellulase transcription responds both to active induction by soluble inducers and global nutritional state through carbon catabolite repression (CCR) [27,78,113,114]. In early studies of cellulase regulation, without concerted control for nutritional state, discrimination between active cellulase induction and relief of CCR was poor [71,73,74]. Over time, appreciation has grown for the widespread effects of nutritional state on global metabolism in general and for expression of plant cell wall polysaccharide degrading enzymes in particular [115]. To control for CCR effects in transcriptional studies, use of chemostat cultures have thus become more common [116,117]. Unfortunately, such culture strategies are extremely limiting in throughput in the best of cases, are particularly difficult for filamentous fungal species, and are impossible with regulatory mutants that fail to grow at all on the inducing carbon source. However, even in batch culture, many obfuscating effects of CCR and variant growth rate can be reduced or eliminated through careful selection of culture conditions.

In this chapter I introduce the experimental framework and culture strategies that I employ throughout this dissertation and establish the appropriate conditions for global comparisons of active induction by cellulose as opposed to carbon catabolite repression. Those conditions established, I then introduce the robust transcriptional response of *N. crassa* to Avicel.

Cellulase transcription is regulated by inducing substrates and strict carbon catabolite repression

To establish standard conditions and carbon sources for these studies on cellulase induction, I tested cellulase activity in cultures on cellulose with a variety of additional carbon sources. In a culture strategy widely used throughout this work, I grew cultures of wild type *N. crassa* (strain OR74 A [118]) in Vogel's minimal medium [119] with 2% sucrose (the historically standard liquid cultivation condition for *N. crassa*) for 16 hours, washed them, and transferred them to fresh media with 1% wt/vol Avicel with or without additional carbon sources (glucose, xylose, xylan, acetate, glycerol or glutamic acid) at 1%, 0.5%, or 0.1% wt/vol. I sampled culture supernatants and assayed accumulating cellulase activity (endoglucanase activity) over a 4-day time period.

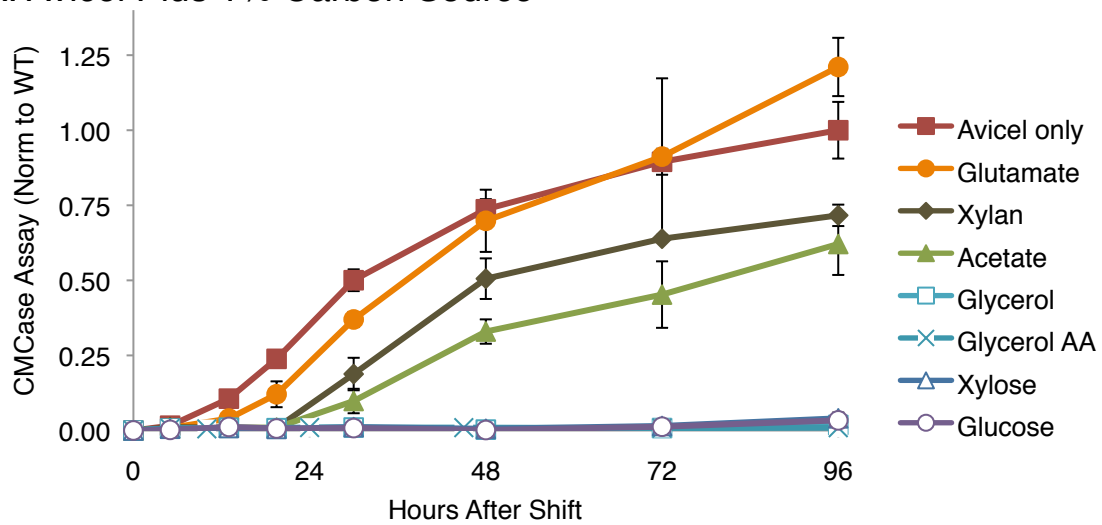
At 1% supplementation (Figure 4A), glucose, xylose, and glycerol completely prevented cellulase accumulation, whereas xylan and acetate had measurably lower enzyme activity than Avicel cultures. Glutamate slightly delayed accumulation early in the culture but later in the time course cellulase activity was enhanced, presumably through the additional carbon available to the culture.

At 0.5% supplementation, CCR effects were reduced, although the relative inhibitory strength of the carbon sources remained similar. Only at this lower concentration was it clear that glycerol is actually more repressing for *N. crassa* cellulases than glucose and xylose (Figure 4B). This strong repression by glycerol was unaffected by the addition of 100 µg/mL ascorbic acid, which greatly increases *N. crassa* growth on glycerol [120].

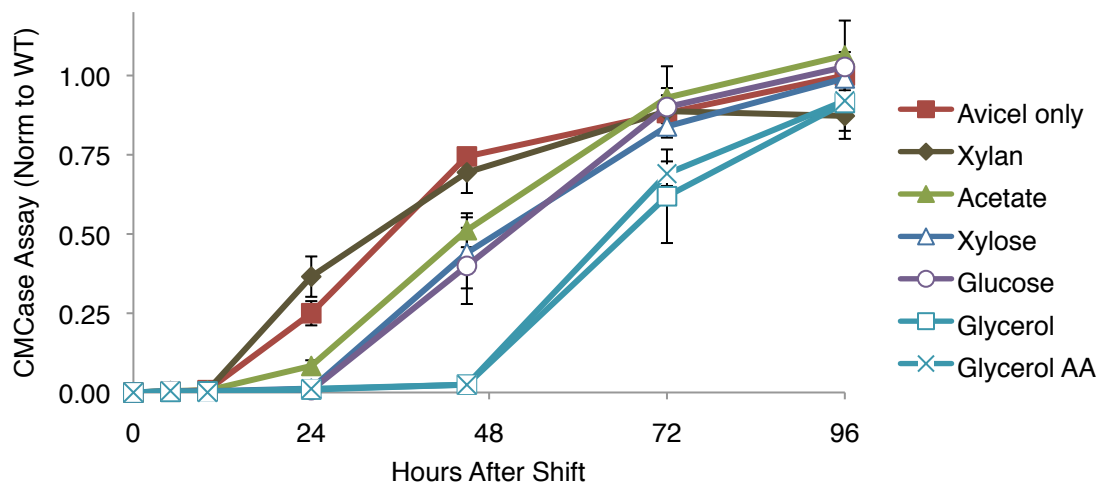
At 0.1% supplementation (Figure 4C), carbon catabolite effects were reduced to a level not detectable in this experiment.

Deletion of the known CCR mediating transcription factor *cre-1* [114] reduced the inhibitory effects of glucose, xylose and glycerol at moderate concentrations (Figure 4D). However, deletion of *cre-1* did not significantly alter the repressive effects of acetate and at higher concentrations, other carbon sources such as glucose, remain inhibitory (Figure 4E),

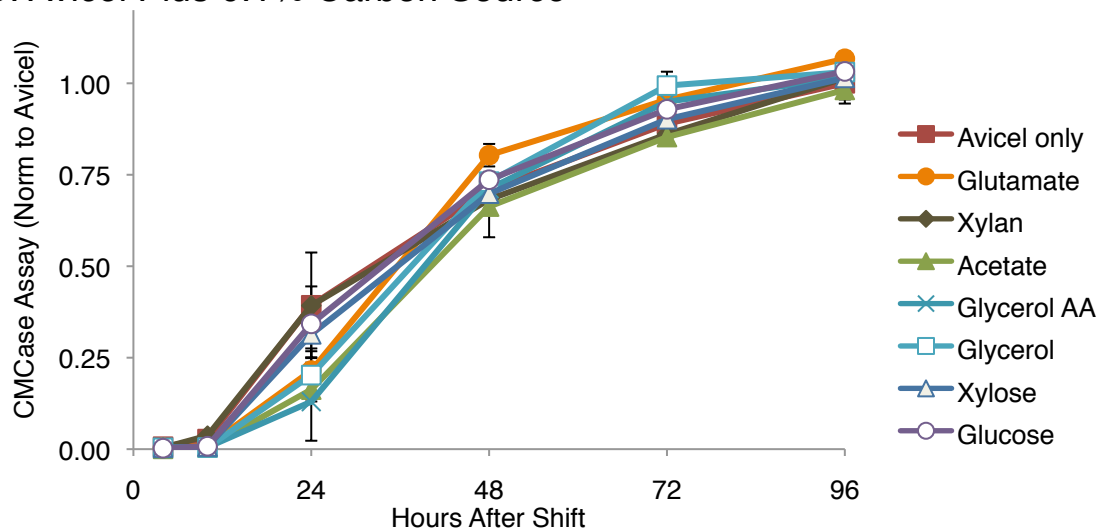
A. Avicel Plus 1% Carbon Source



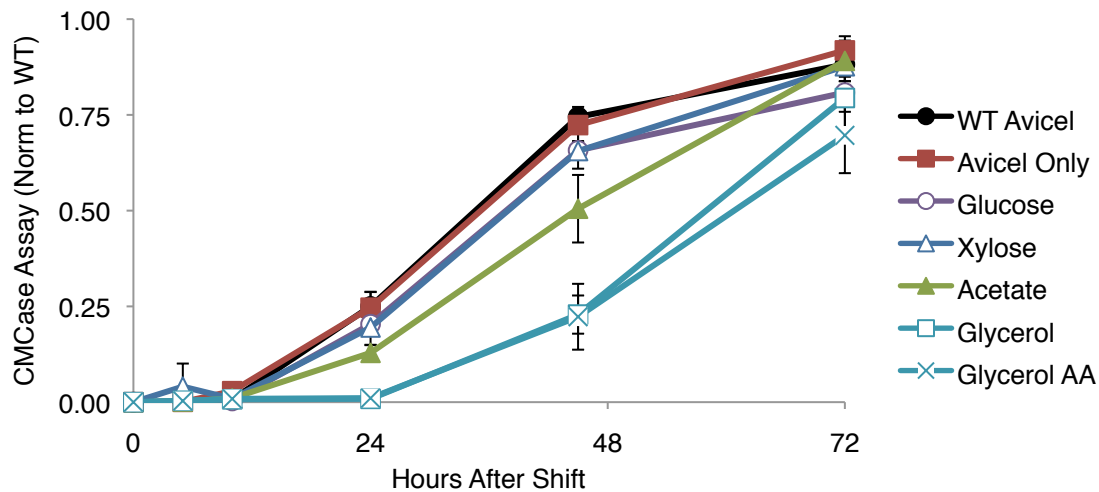
B. Avicel Plus 0.5% Carbon Source



C. Avicel Plus 0.1% Carbon Source



D. $\Delta cre-1$ On Avicel Plus 0.5% Carbon Source



E. $\Delta cre-1$ On Avicel Plus Glucose

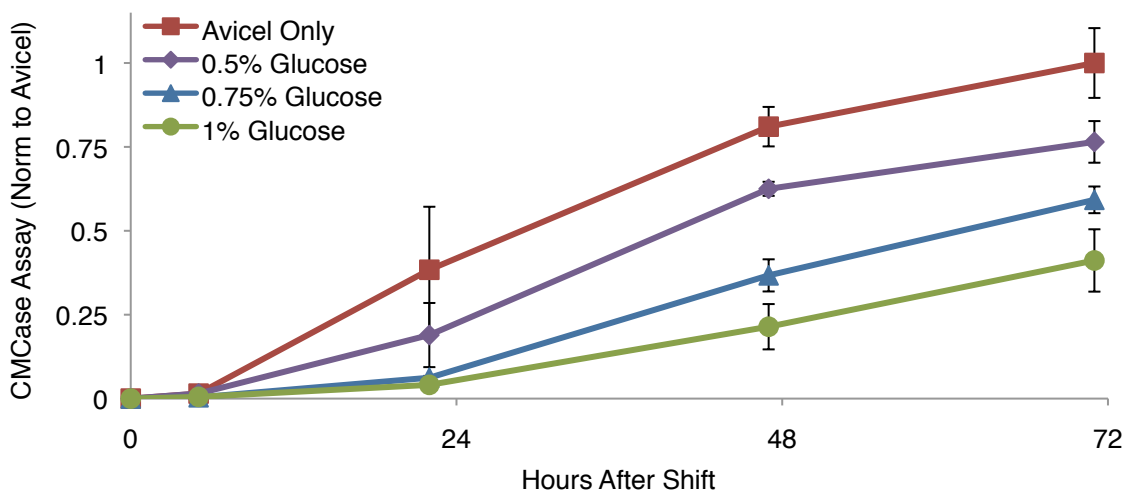


Figure 4. Endoglucanase Activity in *N. crassa* supernatants after shift to Avicel plus repressing carbon source.

(A-C) WT *N. crassa* was grown 24hrs on 2% sucrose, washed and shifted to fresh inducing/repressing media with 1% Avicel (Avicel only) or 1% Avicel plus additional carbon sources.

(A) 1% carbon source added

(B) 0.5% carbon source added

(C) 0.5% carbon source added.

(D) A similar time course with 0.5% added carbon sources in the *cre-1* deletion background.

(E) The $\Delta cre-1$ mutant on 1% Avicel with 0, 0.5, 0.75 and 1% glucose. Supernatants were sampled and tested for endoglucanase activity by the Azo-CMCase activity.

To correlate the CCR effects of the various carbon sources tested above and identify the best "neutral" carbon source for cellulase induction, I carried out a simple growth experiment in cultures with a 1% soluble carbon source (Figure 5). There is an imperfect correlation between growth rate and strength of cellulase repression for the carbon sources tested. While glycerol was the most repressive carbon source, it had the second lowest growth rate. On the other hand, while xylan

exhibited very weak cellulase repression, only glucose and xylose cultures grew faster. For glucose, xylose, acetate and glutamate, however, there was an inverse correlation between growth rate and cellulase accumulation.

These experiments are rudimentary, and the measurement of cellulase activity in culture supernatants is a cumulative, low-resolution assessment of regulation. However, they illustrate the critical balance between inductive and repressive signals in cellulase gene expression. As all transcriptional analysis are comparative, they raise the essential question; what is the most appropriate reference condition for assessing cellulase induction without conflating repressive effects of growth rate and carbon source? To address the questions effectively requires at once a broader (reflective of the entire genome) and more targeted (temporally limited) approach.

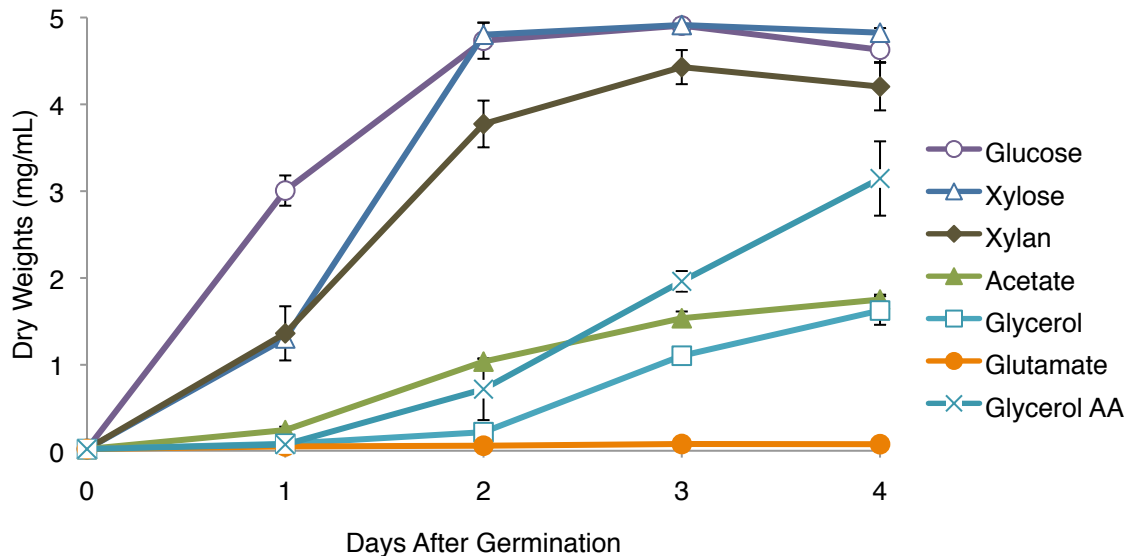


Figure 5. Growth curves on cellulase repressing carbon sources.

The transcriptional response of wild type *N. crassa* in response to cellulose

To directly assess regulation of cellulase transcription, I used next generation RNA sequencing to profile genome-wide mRNA abundance in WT *N. crassa* when exposed to Avicel. I shifted sucrose-grown *N. crassa* to Vogel's medium with 2% Avicel, 2% Sucrose, or no added carbon source and sampled the shifted cultures at 1 hour and 4 hours after the shift. For better temporal information on the cellulase response, I combined these data with previously sequenced samples (prepared by other lab members) from 2% Avicel, 30 minutes and 2 hours after the shift. The 1 and 4 hour samples were collected in biological triplicate, whereas the 30 minute and 2 hour samples were single replicates.

Within 30 minutes of transfer to Avicel, mRNA abundance for cellulase genes increased 2-10 fold (Figure 6A). This level of transcription was maintained until 1 hr after the shift, then cellulolytic genes were further induced several thousand fold. Gene expression patterns for several hemicellulase genes exhibited a similar two-stage induction (Figure 6B). I interpreted this two-stage induction as an initial,

rapid, relief of CCR followed by a slower induction in response to cellobiose and/or other sugars released from the Avicel by the de-repressed cellulase gene expression.

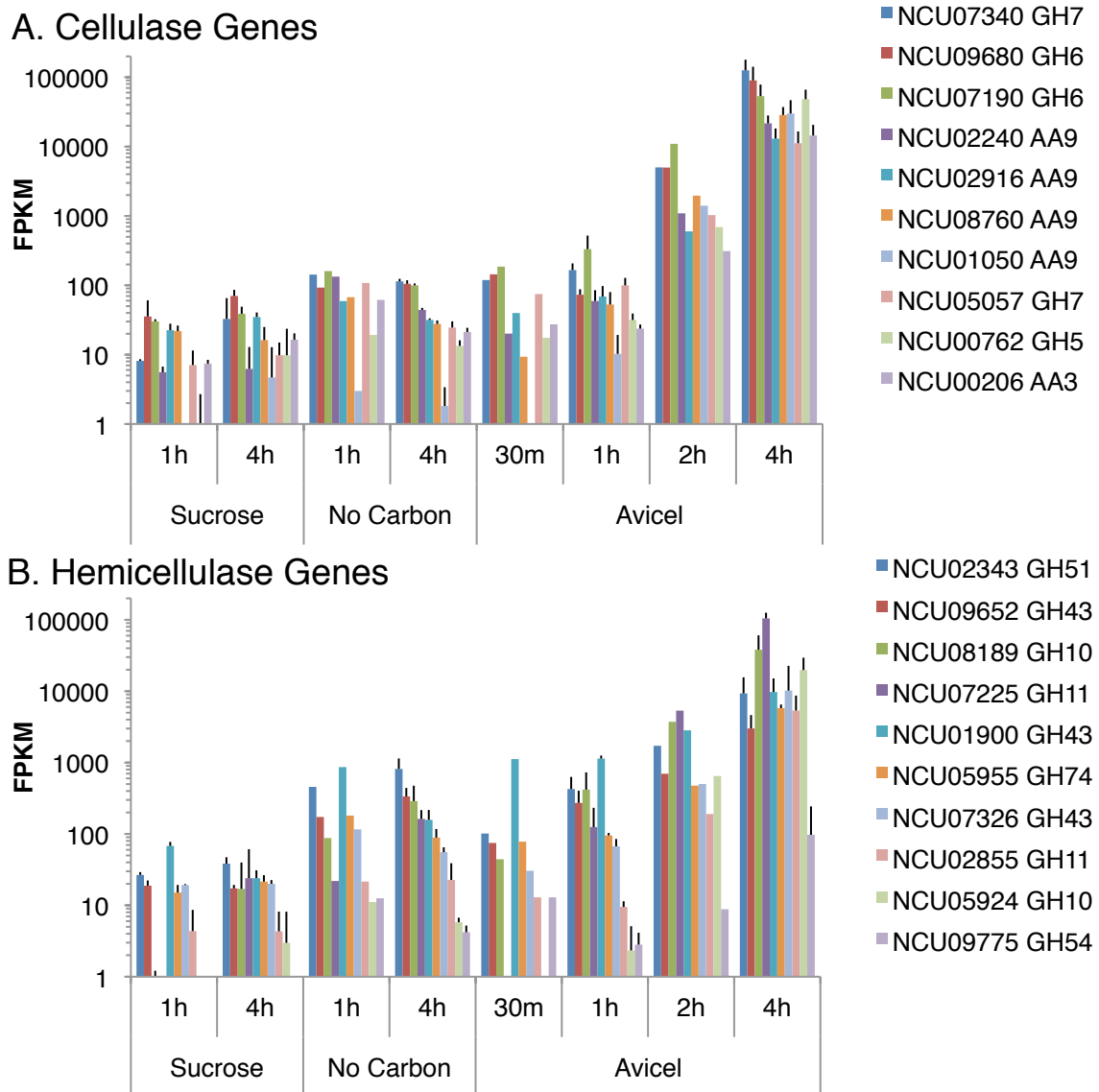


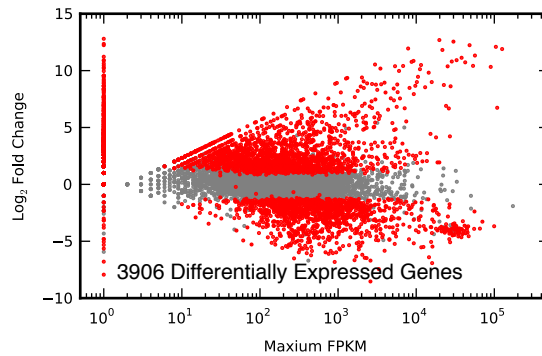
Figure 6. Transcriptional induction by cellulosic substrates in *Neurospora crassa* as measured by RNA-Seq.

Message abundance of 10 cellulase genes (A) and 10 hemicellulase genes (B) at selected times after transfer of a 16 hr sucrose grown culture to either sucrose, no carbon source or Avicel. Abundance expressed in Fragments Per Kilobase unique exon sequence per Megabase of library mapped (FPKM). Error bars show +/- 1 standard deviation for conditions with triplicate measurements. Conditions without error bars are from single libraries. Genes listed by 5-digit NCU number [121] and CAZY class [122].

In order to guide effective analysis of active transcriptional induction by cellulose, as opposed to changes in CCR or generalized stress responses, I compared differentially expressed genes across the three conditions at 4 hours after the shift. Between the Avicel and sucrose cultures 3906 genes were differentially expressed (approximately 41% of the genome, Figure 7A) with functional category enrichments [123], dominated by a general decrease in expression for genes involved in transcription, translation and mitochondrial biogenesis (Figure 7B and C). Between the Avicel and no carbon cultures, only 333 genes were differentially expressed (Figure 7D). This gene set was enriched for induction of C-compound and carbon metabolism, extracellular metabolism and polysaccharide binding functional categories, but with no enrichment in any gene categories for the few down-regulated genes (Figure 7E and F).

Of the 333 genes differentially expressed between Avicel and no carbon conditions, 212 had greater expression on Avicel than under either no carbon or sucrose conditions (Figure 8). I take this gene set as a conservative “Avicel regulon”. The Avicel regulon included 17 of 21 predicted cellulase and 11 of 19 predicted hemicellulase genes [44], 22 genes encoding enzymes with predicted activity on polysaccharides and 56 additional genes encoding proteins predicted to enter the secretory pathway (Dataset S1). Also induced were genes for 13 transporters (including 2 recently identified cellodextrin transporters [46]), 3 components of the ER translocation complex, 2 ER disulfide isomerases, 2 ER chaperones (HSP70 and calnexin), 7 cytosolic enzymes with predicted activity on disaccharides, a homolog to *Aspergilli* and *H. jecorina xlrR/xyr1* [81,124], as well as both *clr-1* and *clr-2* (discussed below). Also included in the Avicel regulon were 92 proteins annotated as conserved hypothetical proteins or with no predicted function, 31 of which are predicted to enter the secretory pathway. The full regulon and mean FPKM 4 hours after a transfer to Avicel or no carbon are listed in Appendix A.

A. Avicel vs Sucrose



D. Avicel vs No Carbon

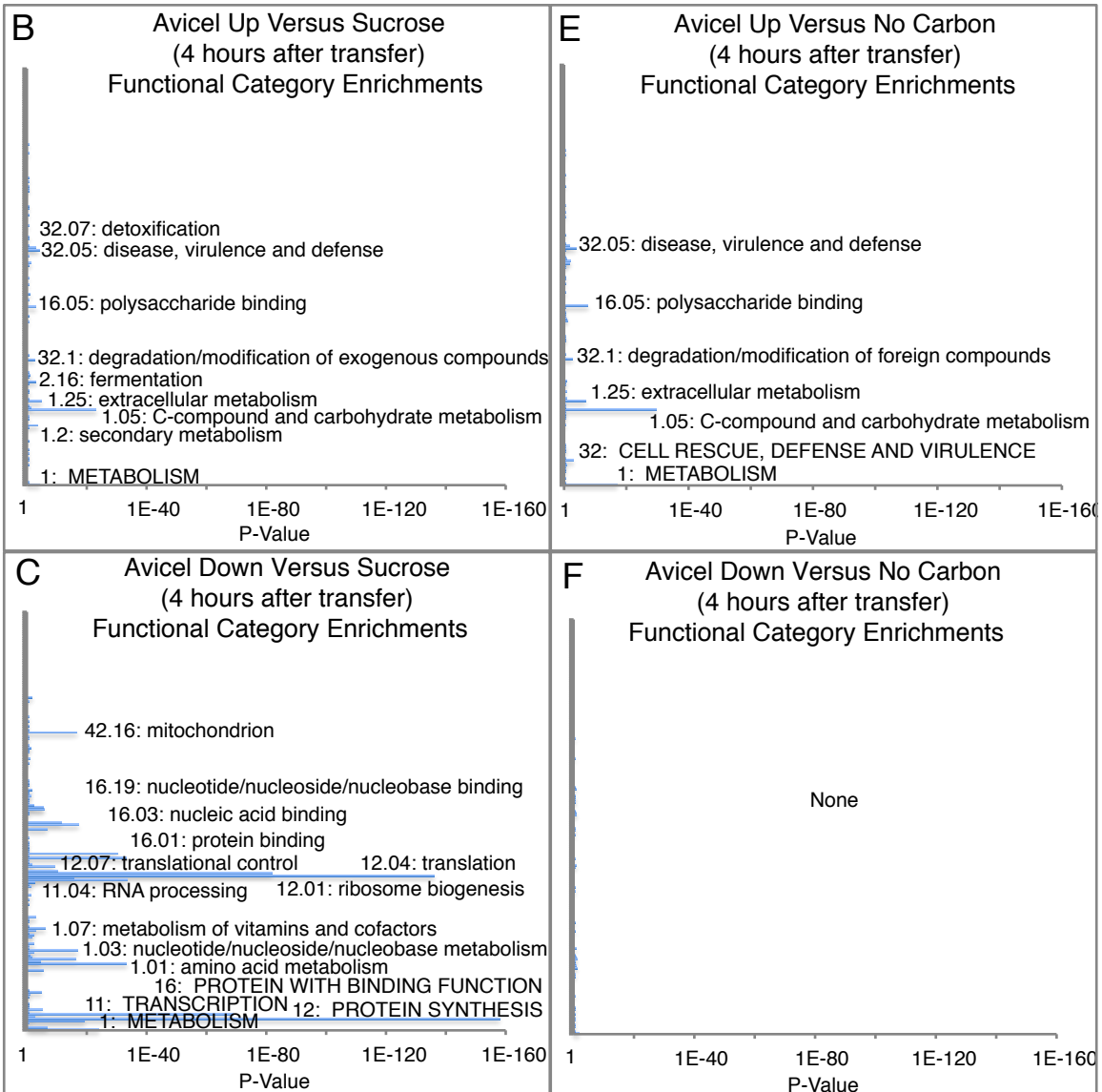
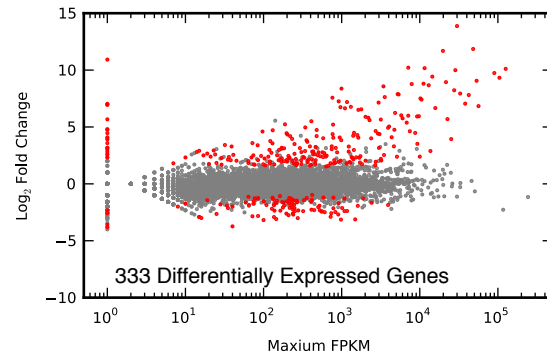


Figure 7. Differentially expressed *N. crassa* genes after shift to either Avicel versus sucrose cultures or Avicel versus no carbon cultures.

(A) Abundance on Avicel vs abundance on sucrose for *N. crassa* transcripts by RNAseq. Log₂ ratio of Avicel/sucrose FPKM vs maximum FPKM in either condition. Genes detectable in one condition but

not the other are plotted along the y-axis with the non-expressed condition set to 1 FPKM. Genes exhibiting differential expression ($P < 0.05$) are plotted in red.

(B) Relative enrichment of functional categories [123] among genes with significantly higher transcript abundance on Avicel than on sucrose. The probability of randomly selecting a subset of genes with the observed functional category enrichment is plotted along the x-axis. Categories with $P < 10^{-2}$ are labeled.

(C) Relative enrichment of functional categories among genes with significantly higher transcript abundance on sucrose than on Avicel.

(D) Abundance on Avicel vs abundance on no carbon for *N. crassa* transcripts by RNAseq.

(E) Relative enrichment of functional categories among genes with significantly higher transcript abundance on Avicel than on no carbon.

(F) Relative enrichment of functional categories among genes with significantly higher transcript abundance on no carbon than on Avicel.

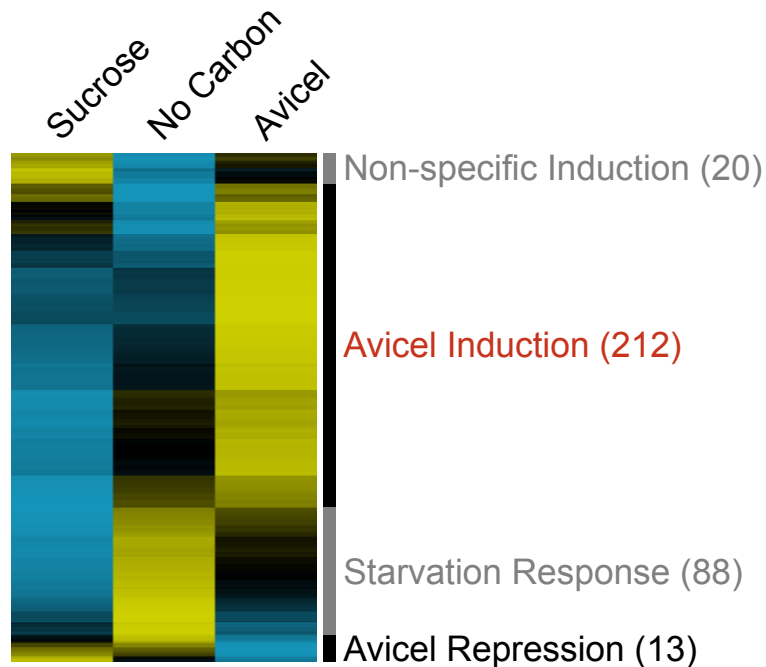


Figure 8. Heat map of expression levels in wild type on sucrose, no carbon and Avicel for 333 genes differentially expressed on Avicel and no carbon 4 hrs after shift.

Twenty genes had the greatest transcript abundance on sucrose and are most likely involved in glucose metabolism. A total of 212 genes were most highly expressed on Avicel and are most likely involved in cellulose degradation. A total of 88 genes were expressed to the lowest level on sucrose and to the highest level on no carbon and are most likely part of a general starvation response. Thirteen genes were expressed at their lowest level on Avicel. These genes have a mixture of expression levels on sucrose and no carbon conditions and relatively small absolute changes in transcript abundance.

Discussion

Carbon catabolite repression is a complex and multifaceted process. With even the very basic growth and cellulase activity experiments in Figure 4 and Figure 5, complex unexpected behavior emerged. In the filamentous fungal community, discussion of CCR often centers on a single repressing transcription factor, CRE-1, ortholog of *S. cerevisiae* Mig1 [114,125]. While CRE-1 is undoubtedly an important mediator of CCR as it affects cellulase transcription, it is clearly not the only one. Deletion of CRE-1 does not completely alleviate repressive effects of glucose and other sugars, and it does not affect repressive effects of some carbon sources, such as acetate, at all. To the extent other players are discussed in current literature [115], there is heavy emphasis on pathways studied almost entirely in *S. cerevisiae* even though these pathways have evolved in the absence of plant cell wall utilization and when comparative studies on orthologs in filamentous fungi have been done, sometimes significant differences in function have sometimes been found. For instance the Swi/Snf complex does not phosphorylate Cre1 in *T. reesei* as it does Mig1 in *S. cerevisiae* [126]. Clearly, many genes characterized in *S. cerevisiae* do show similar behavior when studied in filamentous ascomycetes, but great care must be taken when transferring knowledge between organisms with such different cellular biology and lifestyles. Ultimately, further advances in this area must come from studies in filamentous fungal species without an assumption of conserved components in non-cellulolytic eukaryotic systems. Until the missing components of CCR in filamentous fungi are identified, mastery of this regulatory aspect, and with it mastery of cellulase regulation and secretion will remain out of reach.

In this work, my primary focus will be on active cellulase induction, for which CCR is an obfuscating complication. As such, careful attention to experimental conditions will be required. In most transcriptional studies an obvious parameter to maintain invariant is the carbon source used for culture, but in the experiments to follow, I will explore transcriptional defects in mutant strains unable to utilize Avicel. As my initial studies on WT cultures showed, a rapid shift to Avicel is initially a starvation condition, without active induction of cellulases for the first two hours, followed by a rapid increase in cellulase expression over the 2-4 hour period. The WT response to Avicel and to no carbon conditions differ primarily in expression of actively induced genes and not in global transcriptional profiles altered by growth rate or nutritional state. For experiments requiring prolonged culture, xylan will be used as a nearly — though not entirely — neutral carbon source. Though glycerol has been described as a neutral substrate for *T. reesei* [77] it is wholly inappropriate for use in this regard for experiments with *N. crassa*.

The cohort of approximately 200 *N. crassa* genes induced on Avicel will be discussed in more detail in later chapters as differences in regulation in *N. crassa* and across species in *A. nidulans* begin to elucidate more aspects of their biochemical and biological function. A critical feature of the Avicel regulon that should be addressed here, however, is the induction of not only cellulases, but all major xylanases and the xylose utilization pathway. As we shall see, some of these xylanases are indeed co-regulated with cellulases and other plant cell wall degrading genes as a synergistic complex of complimentary enzymes to degrade

complex substrates. However, several of the Avicel induced xylanases (e.g. *gh10-2* NCU08189 a major endoxylanase) are known to be regulated by XLR-1 [89] and were never detected in proteomic studies of supernatants from *N. crassa* cultivated on Avicel [127].

Coincident with my initial RNAseq studies on WT and mutant *N. crassa* strains, lab member Elizabeth Znameroski established strains and conditions to induce cellulase expression directly with cellobiose. She reduced the CCR that usually obfuscates cellulase induction by direct exposure of *N. crassa* to cellobiose by deleting genes for three of the most abundant β -glucosidases, thereby slowing glucose release to the point that it is consumed before it can accumulate. RNAseq on this mutant exposed to pure cellobiose recapitulated most of the Avicel regulon identified in this chapter, but not several of the Avicel-induced xylanases or the xylose utilization pathway genes [128]. These data together suggest that Avicel is not pure cellulose, but is contaminated with trace amounts of hemicellulose. Compositional analysis by J. Phillip Benz has since confirmed the presence of this contamination, and of similar cross-contamination of plant cell wall polysaccharides in other purified preparations of xylan and pectin (unpublished results). As we shall see, this contamination is not merely a complication, but can actually serve as a useful internal control, by verifying that loss of cellulase induction is not due to a technical issue or a general biological defect preventing any and all robust transcriptional responses to environmental stimuli.

With these constraints, conditions and tools in mind, we are now ready to search for unknown regulators of cellulase gene expression and coherently assess their function.

Acknowledgements

Chaoguang Tian, Jianping Sun, Spencer Diamond, and Morgann Reilly developed culture techniques for *N. crassa* on cellulosic substrates and assays for quantifying cellulase activity in culture supernatants that I have employed or modified for my experiments.

Jianping Sun and James Craig contributed directly to my RNAseq analysis by performing a pilot transfer and RNAseq experiment on Avicel and sucrose media and provided the data for Avicel cultures 30 minutes and 2 hours after transfer.

Charles Hall and David Kowbel built a computational infrastructure, toolset and pipeline for RNAseq analysis [129] that I adapted with my scripts and additions to analyze my data.

Published material

Data in Figure 6, Figure 7, and Figure 8 were initially published in [130]: Coradetti ST, Craig JP, Xiong Y, Shock T, Tian C, et al. (2012) Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *Proc Natl Acad Sci U S A* 109: 7397-7402.

Methods

Strains and culture conditions.

N. crassa OR74A, FGSC 2489 [131] was cultivated on Vogel's minimal medium (VMM) [132] and 2% sucrose, the historically standard growth medium for *N. crassa*. Cultures were incubated on agar slant tubes for 2 days at 30C in the dark to stimulate mycelia extension, then 4-10 days at 25C in the light to obtain conidia. Except for this initial incubation at 30C, all experiments were carried out at 25C in constant light to dampen circadian rhythms. Unless specified otherwise, all carbon sources were 2% wt/vol. All liquid culture was carried out with constant shaking at 200rpm.

For growth curves and enzyme activity assays in this chapter, conidia were inoculated at 5×10^5 conidia/mL into 3mL VMM sucrose in 24-well plates (GE Healthcare UK Limited) and grown for 24 hours. Mycelia were then washed in VMM with no carbon and re-suspended in 3mL VMM with 0%-2% carbon source. Supernatants were sampled up to 96 hours after transfer and stored at 4C for analysis. For growth assays, individual cultures were harvested by filtering on Whatman (#1) filter paper, lyophilized and weighed. As such, growth curves do not represent a sampling of a single continuous culture, but average final weights from cultures stopped at each time point.

For transcriptional profiling, conidia were inoculated at 1×10^6 conidia/mL into 100mL VMM sucrose in 250mL shake flasks and grown 16 hours. Cultures were then washed in VMM with no carbon source and transferred to 100mL fresh VMM with sucrose, cellulose (Avicel® PH-101, Sigma Aldrich) or no carbon added. At 30 minutes to 4 hours after the media shift, mycelia were harvested by filtering on Whatman (#1) filter paper and flash frozen with liquid nitrogen. Total RNA was extracted with TRIzol reagent (Life Technologies) as described in [44] and further purified with RNeasy kits (Qiagen).

RNA sequencing and transcript abundance.

Libraries were prepared with standard protocols from Illumina Inc (San Diego, CA) and sequenced on the Illumina Genome Analyzer IIx and HiSeq 2000 platforms. RNA samples were reverse transcribed and prepared for high throughput sequencing with protocols adapted from Illumina Inc. Briefly, mRNA was purified with Dynabeads® Oligo dT magnetic beads (Illumina). Purified mRNA was fragmented with buffered zinc solution from Ambion (Cat # AM8740). First and second strand cDNA synthesis was carried out using Superscript II Reverse Transcriptase (Invitrogen) and DNA pol I (Invitrogen) and random primers. Illumina sequencing adapters were then ligated to the cDNA, 200bp fragments were purified by gel electrophoresis, and PCR enriched with Pfx DNA polymerase (Invitrogen). Libraries were sequenced on the Illumina Genome Analyzer IIx or HiSeq 2000 DNA sequencing platforms.

Sequenced libraries were mapped against predicted transcripts from the *N. crassa* OR74A genome [121] (version 10) with Tophat v1.2.0 [133]. Transcript abundance (FPKM) was estimated with Cufflinks v0.9.3 [134] using upper quartile

normalization and mapping against reference isoforms. Profiling data are available as a supplemental Dataset (S1) and at the GEO (<http://www.ncbi.nlm.nih.gov/geo/:SeriesRecord/GSE35227>).

Independent triplicate cultures were sampled and analyzed for the WT strain on Avicel, sucrose and no carbon at 4 hrs after the media shift. Raw counts for reads mapping to unique exons were tallied with HTSeq [135] and used as inputs for both the EdgeR and DESeq software packages. Genes with a multiple-hypothesis-adjusted p-value below 0.05 from both models were taken as differentially expressed between conditions.

FPKM of genes in the Avicel regulon were hierarchically clustered with the Cluster 3.0 software suite (<http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm>). Conditions compared were mean FPKM from triplicate WT cultures on Avicel, sucrose, and on no carbon. Prior to clustering, FPKMs were log transformed, normalized across strains/conditions and centered on the geometric mean across strains/conditions on a per-gene basis. The average linkage method was used for cluster generation, with uncentered correlation as the similarity metric.

Functional category analysis.

Lists of differentially expressed genes were matched against the MIPS Functional Category Database [123]. Significance of enrichment for functional categories was calculated with a hypergeometric function as described in [136].

Enzyme activity assays.

Relative enzymatic activity of culture supernatants was assayed with Remazolbrilliant Blue R conjugated CMC kits from Megazyme (Wicklow, Ireland). Briefly, in these assays enzyme activity is measured by the release of ethanol soluble short-chain dye-conjugated oligomers from the long-chain substrate, followed by relative quantification with a spectrophotometer (Spectramax M2 manufactured by Molecular Devices).

Chapter 3 :

Screening the *N. crassa* Deletion Collection for Novel Transcription Factors Regulating Cellulose Utilization

Abstract

In order to identify transcription factors mediating cellulase gene induction, the *N. crassa* deletion collection was screened for transcription factor mutants with deficient growth on cellulose. Twenty-nine mutants were found to have repeatable phenotypes ranging from slight decreases in growth and enzyme activity in culture supernatants to qualitative loss of growth and cellulase secretion. Twelve of these mutants had broad defects in cellulase transcription as detected by RNA microarray analysis. Three genes corresponding to deletions of NCU02017, NCU07705, and NCU08042 were identified as likely direct regulators of cellulase induction based on a complete lack of cellulase induction in the respective deletion mutants.

Introduction

My goal in this work is to explore — and ultimately manipulate — regulation of cellulase secretion in filamentous fungi. As my collaborators and I began this work, it was clear that several fungal regulators of cellulase expression were not yet identified. In *A. niger* and *T. reesei*, it was known that expression of cellulase and hemicellulase genes requires the conserved fungal transcription factor XlnR/XYR1 [81,137-139]. Yet it was also known that although *xlnR/xyr1* homologs are important for hemicellulase induction in *Fusarium* species and in *Neurospora crassa*, they do not play a significant role in cellulase gene expression [89,97,140]. Other genus-specific transcriptional activators (such as ACE2 in *H. jecorina*) appeared to have some specificity for particular inducers, but deletion strains for these transcription factors still induced significant cellulase secretion [141]. These findings suggested a complex network of transcriptional activators is required for cellulase and hemicellulase gene induction and secretion in filamentous fungi. Mastery of that network would require identification of additional components as well as a detailed understanding of their interactions and how they vary among fungal species.

Indeed, preliminary transcriptional studies of *N. crassa* identified an uncharacterized predicted transcription factor that was induced on cellulosic substrates [44]. Preliminary studies of a deletion mutant for this transcription factor demonstrated a strong growth defect on cellulose, further motivating a systematic search for more such regulators. To identify novel transcription factors required for cellulose degradation in an unbiased fashion, we screened the *N. crassa* transcription factor deletion set [131] for mutants with deficient growth on Avicel. In this chapter I summarize that screen and introduce several transcription factors required for robust cellulase gene expression. Two of these transcription factors, encoded at loci NCU07705 and NCU08042 (*clr-1* and *clr-2*, respectively). Had

uniquely critical and specific functionality in cellulase secretion. Exploration of that functionality is the major focus of the remaining chapters of this dissertation.

Growth and enzyme activity of transcription factor deletion mutants on cellulose

The transcription factor deletion strain screen was carried out in three phases; an initial growth screen, a screen of enzyme activity, and finally a screen of global transcriptional profiles. In the first phase (performed by laboratory member Terry Shock), 269 genes with predicted DNA binding domains were selected from the *N. crassa* deletion collection and screened for visually apparent growth phenotypes on Vogel's media with 2% Avicel in 5mL liquid culture. This initial screen yielded 42 candidate strains for more thorough examination.

In the second phase, I cultured the 42 candidate strains on sucrose, xylan (a major component of hemicellulose) and Avicel, visually monitored growth (Figure 9) and assayed accumulated enzyme activity in culture supernatants. Cellulase activity was monitored by glucose and cellobiose release from Avicel and reducing sugars release from xylan (Figure 10). This second round of screening confirmed a spectrum of phenotypes for 29 candidate deletion strains. Seven deletion strains had severe growth and activity defects on Avicel but had WT growth on other carbon sources — with one exception, Δ NCU02017, which had lowered conidia production, and altered colony morphology. Ten strains had reduced growth and enzyme activity on Avicel, but also had pronounced pleiotropic defects including altered morphology, low conidiation, altered carotenoid production, and generally slower growth on all carbon sources. Seven strains had mild growth and activity defects. Thirteen strains had no consistent activity and growth phenotypes. Five strains had apparently increased growth or enzyme activity. Phenotypes are summarized in Table 2. Note that the enzyme assays employed in this phase were quite variable and few biological replicates were performed, so results should be treated as qualitative, rather than quantitative.

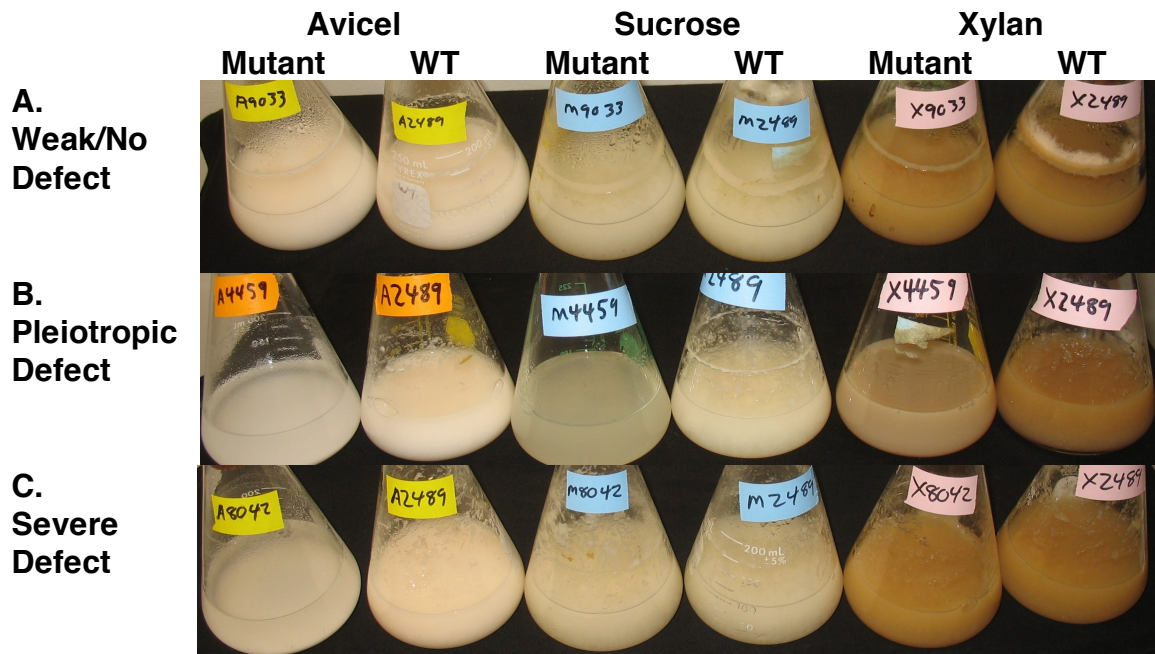


Figure 9. Visual inspection of transcription factor deletion mutants growing on Avicel, sucrose and xylan.

(A) Representative culture (Δ NCU09033) of a mutant with no obvious growth phenotype on Avicel (left), sucrose (middle), or xylan (right).

(B) Representative culture (Δ NCU04459) of a mutant with general growth defects on all carbon sources.

(C) Representative culture (Δ NCU08042) of a mutant with specific growth defects on Avicel

Table 2. Phenotypes of transcription factor deletion strains selected for secondary screening

Category	Locus	Annotation	Visual Growth			Enzyme Activity on Avicel (% WT)		
			Avi	MM	Xyl	Gluc	Cell	Xyln
Severe	NCU02017	<i>ada-2</i>	-	++	oo	2	1	0
	NCU03725	<i>vib-1</i>	-	++	+	1	2	1
	NCU07705	hypothetical	-	++	++	0	3	12
	NCU08042	hypothetical	-	++	++	1	0	14
	NCU06407	<i>vad-3</i>	o	++	++	34	50	28
	NCU10006	hypothetical	+	++	++	17	5	29
	NCU05685	hypothetical	+++	++	++	11	6	34
Pleiotropic	NCU06205	<i>rco-1</i>	o	o	oo	5	6	3
	NCU04459	SAGA subunit	-	+	-	2	6	14
	NCU01461	hypothetical	-	o	+	4	5	16
	NCU05891	arid/bright domain	+	+	+	13	19	18
	NCU00090	<i>pacc-1</i>	+	+	-	33	47	22
	NCU03277*	<i>pex10*</i>	o	o	o	13	45	51
	NCU03043	<i>acon-4 FlbC homolog</i>	o	o	+	103	34	30
	NCU03184	FlbC homolog	oo	ooo	o	25	247	35
	NCU06411	<i>vad-4</i>	o	+	o	176	97	47
Mild Defect	NCU00019	FKH1	o	+	o	13	301	126
	NCU06487	hypothetical	+	++	++	22	17	18
	NCU09068	<i>nit-2</i>	+	++	+	39	71	36
	NCU07039	<i>Asd-4</i>	+	++	++	46	26	76
	NCU02413	<i>rrg-2</i>	++	+	+	59	79	70
	NCU05242	C2H2 zinc finger	++	++	++	62	187	37
	NCU02671	<i>msn-1</i>	+	++	++	94	75	137
	NCU07952	C2H2 zinc finger	+	++	++	67	159	88
Weak/No Defect	NCU05207*	MFS transporter	++	++	++	46	41	99
	NCU07728	<i>sre</i>	++	++	++	80	52	90
	NCU00100	hypothetical	++	++	++	68	66	99
	NCU08317	hypothetical	++	++	oo	99	73	60
	NCU05024	hypothetical	++	++	++	68	65	105
	NCU07863	<i>vid21</i>	++	++	++	51	79	109
	NCU06763	hypothetical	++	++	++	42	107	109
	NCU09495	<i>set-6</i>	++	++	++	103	90	110
	NCU06145	RING-6	++	++	++	97	119	104
	NCU06920	hypothetical	++	++	++	106	118	107
	NCU07544	<i>upl-1</i>	++	++	++	332	13	33
	NCU09846	hypothetical	++	++	++	138	115	128
Increased Growth or Activity	NCU09033	RhaR ortholog	++	++	++	335	178	81
	NCU03206	C2H2 zinc finger	+++	++	++	223	193	111
	NCU01640	<i>rpn-4</i>	+++	++	++	106	57	123
	NCU08744	hypothetical	+++	++	++	467	285	62
	NCU09915	<i>ndt80</i>	+++	++	++	282	117	132
	NCU06173	hypothetical	++	++	+++	59	88	81

Key: Avi: growth on avicel, MM: growth on sucrose, Xyl: growth on xylan, Gluc: glucose release from Avicel, Cell: cellobiose release from Avicel, Xyln: reducing sugars released from xylan, ++ WT growth, +reduced growth, +++enhanced growth, o/oo/ooo reduced/WT/enhanced growth with morphological changes. *Annotation and/or gene model has changed since this screen was performed. No longer predicted to bind DNA.

Confirmation of defects in cellulase transcription with microarray analysis

To further confirm results from previous rounds and to differentiate transcription factors with a direct effect on transcriptional induction from those with more indirect effects on growth and enzyme secretion, I subjected the 29 strains with apparent growth and enzyme activity phenotypes on Avicel to a third round of screening by transcriptional profiling with DNA microarrays. I grew strains on sucrose and transferred them to Avicel for four hours to induce cellulase expression. I used a rapid, in-house, oligonucleotide microarray system [142] to compare cellulase transcription between the mutant strains on Avicel and the WT strain in the same condition.

Expression of the Avicel regulon (see Chapter 2) in the profiled mutants is shown in Figure 10. Initially, I analyzed a single biological replicate for each mutant. For mutants with a specific Avicel growth phenotype that also had a substantial (10-fold or more) decrease in transcript abundance for multiple known cellulases and/or hemicellulases, a second biological replicate was analyzed to confirm the phenotype. Considering the limits of this microarray system (a single hybridization sequence per gene and common loss of data from a significant number of genes due to optical noise) and the low number of biological replicates analyzed, these data are not sufficient to make conclusions on the expression of individual genes. Rather they serve as a direct and broad confirmation of a deficient response to cellulose.

Hierarchical clustering of arrays by the expression of the Avicel regulon identified two clusters of primary interest (array tree in red, Figure 10, right). A cluster of arrays from five of the mutants with severe growth defects (Δ NCU07705, Δ NCU08042, Δ NCU02017, Δ NCU03725, and NCU06407) had a severe deficiency in induction of the Avicel regulon — up to three orders of magnitude lower expression in the mutant on Avicel for some genes. A second cluster comprised of seven mutants mostly with pleiotropic defects (Δ NCU04459, Δ NCU05891, Δ NCU03043, Δ NCU03184, Δ NCU02671, Δ NCU03277, and Δ NCU01461) had generally lower cellulase expression than WT, but only by one order of magnitude or less. I interpreted these data as identifying a small set of mutants with strong, and likely direct, deficiencies in cellulose sensing; and another set of mutants with indirect effects on cellulase sensing or general defects in modulation of transcription. Other strains with Avicel growth and activity phenotypes might either have defects in transcription of a small number of genes that are critical for cellulase utilization (thus not amenable to confident discovery with this broad and rapid survey) or they may have transcriptional effects that then have secondary effects on post-transcriptional aspects of cellulase transcription and enzyme secretion.

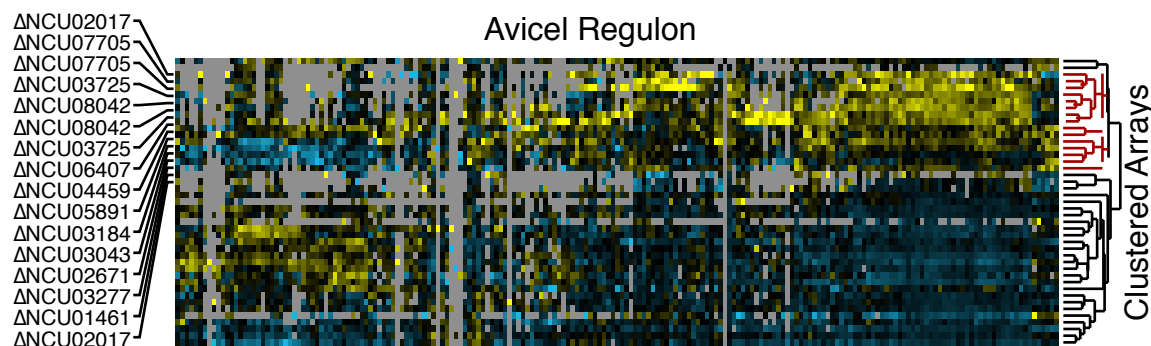


Figure 10. Heat map of hierarchical clusters of genes in the *N. crassa* Avicel regulon from microarray transcriptional profiles of transcription factor deletion strains switched to Avicel. Fold difference wild type on Avicel divided by mutant on Avicel were log transformed, normalized and centered before hierarchical clustering by average linkage, thus yellow (high) values represent genes with deficient transcriptional abundance on Avicel. Grey points represent genes with no data in that sample. Clusters of mutants with broad deficiency in cellulase induction are highlighted in red (right) and mutant strains listed (left).

Discussion

While this screen was broad and unbiased, it was far from a complete accounting of transcription factors with roles in cellulose degradation. Firstly, not all transcription factor deletion mutants are viable as homokaryotic lines in the deletion collection. Secondly, the first round of this screen, a simple and qualitative assessment of growth in a single condition, was sensitive only to mutants with particularly strong phenotypes. In fact, previous experiments with deletion mutants for major cellulases in *N. crassa* would suggest that in most cases even completely abolishing the activity of a single cellulase enzyme has a relatively mild effect on overall growth and cellulase depolymerization on Avicel [143]. A screen that depends on visible growth differences between strains can only identify mutants with deficient expression of several cellulases or at least the few most critical cellulases. Some regulators may induce only a few genes in response to a particular molecular signal. Finally, even if such mutants with a small regulon of deficient induction had sufficient growth defects as to be identified in initial rounds of screening, the limitations of the final transcriptional profiling applied made identifying mutants with only a few deficiently induced genes unlikely. This was not an authoritative and complete survey of transcriptional regulation in response to cellulose and likely much remains to be discovered.

Such limitations aside, this screen successfully identified approximately two-dozen poorly characterized transcription factors with a spectrum of effects for growth and activity on Avicel. Although strains containing deletions of genes involved in plant cell wall degradation in *T. reesei* and *Aspergilli* were present in the collection, including homologs to *xlnR* [81] (NCU06971), *ace1* [85] (NCU09333) and *hap2* [144] (NCU03033), in *N. crassa* these mutants showed no visual growth defects on Avicel. Other deletion strains for transcription factors with a known influence on cellulase production were identified, including *nit-2* (NCU09068, a homolog to the nitrogen regulator *areA*) [145], *pacC* (NCU00090, pH sensing) [146] and *cre-1* (NCU08807, carbon catabolite repression [147]). These strains had relatively modest changes in cellulase activity on Avicel.

Some deletion mutants with major pleiotropic defects were particularly deficient at growth on Avicel. One would assume that radically shifting metabolic state and protein expression profiles from rapid growth on sucrose media to a demanding near starvation state on Avicel requires a robust mycelial colony. Four pleiotropic mutants with attenuated, but not eliminated transcriptional responses to Avicel were homologous to proteins with known or suspected roles in vegetative development or the cell cycle (NCU03043, NCU03184, NCU06411, and NCU01461). Thus at this stage it is unclear if these transcription factors directly impact cellulose sensing or if they are simply required for a colony healthy enough to mount a robust transcriptional response to environmental stimuli.

Another transcription factor necessary for full cellulase induction was NCU04459, which has homology to a component of the SAGA complex, a conserved eukaryotic multi-protein complex with diverse roles in transcriptional activation [148]. Among the SAGA complex's many proposed functions in *S. cerevisiae* is the tethering of galactose-regulated loci to the nuclear pore complex upon transcriptional activation [149] and may play a role in rapid repression of those genes [150]. What role such localization plays in cellulase gene regulation in filamentous fungi is unknown.

A deletion mutant for NCU05891, ortholog of *S. cerevisiae* Swi1, also had attenuated, but not eliminated cellulase induction. The Swi/Snf complex is required for transcriptional activation of several *S. cerevisiae* genes including some enzymes required for alternative metabolism [151]. In their review on the Swi/Snf complex [152], Sudarsanam and Winston noted that many inducible *S. cerevisiae* genes have overlapping regulation by Swi/Snf and the SAGA complex. This theme of layered regulation by Swi/Snf and SAGA may extend to cellulase genes in filamentous fungi.

One deletion mutant, ΔNCU06407, had a severe and seemingly specific growth and cellulase induction deficiency on Avicel, but when I crossed the strain back to the WT genomic background, the loss of cellulase activity phenotype did not segregate with the hygromycin resistance encoded in the deletion cassette (data not shown). I thus abandoned that strain as an immediate target for further analysis as the source of the phenotype may be an additional mutation or mutations not directly associated with the gene deletion. Bulk segregant analysis techniques discussed in Chapter 6 may be useful in ultimately identifying the source of the mutant phenotype, but as of this writing it has not been explored further.

Another severe mutant (ΔNCU03725) was a deletion for *vib-1* (vegetative incompatibility blocked 1), a putative transcription factor required for regulated cell death after fusion of genetically incompatible hyphae [153]. Since this screen was conducted, Xiong et al investigated the role of *vib-1* in cellulase regulation and proposed that it may be required for effective relief of carbon catabolite repression upon exit from growth on preferred carbon sources [107].

Of the remaining three strains with severe defects in cellulase induction, two stood out. The deletion strains for the conserved hypothetical proteins NCU07705 and NCU08042 had the most severe and consistent cellulase growth defects of any strains tested. NCU07705 had been identified as strongly induced by growth on *Miscanthus* in a previous study [44]. Both NCU07705 and NCU08042 belong to a conserved fungal family of transcriptional regulators with a DNA binding domain

based on a six-cysteine motif which coordinates two zinc ions to interface with polar residues on DNA and are thus called zinc binuclear cluster proteins (ZNBCs). Many ZNBCs are thought to regulate genes required for the regulation of alternative carbon sources [90], for example the prototypical family member *S. cerevisiae* Gal4 [154] and the *T. reesei* cellulase regulator Xyr1 [87]. The strong, cellulose-specific phenotype exhibited by the NCU07705 and NCU08042 deletion strains, together with their homology to known regulators of alternative carbon metabolism, strongly suggested these two unknown transcription factors play a central role in cellulose sensing in *N. crassa*. I provisionally named them *clr-1* and *clr-2*, respectively, for cellulose degradation regulator 1 and 2. Deeper characterization of *clr-1* and *clr-2* follows in the next chapters.

NCU02017 also remains an interesting gene for further analysis. Though I crudely categorized this mutant as having a relatively specific phenotype, it does have a conidiation defect suggesting broader effects than cellulase regulation exclusively. Some broader phenotypes might be expected given that the *S. cerevisiae* ortholog of NCU02017, *NCB2*, is an essential gene and component of the NC2 complex, which interfaces with basal transcription factors and modulates expression of a large number of genes [155]. The ortholog of NCU02017 in *T. reesei* was found to contain a non-synonymous point mutation in the hyper-secreting strain Rut-C30 [61], although the effect of this individual mutation amongst hundreds of other mutations is not understood. With the number of mutations that have accumulated in strain Rut-C30 over decades of mutagenesis and selection, it is very possible that this particular mutation is merely coincidental and does not contribute to cellulase hyper-secretion, but it does suggest the possibility of a conserved role for this transcription factor across species. Further investigations into NCU02017's function in *N. crassa* continue, but remain preliminary and are not be addressed in this dissertation.

Acknowledgements

Chaoguang Tian first identified NCU07705 as gene of interest in transcriptional profiling experiments of *N. crassa* grown on *Miscanthus*.

Terry Shock performed the initial selection of putative transcription factors, first round of screening and selection of candidate strains for further screening.

Published material

A brief summary of the transcription factor screen, without details on preliminary transcriptional profiling or the identification of novel cellulase regulators besides CLR-1 and CLR-2 was initially published in [130]:

Coradetti ST, Craig JP, Xiong Y, Shock T, Tian C, et al. (2012) Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *Proc Natl Acad Sci U S A* 109: 7397-7402.

Methods

Strains and culture conditions

Proteins with predicted DNA-binding domains were selected from the *N. crassa* genome. Deletion strains for 269 of these possible transcription factors were available in the *N. crassa* deletion collection [131,156] from the Fungal Genetics Stock Center (FGSC) [157] and were successfully revived from frozen stocks. Strains were grown on VMM sucrose slants [132] for ~ 7 days. Conidia from each deletion strain were harvested in water and inoculated into a 10 mL tube containing 2% Avicel (Avicel® PH-101, Sigma Aldrich, MO) media to a final concentration of 1×10^6 conidia per mL. Any deletion strains showing reduced growth over a 7 day time period by visual inspection were repeated for confirmation. This initial screen yielded 42 candidate strains for further characterization. Candidate strains were again inoculated at 1×10^6 conidia per mL in Vogel's minimal media (VMM) with 2% sucrose, xylan (beechwood xylan, Sigma Aldrich) and Avicel. At 2 days (sucrose) or 6 days (Avicel and xylan) after inoculation, supernatants were harvested and assayed for enzyme activity.

Enzyme activity assays

Cellulase assays were performed as described in Tian *et. al.*, [143]. In brief, one volume of 7-day culture supernatant was mixed with one volume of substrate solution containing 5 mg/ml Avicel and 100 mM sodium acetate buffer, pH 5.0 at 37°C for 5 hrs with shaking. Then, glucose and cellobiose concentration were measured by coupled enzyme assays. Similarly, xylanase activity was assayed by mixing one volume of culture supernatant with 9 volumes of substrate solution containing 10mg/mL beechwood xylan and 50 mM sodium acetate, pH 5.0 at 50C for 5 minutes. Released reducing sugars were then measured with the DNS method [158].

Transcriptional profiling with oligonucleotide microarrays

N. crassa cultures were grown for 16 hrs on VMM sucrose and washed with VMM without added carbon. Mycelia were re-suspended in 100 mL VMM with 2% Avicel. At 4 hours after the shift, mycelia were harvested by filtering on Whatman (#1) filter paper and flash frozen in liquid nitrogen. Total RNA was extracted with TRIzol reagent (Life Technologies) as described in [44] and further purified with RNeasy kits (Qiagen). ChipShot™ Indirect Labeling/Clean-Up System (Promega, catalog no. Z4000) and CyDye Post-Labeling Reactive Dye Pack (GE, catalog no. RPN5661) were used to synthesize and label cDNA according to the manufacturer's instruction, except that 10µg of total RNA was used as the starting material. The Pronto! Hybridization Kit (Corning, catalog no. 40076) was used for microarray hybridization according to the manufacturer's specifications. For deletion mutants with biological replicates, a dye-swap used to control for differential hybridization the C3 and C5 dyes to different samples. Data analyses were performed as previously described[142]. A GenePix 4000B scanner (Axon Instruments) was used to acquire images, and GenePix Pro6 software was used to quantify hybridization signals and collect the raw data.

Hierarchical clustering

For each experiment, normalized intensity for each gene in the WT Avicel sample was compared to the mutant sample. These WT/mut ratios were hierarchically clustered with the Cluster 3.0 software suite [159]. Prior to clustering, FPKMs were log transformed, normalized across strains/conditions and centered on the geometric mean across strains/conditions on a per-gene basis. The average linkage method was used for cluster generation, with uncentered correlation as the similarity metric. Clusters were visualized with Javatreeview [160].

Chapter 4

Cellulose Degradation Regulators CLR-1 and CLR-2

Abstract

A screen of the *Neurospora crassa* transcription factor deletion collection identified two uncharacterized zinc binuclear cluster transcription factors (*clr-1* and *clr-2*) that were required for growth and enzymatic activity on cellulose, but were not required for growth or hemicellulase activity on xylan. Transcriptional profiling demonstrated that *clr-1* and *clr-2* were required for the bulk of that *N. crassa* transcriptional response to cellulose including induction of all major cellulase and some major hemicellulase genes. Functional CLR-1 was necessary for expression of *clr-2* and efficient cellobiose utilization. Constitutive expression of CLR-2 was sufficient for cellulase secretion and activity, but not CLR-1, implying a requirement for post-translational modification of CLR-1 as the terminal output of a cellobiose sensing system.

Introduction

An unbiased screen of putative transcription factors in the *N. crassa* deletion collection [161] identified several transcription factors necessary for robust induction of cellulases in response to cellulose (Avicel). Two transcription factors in particular, *clr-1* (NCU07705) and *clr-2* (NCU08042) were essential for cellulase expression on Avicel. To better understand environmental cellulose sensing in filamentous fungi and to enable rational engineering of industrial enzyme production strains, I further characterized these transcription factors and their roles in a cellulose-sensing pathway.

CLR-1 and CLR-2 are essential regulators of cellulose degradation in *N. crassa*

I cultured deletion strains for *clr-1* and *clr-2* (Δ *clr-1* and Δ *clr-2*) from conidia for 16 hr on sucrose and then transferred them to Avicel for 24 to 48 hrs. The mutant strains showed no detectable cellulase activity when assayed for cleavage of carboxymethyl cellulose (CMC) (Figure 11A). The Δ *clr-1* and Δ *clr-2* mutants secreted approximately 5% of total protein as compared to wild type (WT) strains under cellulolytic conditions (Figure 11C) and displayed only trace levels of xylanase activity. However, when I grew the Δ *clr-1* and Δ *clr-2* mutants on sucrose for 16 hr and then switched them to xylan, they displayed near WT hemicellulase activity and WT levels of protein secretion (Figure 11C) as well as similar banding patterns on SDS-PAGE gels (Figure 12). The loss of cellulase induction in response to Avicel segregated with the hygromycin marker used to disrupt NCU07705 and NCU08042 (data not shown) and was complemented by introduction of a WT copy of NCU07705 or NCU08042 at the *his-3* locus (Figure 13). Neither WT strains nor the *clr* mutants showed cellulolytic activity when exposed to hemicellulosic substrates (Figure 11A).

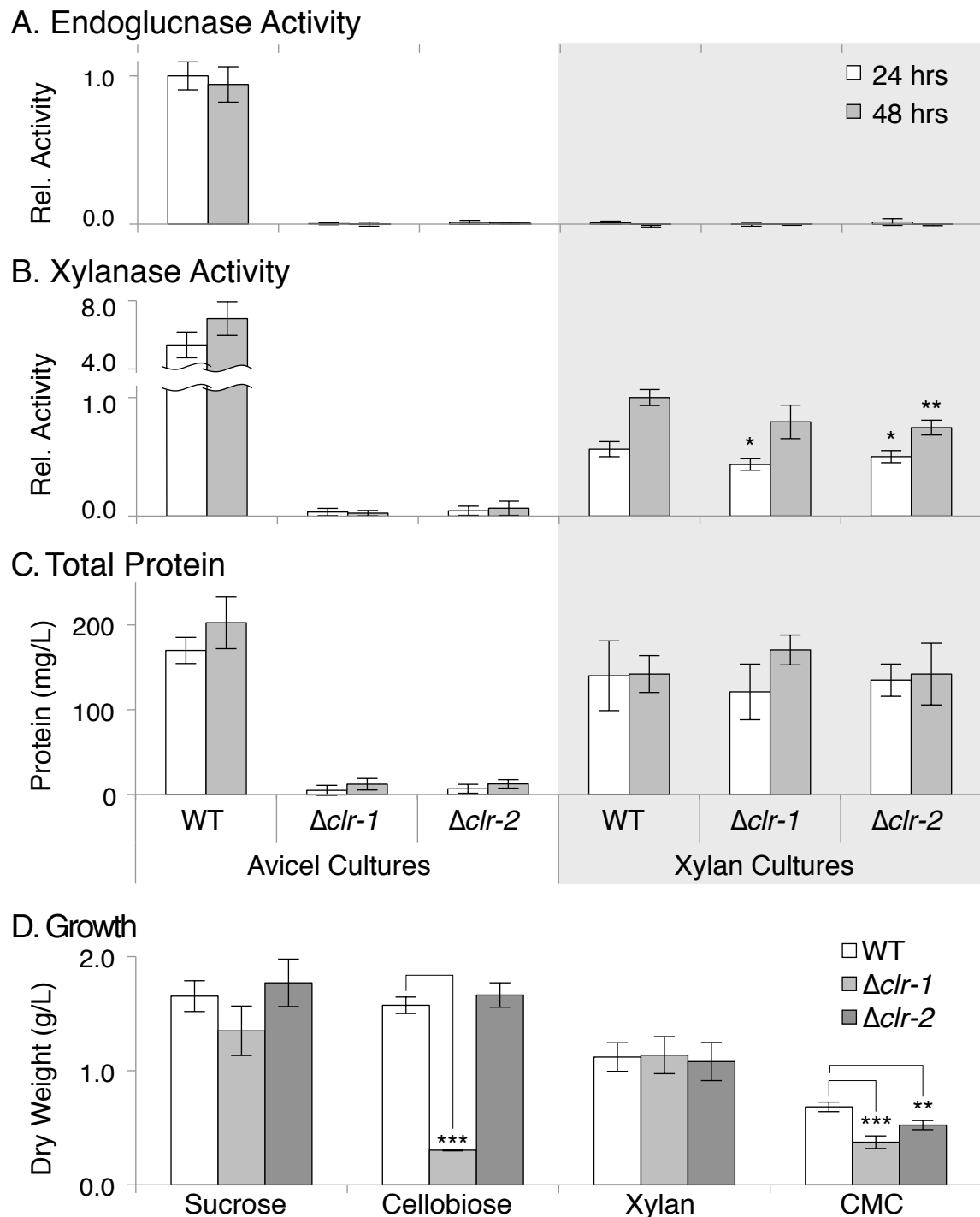


Figure 11. Growth and activity phenotype of *clr* deletion strains.

(A-C) Endoglucanase activity, xylanase activity and total protein in supernatants from WT and *clr* mutant cultures grown on sucrose for 16 hr and subsequently transferred to Avicel or xylan for an additional 24 to 48 hrs.

(D) Mycelia dry weights from cultures on soluble carbon sources (0.5% w/v except for CMC, which was 2%).

The $\Delta clr-1$ and $\Delta clr-2$ mutants showed an identical phenotype when grown on sucrose (WT growth), xylan (WT growth) and cellulose (no growth). The $\Delta clr-1$ mutant was deficient for growth on cellobiose, yet the $\Delta clr-2$ mutant grew similarly to WT on cellobiose (Figure 11D). These data suggested that CLR-1 and CLR-2 may differentially regulate gene sets associated with plant cell wall degradation, in particular genes encoding β -glucosidase enzymes, which are involved in hydrolysis of cellobiose to glucose (Figure 1).

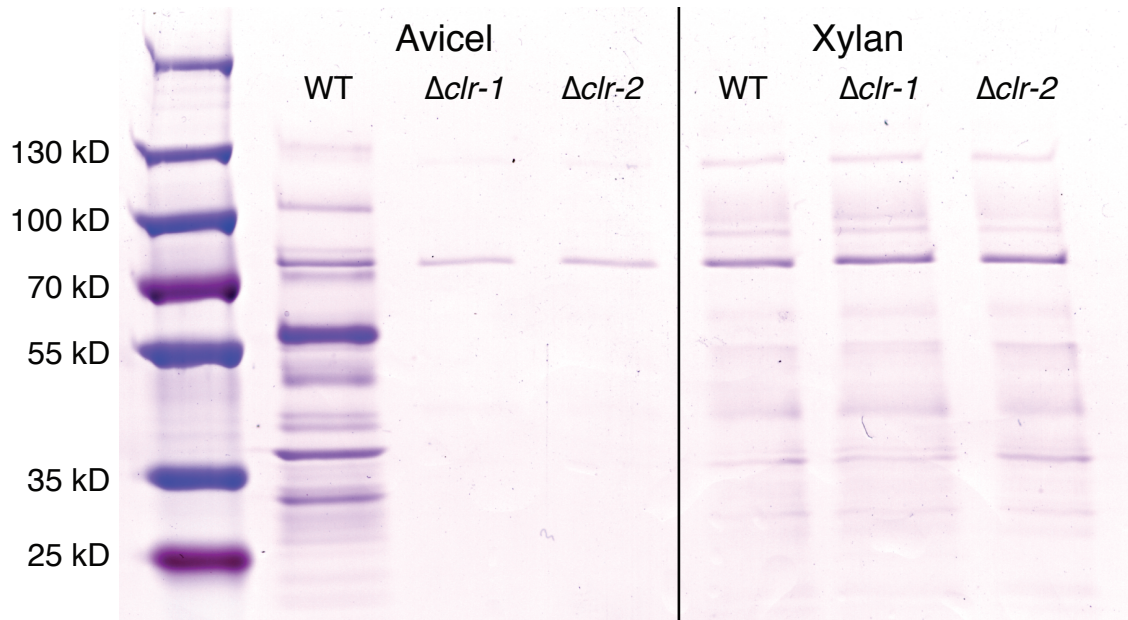


Figure 12. Secreted proteins in $\Delta clr-1$ and $\Delta clr-2$ culture supernatants

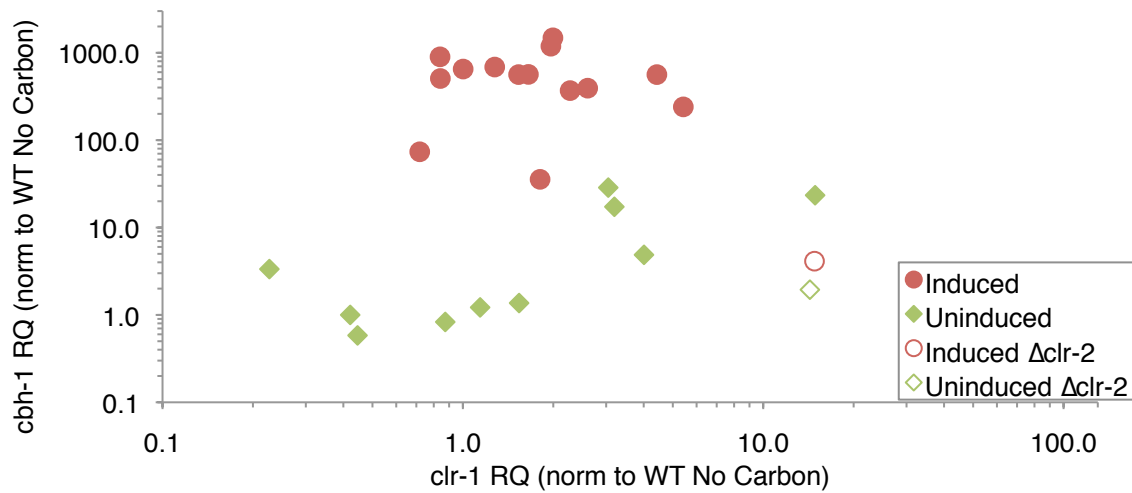
SDS-PAGE of WT, $\Delta clr-1$ and $\Delta clr-2$ culture supernatants 48 hrs after shift to Avicel or xylan media from a 16 hr sucrose-grown culture. Equal volumes loaded. The broad spectrum of proteins secreted by WT *N. crassa* on Avicel was absent in the *clr* deletion mutants. There was a conspicuous band at approximately 75 kDa, which corresponds to the size of NCW-1 (NCU05137), identified in [44] as secreted when WT was grown on either Avicel or *Miscanthus*. The secretome of WT and *clr* deletion strains on xylan was identical.

Inducer-independent expression of cellulases in *N. crassa*

I performed a series of experiments to test if expression of one or both transcription factors was sufficient for cellulase expression in non-inducing conditions. I utilized a strain constructed by Jianping Sun with GFP tagged *clr-1* under control of the *ccg-1* promoter at the *his-3* locus (strain *Pccg-1:clr-1-GFP*). Under culture conditions of vegetative growth in constant light and temperature, as in these experiments, circadian rhythm is dampened [162] and *Pccg-1* is essentially a constitutive promoter. I constructed a similar strain with untagged *clr-2* also under regulation of the *ccg-1* promoter at the *his-3* locus (strain *Pccg-1:clr-2*). By crossing these strains with the deletion mutants and/or forming heterokaryons with the WT or *clr-1/2* deletion strains I created a line of strains with a range of effective copy numbers of *clr-1* and *clr-2* (materials and methods). I used these strains to test transcription of the major cellulase *cbh-1* (NCU07340) as a function of

clr-1 or *clr-2* transcription. Increasing *clr-1* abundance in non-inducing conditions was insufficient to recapitulate induced levels of *cbh-1* (Figure 13A). A weak induction of *cbh-1* was achieved in the most highly *clr-1* expressing strains, but that induction required a functional copy of *clr-2*. Regardless of *clr-1* abundance, inducing conditions (a shift to Avicel) was required to achieve high *cbh-1* expression. However, constitutive expression of *clr-2* under non-inducing conditions was sufficient to drive Avicel-like abundance of *cbh-1* (Figure 13B).

A. *clr-1* vs *cbh-1* Transcript



B. *clr-2* vs *cbh-1* Transcript

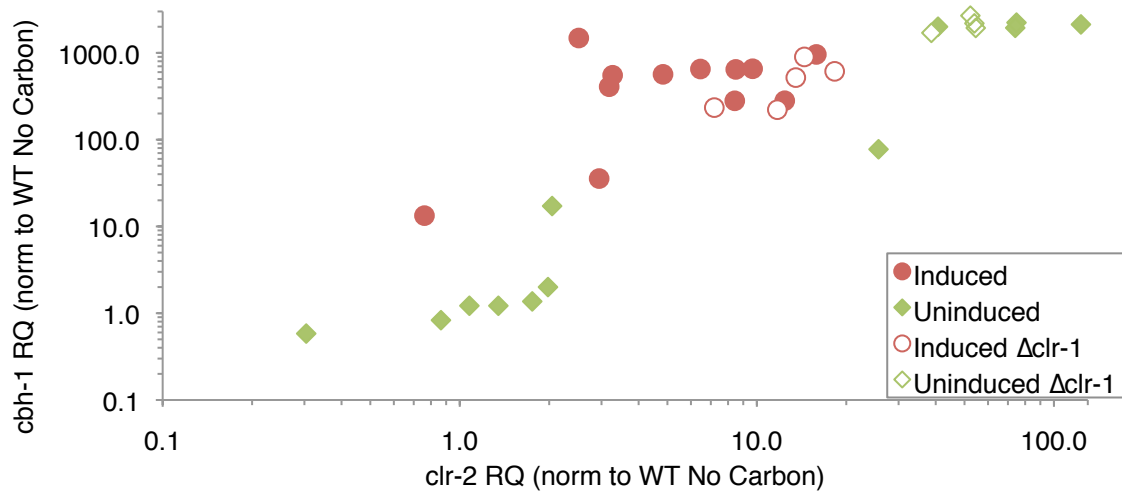


Figure 13. Cellulase transcription as a function of *clr-1* and *clr-2* transcription

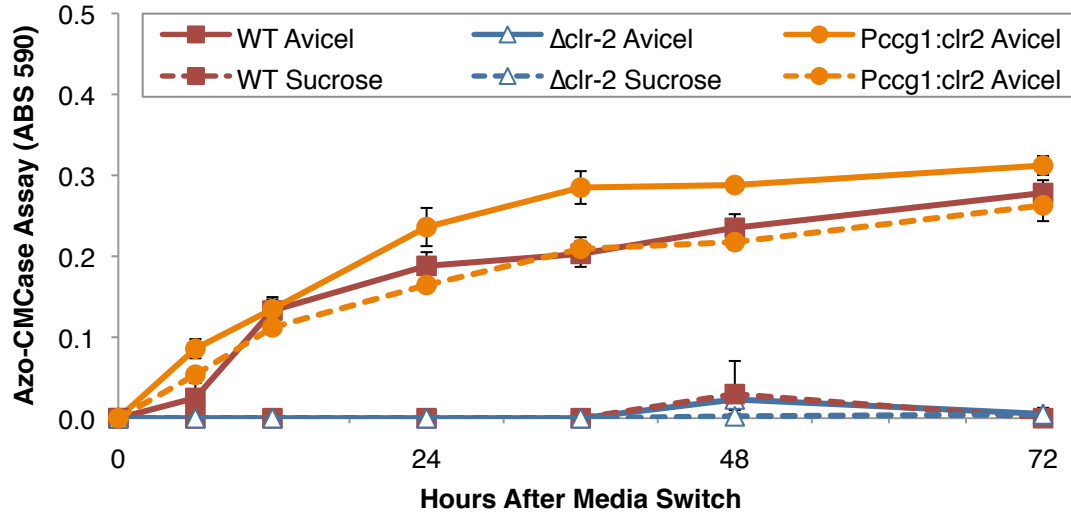
(A) Relative expression of the major cellulase *cbh-1* (*gh7-1*) in response to manipulated *clr-1* expression.

(B) Relative expression of *cbh-1* in response to manipulated *clr-2* expression.

Transcript abundance by RT-PCR. Relative abundance to actin (*act* NCU04173) for each gene was normalized to the WT strain on no carbon media.

To test whether *clr-2* expression alone was sufficient for cellulase enzyme secretion and activity in *N. crassa* cultures, I shifted cultures of WT, Δ *clr-2* and *Pccg-1:clr-2* strains to Avicel, sucrose and no carbon media. Endoglucanase activity was similar in the *Pccg-1:clr-2* sucrose cultures to the WT on Avicel (Figure 14A). The profile of secreted proteins in the *Pccg-1:clr-2* sucrose and no carbon culture supernatants were also similar to those of WT strains on Avicel at 24 hours after a media shift (Figure 14B), suggesting a broad induction of the Avicel regulon.

A. Endoglucanase Activity



B. SDS-PAGE

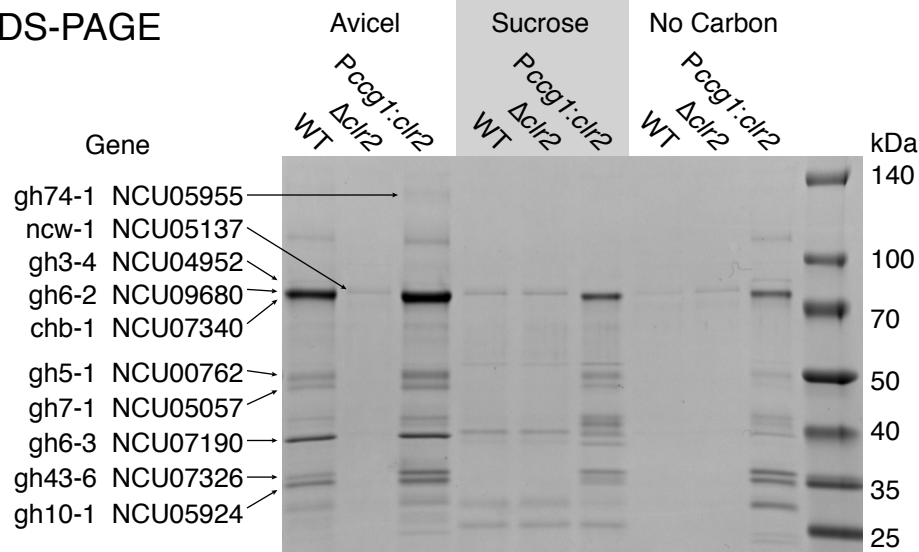


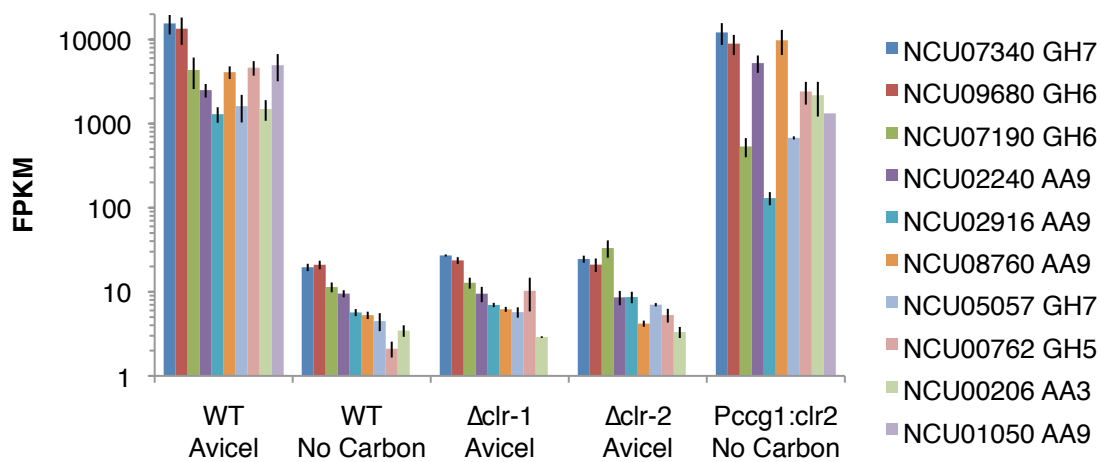
Figure 14. Enzyme activity and protein secretion in the *Pccg-1:clr-2* mutant

(A) A time course of Endoglucanase activity of supernatants from WT, Δ *clr-2* and *Pccg-1:clr-2* cultures after shift from sucrose pre-grown cultures (16 hrs) to either Avicel or sucrose. (B) Coomassie stained SDS PAGE gel of culture supernatants from WT, Δ *clr-2* (Δ) and *Pccg-1:clr-2* strains shifted to Avicel (4 days), sucrose (2 days) or media lacking a carbon source (2 days). Secreted proteins labeled from molecular weights and intensity patterns are derived from [127,128,143]. Note similarity of secreted protein profile of the *Pccg-1:clr-2* strain grown on sucrose or no carbon as compared to Avicel-exposed WT and *Pccg-1:clr-2* cultures.

Transcriptional profiles of *clr* deletion mutants.

To compare expression of the Avicel regulon in the WT and deletion strains, I subjected the $\Delta clr-1$ and $\Delta clr-2$ mutants, as well as the *Pccg-1:clr-2* mutant, to transcriptional profiling with RNA-Seq (Figure 15, Figure 16). At 4 hrs after transfer to Avicel media, both deletion strains had transcriptional profiles more similar WT no carbon samples than to WT Avicel samples. Both mutants failed to induce transcription of all major cellulases (Figure 15A) and five of the ten most highly expressed hemicellulase genes (Figure 15B). At 4 hours after transfer to no carbon media, the *Pccg-1:clr-2* mutant had a WT Avicel like profile, with strong induction of all major cellulases and 6 major hemicellulases (Figure 15).

A. Cellulases



B. Hemicellulases

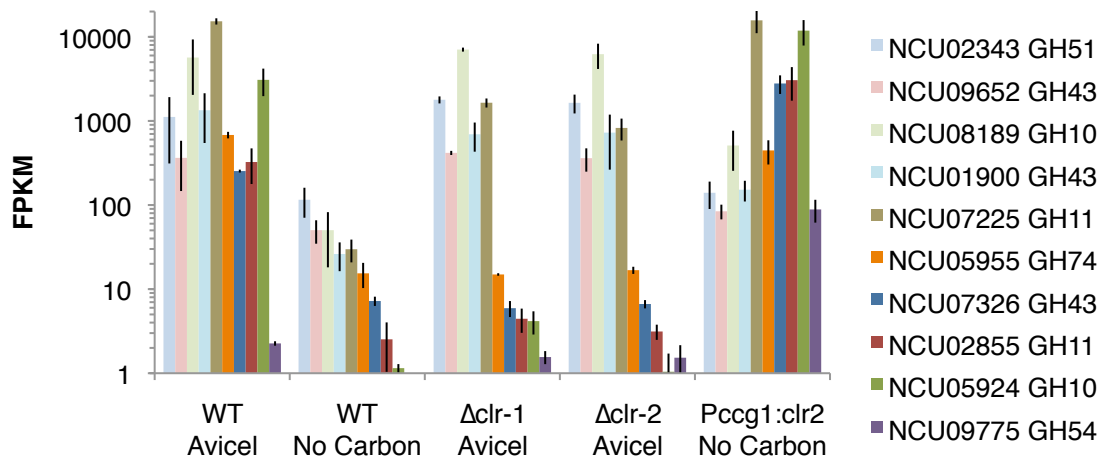
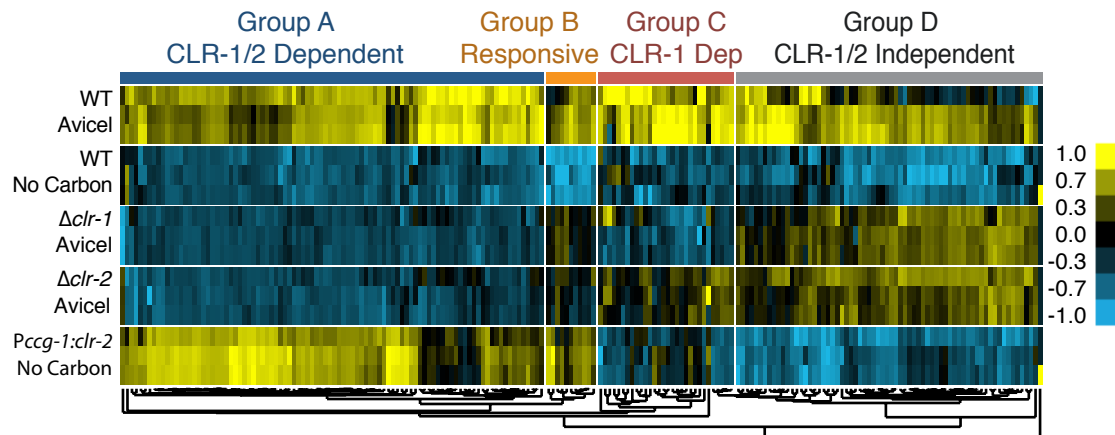


Figure 15. Transcript abundance for major cellulases and hemicellulases in *clr-1* and *clr-2* deletion mutants

FPKM of 10 cellulase genes (A) and 10 hemicellulase genes (B) in WT and *clr* mutant strains when exposed to Avicel for 4 hrs.

Hierarchical clustering of expression patterns for 203 genes specifically induced on Avicel (genes identified in Figure 8, with 9 of the 212 genes removed due to recent changes in gene models from the *N. crassa* genome version 10 (used for the analysis in Chapter 2) to version 12, all with relatively low absolute expression) revealed four major groups of genes with similar expression patterns. A summary of the Avicel regulon and dependence of major categories on *clr-1* and/or *clr-2* is presented in Figure 16.

A. Hierarchical Clustering of Avicel Induced Genes



B. Gene Function and CLR-1/2 Dependence

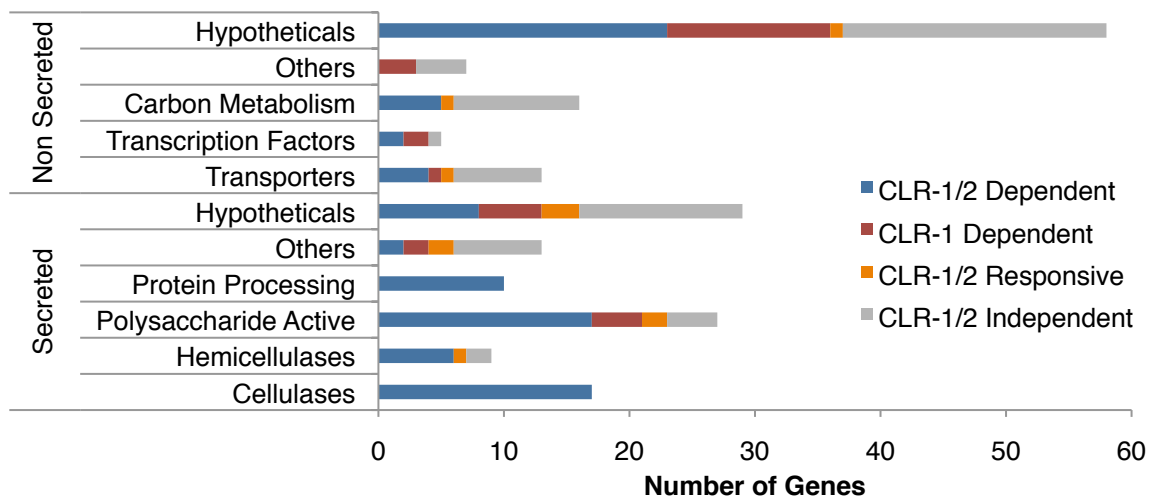


Figure 16. The *clr* regulon.

(A) Expression patterns for 203 Avicel-induced in WT and the *clr* mutants cluster in 4 major groups..
 (B) Distribution of Avicel-induced genes in selected functional categories [123].

Group A (CLR-1/2 dependent) consisted of 94 genes that showed no induction in both $\Delta clr-1$ and $\Delta clr-2$ mutants and were strongly induced in the *Pccg-1:clr-2* mutant on no carbon media. This group included 16 cellulases, 6 hemicellulases, 17 other genes encoding enzymes with activity on polysaccharides and 4 other enzymes predicted to enter the secretory pathway. It also includes 10 proteins with likely function in protein secretion including components of the ER

translocation complex, disulfide isomerizes, and ER chaperones (HSP70 and calnexin). The genes in group A were as whole the most strongly induced cluster in the Avicel regulon.

Group B (CLR-1/2 responsive) consisted of 11 genes with reduced, but not eliminated induction in both deletion mutants and strong induction in the *Pccg-1:clr-2* mutant. This group included a major hemicellulase (NCU07225, GH11), two secreted lipases (NCU03322 and NCU04475) and two acetyl xylan esterases (NCU04870 and NCU05159).

Group C (CLR-1 dependent) consisted of 30 genes with no carbon-like expression in the Δ *clr-1* mutant on Avicel, generally lower (but not abolished) induction in the Δ *clr-2* mutant, and weak or no induction in the *Pccg-1:clr-2* mutant. This was dominated by 18 hypothetical proteins but also included the major cell wall associated β -glucosidase NCU08755, the xylanase regulator *xlr-1* (NCU06791) and *vib-1* (NCU03725) a regulator involved in programmed cell death but also regulation of carbon catabolite repression[107].

Group D (CLR-1/2 independent) consisted of 94 genes with substantial induction in both deletion mutants on Avicel and no carbon like expression in the *Pccg-1:clr-2* mutant. Some of these genes did have significantly lower expression in the deletion mutants, but they still had substantially greater transcript abundance than on no carbon. This cluster included two major hemicellulase genes and several genes with predicted function related to xylose or other pentose utilization as well as NCU08114, a high affinity cellodextrin and xylodextrin transporter [46,51]. It included no major cellulases.

Given that the Δ *clr-1* mutant grew poorly on cellobiose (Figure 11) and that cellobiose is likely the signaling molecule inducing cellulase expression (or the source of a trans-glycosylation product that is the inducer) [128], expression of genes required for cellobiose and cellobionic acid utilization was of particular interest. Figure 17 shows transcript abundance for three major β -glucosidases (*gh1-1* NCU00130, *gh3-3* NCU08755 and *gh3-4* NCU04952), a cellobionic acid phosphorylase (NCU09425) [54], and three cellodextrin/cellobionic acid transporters (*cdt-1* NCU00801, *cdt-2* NCU08114, *cbt-1* NCU05853) [163,164]. In general, cellobiose utilization genes are more highly expressed in starvation conditions than cellulases and more weakly induced on Avicel (compare Figure 15 and Figure 17). Five major cellobiose/cellobionic acid utilization genes were more strictly reliant on *clr-1* than *clr-2*, with higher expression in the Δ *clr-2* mutant than the Δ *clr-1* mutant and lower expression in the *Pccg-1:clr-2* mutant than the WT strain on Avicel. Two genes — one major β -glucosidase (*gh3-4* NCU04952, the major secreted β -glucosidase) and one cellodextrin transporter (*cdt-1* NCU00801) — were relatively more responsive to *Pccg-1:clr-2* and thus have a more cellulase-like expression pattern. These data suggest the former genes are more critical for cellobiose utilization whereas the later may be more critical for efficient cellulose degradation than cellobiose utilization. Most likely NCU04952 and NCU00801 help reduce local cellobiose concentrations and thus relieve product inhibition on cellobiohydrolases [165].

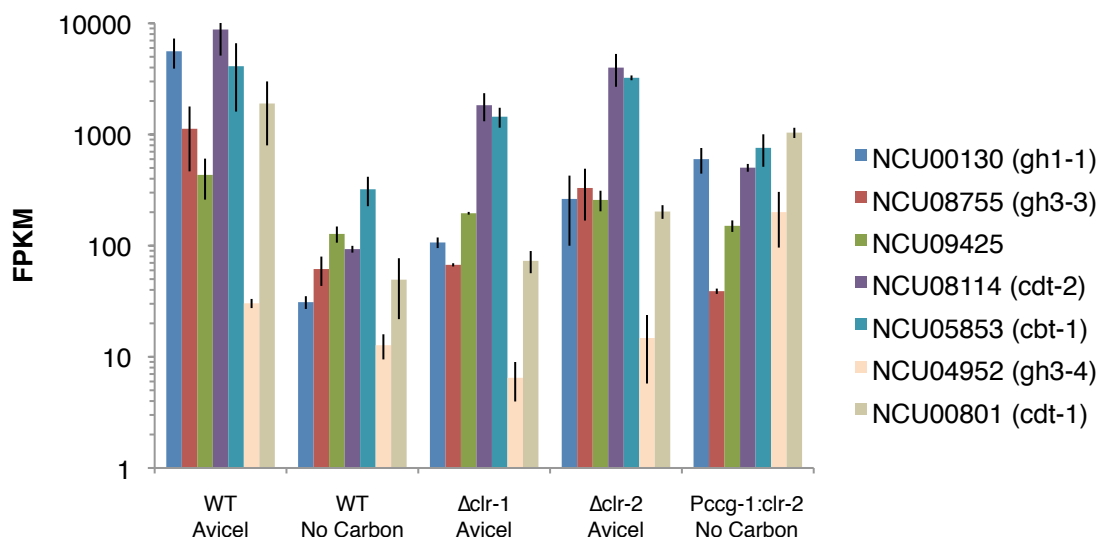


Figure 17. Transcript abundance for cellobiose and cellobionic acid utilization genes
 FPKM at 4 hours after media shift. Dark bars: genes with relatively high expression in the Δ clr-2 mutant on Avicel. Light bars: genes with relatively high expression in Pccg-1:clr-2.

Cellulase induction by cellobiose requires CLR-1 and CLR-2

Given that it had been recently shown that cellobiose could induce the major component of the Avicel regulon (if β -glucosidase activity was drastically reduced by deleting three major β -glucosidase genes) [128], my results showing that *clr-1* and *clr-2* were critical for cellulase induction by Avicel immediately raised the question: are *clr-1* and *clr-2* also critical for induction by cellodextrins? To address this question, I collaborated with Elizabeth Znameroski and Jordan Tsai, who crossed the *clr-1* and *clr-2* deletions into their triple β -glucosidase deletion strain (Δ β G strain). I then transferred sucrose grown cultures to 0.2% cellobiose and probed cellulase expression with RT-PCR. Induction of the major cellulase *cbh-1* (NCU07340) by cellobiose was abolished in the Δ β G Δ clr-1 and Δ β G Δ clr-2 strains (Figure 18). Furthermore, the cellodextrin and xylodextrin transporter *cdt-2* NCU08114 was induced by cellobiose in the Δ β G and Δ β G Δ clr-2 strains, but not in the Δ β G Δ clr-1 strain, both confirming independent induction of this transporter by cellobiose and confirming that this induction was dependent on *clr-1* but not *clr-2*.

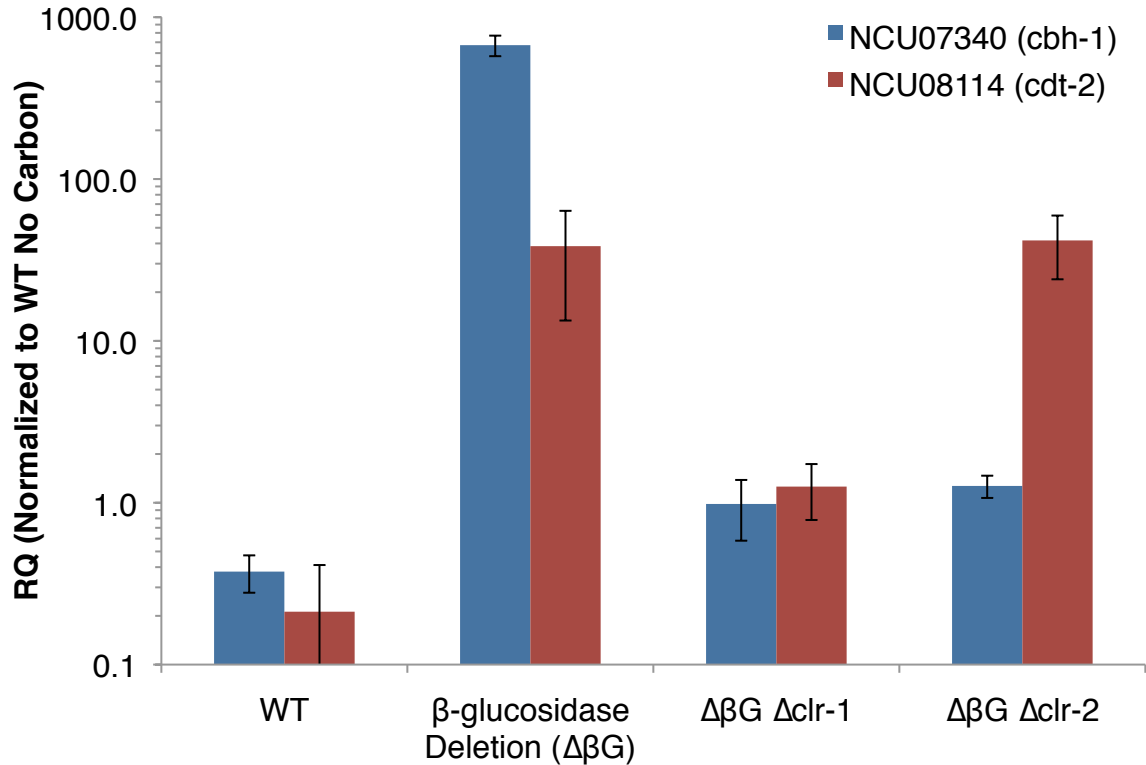


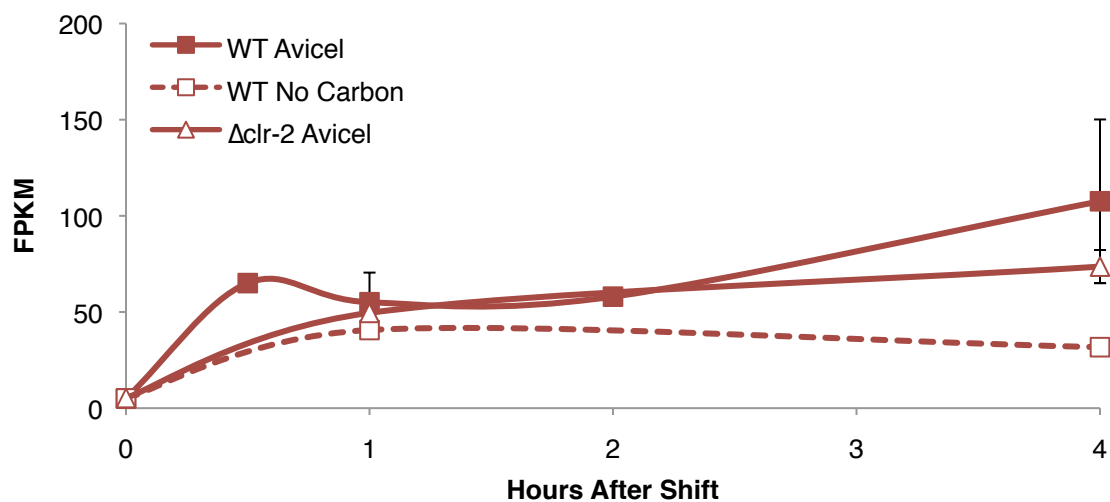
Figure 18. Cellulase induction by cellobiose in $\Delta clr-1$ and $\Delta clr-2$ mutants

Abundance of *cbh-1* and *cdt-2* transcript relative to *act* after switch to cellobiose. Transcript abundance by RT-PCR. All samples normalized to WT switched to no carbon.

Transcriptional regulation of *clr-2* by *clr-1*.

When wild type *N. crassa* was transferred to Avicel, *clr-1* and *clr-2* initially showed an expression level similar to no carbon controls (Figure 19). Between two and four hours after a shift to Avicel, both *clr-1* and *clr-2* transcripts became significantly more abundant. While *clr-1* was more highly expressed in the early de-repressed phase, *clr-2* was more abundant in the later inductive phase. In the $\Delta clr-1$ mutant, expression of *clr-2* was effectively abolished. Expression levels of *clr-1* were unaffected in the $\Delta clr-2$ mutant. These data suggest that *clr-2* was under transcriptional regulation by CLR-1 and that the relative abundance of *clr-1* transcript under de-repressed conditions was insufficient for induction of *clr-2* expression, just as forced transcription of *clr-1* was insufficient for *cbh-1* induction.

A. *clr-1* Transcript Abundance



B. *clr-2* Transcript Abundance

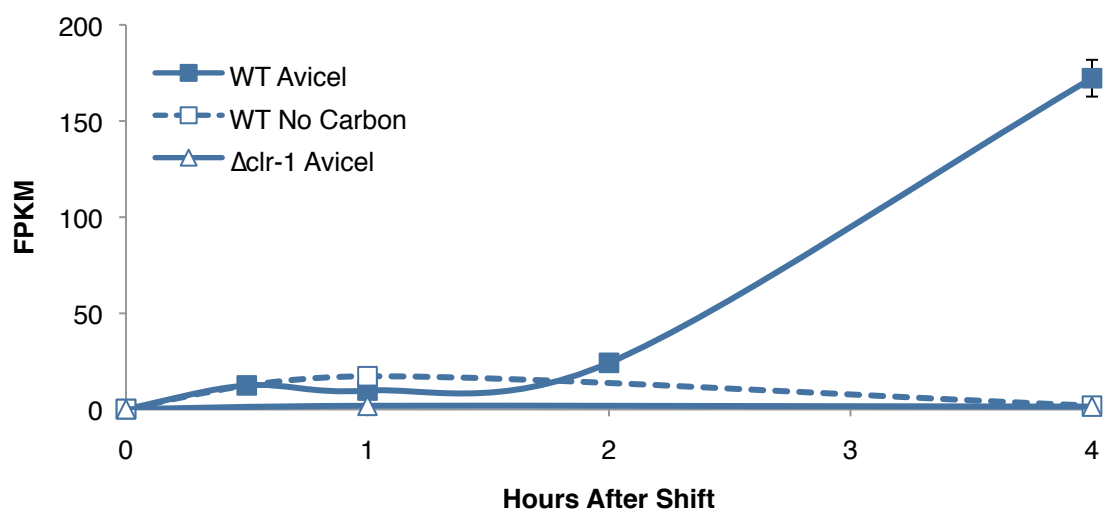


Figure 19. Transcript abundance of *clr-1* and *clr-2* after a shift to Avicel or no carbon. (A) FPKM for *clr-1* in WT *N. crassa* transferred to Avicel or no carbon media and for the *clr-2* deletion mutant. (B) FPKM for *clr-2* in WT *N. crassa* transferred to Avicel or no carbon media and for the *clr-1* deletion mutant.

Discussion

By utilizing the well-developed genomic resources for *N. crassa*, I identified two novel and conserved transcription factors that are key transcriptional regulators of cellulase gene expression. Despite the complete loss of induced cellulase gene expression and cellulase secretion in $\Delta clr-1$ and $\Delta clr-2$ strains, growth and hemicellulase secretion were normal on xylan. These results were in contrast to the function of XlnR/XYR1 in *A. niger* and *T. reesei*, where this transcription factor was necessary for regulation of all cellulase and hemicellulase genes [27,124]. However, an XlnR/XYR1 homolog was not necessary for cellulase gene expression or activity in *N. crassa* or *Fusarium* species 89 [97,140]. In those species, XlnR/XYR1 homologs regulated the expression of hemicellulase genes and were required for utilization of hemicellulose. These data suggest that some fungal species have evolved independent mechanisms for expression of cellulase and hemicellulases genes in response to distinct inducers from cellulose or hemicellulose. CLR-1 and CLR-2 are apparently components of such a specific pathway for cellulose sensing in *N. crassa*.

Cellobiose and/or its transglycosylation products have been considered the primary inducing molecules for fungal cellulases [72,128]. In this study, the $\Delta clr-1$ mutant grew poorly on cellobiose. Some cellobiose utilization genes were more strongly dependent upon *clr-1* for their induction, particularly by cellobiose. I interpreted these data to suggest that CLR-1 is a crucial element in *N. crassa*'s cellobiose sensing mechanism [128,166]. Expression of *clr-2* was also dependent on *clr-1* and expression of *clr-2* was sufficient for robust cellulase expression, suggesting it serves as a secondary activator. Expression of *clr-1* transcript was insufficient for robust cellulase gene expression; post transcriptional regulation or post translational modification of CLR-1 is likely required for full cellulase induction. These results are summarized in Figure 20, a model for regulation of cellulose and xylan degradation in which CLR-1 and XLR-1 regulate the utilization of the two polymers, respectively.

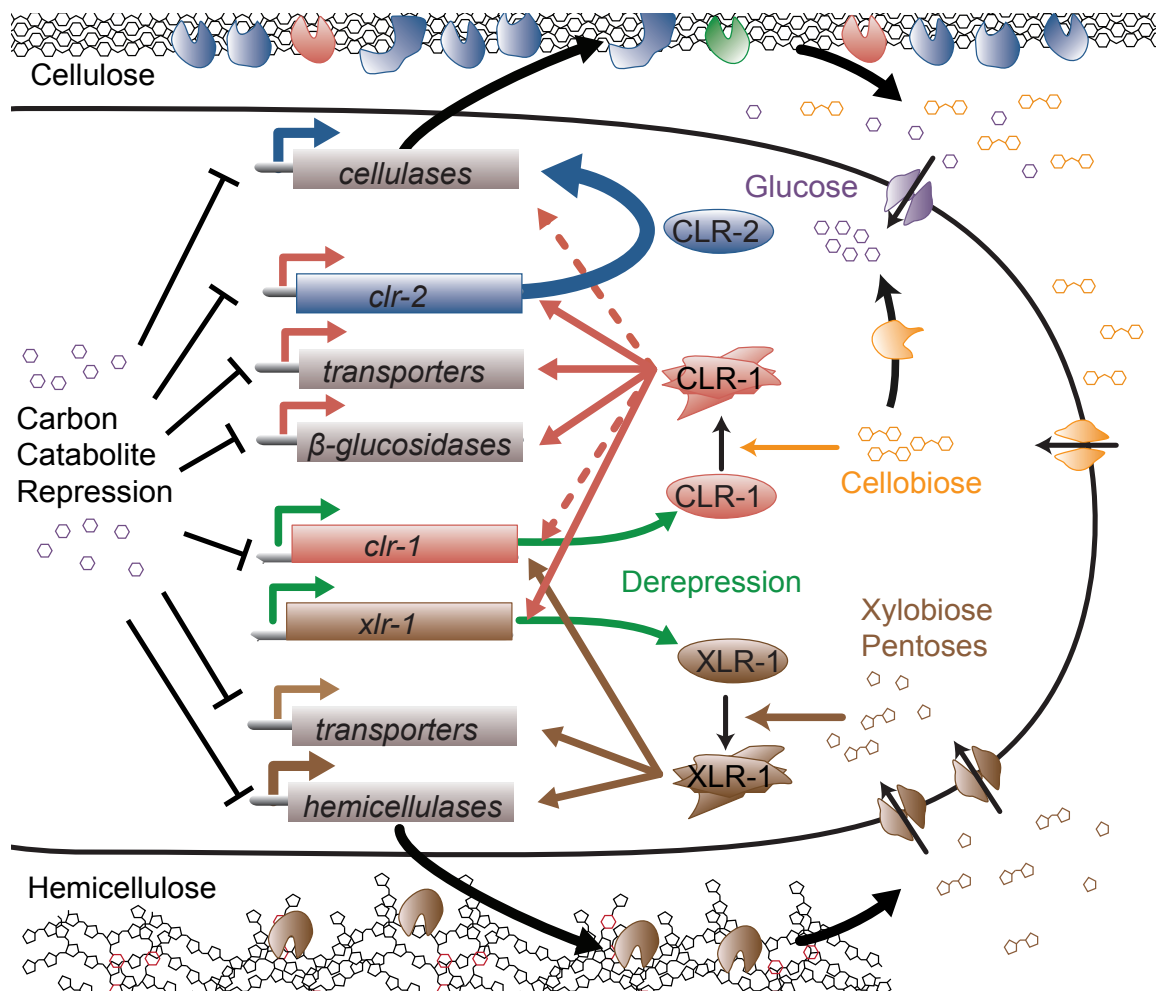


Figure 20. Model for regulation of cellulose and xylan degradation in *N. crassa*

Some key genes (e.g. the oligosaccharide transporter *cdt-2* NCU08114) are under co-regulation by these two transcription factors and they may modulate each other's expression (*xlr-1* expression on Avicel was lower in the Δ *clr-1* mutant). In this way each response regulator poises the cell for the other regulator to detect its respective inducing molecule, but largely they regulate distinct sets of genes targeted to their respective polymers.

The combination of a response regulator (CLR-1) in a feed-forward loop with second transcription factor may improve the CLR-1/CLR-2 system's robustness to noisy signals [167] and thus help avoid costly expression of secreted enzymes in environments without sufficient local cellulose to reward such an intensive investment of energy and nutrients in the form of secreted proteins. In addition to any such signal processing advantages, the separation of cellobiose utilization genes under direct regulation of CLR-1 and cellulase genes under more prominent regulation of CLR-2 allows for the decoupling of cellulase production and cellobiose utilization even as cellobiose is used as the signaling molecule for cellulase production. When carbon is scarce, environmental cellobiose will induce expression of cellulases to release cellobiose, transporters to import it and β -glucosidases to hydrolyze it. At relatively high cellobiose flux, repressing glucose and other

metabolic intermediates will accumulate and activate carbon catabolite repression. Both CLR-2 and the cellulase genes themselves are subject to direct regulation by repressive transcription factors such as CRE-1 [114], thus that repression will have a relatively stronger effect on the cellulases than the cellobiose utilization genes. Cellobiose metabolism will continue relatively unhindered even as cellulase secretion is drastically reduced.

With these data I have shown that the majority of the Avicel regulon identified in Chapter 2 requires *clr-1*, *clr-2* or both to reach WT induced expression levels. This gene set likely comprises the essential core enzymatic components for efficient degradation of cellulose from complex plant substrates. It not only includes all highly expressed cellulase genes, but many hemicellulase genes, reinforcing the notion that complimentary enzymatic activities are required to effectively degrade complex substrates [168]. In addition, dozens of secreted proteins with no predicted function or only a broad homology-based annotation showed *clr*-dependent induction. This collection of proteins provides a rich source for identification of new enzymes with synergistic activities on complex substrates, as recently demonstrated [47,48]. Also included in this gene set were some well-characterized proteins with functions in secretion, suggesting that modification of the *N. crassa* secretory pathway is needed to accommodate the dramatic increase in protein secretion that accompanies cellulosic metabolism.

A minority of genes induced by Avicel did not require *clr-1* or *clr-2* for that induction. Among these genes are several hemicellulase genes, such as *gh10-2* NCU08189 and *gh51-2* NCU02343 and components of the xylose utilization pathway such as xylose reductase NCU08384 and xylitol dehydrogenase NCU00891, all of which have been shown to be under regulation of XLR-1 in response to induction by xylose [89]. Together these data suggest that while Avicel is a relatively pure model cellulose, it still includes some hemicellulosic contamination, an inescapable issue with any purified plant cell wall component.

Expression of CLR-2 alone was sufficient for robust cellulase expression in the absence of inducing cellobiose or Avicel. This result suggested that manipulation of CLR-2 homologs in some industrial cellulase production strains may prove useful for improved cellulase production and allow process development in a wider range of culture conditions, either of which could substantially reduce the cost of industrial cellulase production and thus lower the barrier to implementation of the biorefinery concept on sustainable cellulosic feedstocks. The realization of that potential rests on conservation of function for CLR-1 and CLR-2 homologs in other species, which will be the subject of Chapter 5.

These data also suggest that CLR-1 plays an intimate role in cellobiose sensing. The final activation of CLR-1 is the output of a cellobiose sensing system in *N. crassa* that has not been elucidated. The mechanism of this system may inform our understanding of how fungi interact with their local environment and further facilitate rational engineering of filamentous fungi for biotechnological applications. A series of experiments to elucidate this sensing system is the subject of Chapter 6.

Acknowledgments.

Jordan Tsai constructed the β -glucosidase deletion mutants with additional deletions of *clr-1* or *clr-2* (Figure 18) under the direction of Elizabeth Znameroski, who kindly shared them with me.

Jianping Sun constructed the *Pccg-1:clr-1*-GFP strain employed in Figure 13.

Published material

Data in Figure 11, Figure 12, Figure 14, Figure 15, Figure 16, and Figure 19 were originally published in [39] (WT, Δ *clr-1*, and Δ *clr-2* data) and [169] (*Pccg-1:clr-2* data):

Coradetti ST, Craig JP, Xiong Y, Shock T, Tian CG, et al. (2012) Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *Proc Natl Acad Sci U S A* 109: 7397-7402.

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.

Methods

Strains.

The wild type reference strain and background for all *N. crassa* mutant strains was FGSC 2489 [131]. Deletion strains for *clr-1* and *clr-2* (FGSC 11029 and FGSC 15835, respectively) were obtained from the Fungal Genetics Stock Center (<http://www.fgsc.net/>) [157]. The *clr-2* mis-expression strain (*Pccg-1:clr-2*) was constructed by transforming a *his-3 rid-1 Δ sad-1; Δ clr-2 mat A* mutant with pSC04 (appendix C), a variant of plasmid pMF272 [170], where the open reading frame and 3' untranslated region (UTR) of *clr-2* were placed under the control of the promoter and 5' UTR of the clock controlled gene 1 (*cgc-1*). Transformants were selected for histidine prototrophy and backcrossed to a *his-3; Δ clr-2 mat a* strain to obtain a *his-3::pccg-1-clr-2 rid-1 Δ sad-1; Δ clr-2 mat A* homokaryotic strain.

Culture conditions.

N. crassa cultures were grown on Vogel's minimal medium (VMM) [132]. Carbon sources were 2% w/v unless otherwise noted. Strains were inoculated into 3 mL VMM slants and grown at 30°C in the dark for 48 hrs, then at 25°C in constant light for 4-10 days to stimulate conidia production. Conidia were inoculated into 100 mL liquid media at 10^6 conidia/mL and grown at 25°C in constant light and shaking (200 rpm).

Media shift experiments.

N. crassa cultures were grown 16 hrs on VMM, washed with Vogel's [132] without a carbon source, then Mycelia were re-suspended in 100 mL Vogel's with 2% carbon source (sucrose, cellulose (Avicel® PH-101, Sigma Aldrich, MO) or hemicellulose (beechwood xylan, Sigma Aldrich, MO)) or with no carbon source added. Total RNA was extracted as described in [44]. For enzyme activity assays,

culture supernatants were sampled at 4-96 hrs and stored at 4°C for analysis within 24 hrs of collecting the final time point.

RT-PCR

RT-PCR was performed with the EXPRESS One-Step SYBR® GreenER™ Kit, with premixed ROX (Invitrogen #11790-200) using the standard cycling protocol in a 20µl reaction with 5ng total RNA as the input. Amplification was carried out on a CFX Connect™ Real-Time PCR Detection System (Biorad). Primers were tested for similar efficiencies and specificities were checked with melt-curve analysis. The Δ CQ method was used for transcript abundance with actin (NCU04713) as the reference gene. Primer sequences are listed in Appendix B.

RNA Sequencing and Transcript Abundance

Libraries were prepared with standard protocols from Illumina Inc (San Diego, CA) and sequenced on the Illumina Genome Analyzer IIx and HiSeq 2000 platforms. Libraries were prepared as described in Chapter 2 and mapped against predicted transcripts from the *N. crassa* OR74A genome [118] from the Broad Institute (www.broadinstitute.org) version 12 annotation 7 with Tophat v2.0.4 [133]. Transcript abundance (FPKM) was estimated with Cufflinks v2.0.2 [171] mapping against reference isoforms. Profiling data are available at the NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>; [Series Record GSE44100](#)).

FPKM of genes in the Avicel regulon were hierarchically clustered with the Cluster 3.0 software suite [159]. Prior to clustering, FPKMs were log transformed, normalized across strains/conditions and centered on the geometric mean across strains/conditions on a per-gene basis. The average linkage method was used for cluster generation, with uncentered correlation as the similarity metric. Clusters were visualized with Javatreeview [160].

Enzyme Activity Assays

Relative enzymatic activity of culture supernatants was assayed with Remazolbrilliant Blue R conjugated CMC and xylan (birchwood) kits from Megazyme (Wicklow, Ireland). Total protein was determined with the Bradford assay (BioRad). SDS-PAGE gels were performed with Criterion™ Tris-HCl Precast Gels (4-15% Gradient) (BioRad)

Statistical significance tests

Unless otherwise noted all statistical significance tests were done with a one-tailed homoscedastic (equal variance) T-test. One star: $P < 0.05$, two stars $P < 0.01$, three stars $P < 0.001$.

Chapter 5

Conservation of CLR-1 and CLR-2 Function in *Aspergillus nidulans*

Abstract

In *Neurospora crassa*, the transcriptional regulators CLR-1 and CLR-2 are required for cellulolytic gene expression and cellulose deconstruction. To assess conservation and divergence of cellulase gene regulation between fungi from different ecological niches, I compared CLR-1 and CLR-2 function with that of their orthologs (ClrA and ClrB) in the distantly related species, *Aspergillus nidulans*. Transcriptional profiles induced by exposure to crystalline cellulose were similar in both species. ClrA had a dramatically reduced regulon as compared to CLR-1 with only one major cellulase dependent on ClrA for full induction. Approximately 50% of the cellulose-responsive genes showed strict dependence on functional CLR-2/ClrB, with a subset of 28 genes encoding plant biomass degrading enzymes that were conserved between *N. crassa* and *A. nidulans*. Constitutive expression of ClrB was not sufficient to drive cellulase gene expression in *A. nidulans*, but did sensitize the strain to induction by mannobiose. Manipulation of *clr-2* orthologs among filamentous fungi may enable regulated cellulosic enzyme production in a wide array of culture conditions and host strains, potentially reducing costs associated with enzyme production for plant cell wall deconstruction. However, this functionality may require additional engineering in some species.

Introduction

In all filamentous fungi studied to date, expression of cellulolytic genes requires both relief from CCR and the presence of an inducing molecule, such as cellobiose in *N. crassa* [128]; sophorose, lactose or xylose in *Trichoderma reesei*; [87,124,172-174] or xylose in several *Aspergillus* species [81,137,175]. Several independent lines of mutagenesis have produced strains with increased cellulase production and decreased CCR [176,177], but constitutive, robust production of cellulases without inducing molecules remained elusive over 30 years of mutagenesis based research. Recently, a constitutively active mutant of *T. reesei* Xyr1 was isolated through random mutagenesis [63], thereby opening up new possibilities for process development using that species for industrial enzyme production.

In Chapter 4, studies in *N. crassa* identified two new transcription factors, CLR-1 and CLR-2, which regulate cellulase transcription in response to environmental cellulose, likely through cellobiose sensing. Expression of CLR-2 was sufficient to induce cellulase secretion in non-inducing conditions. This result suggested that rational manipulation of CLR-2 orthologs in other species of filamentous fungi could enable greater production of cellulases in a wider range of growth conditions without the laborious mutagenesis and screening processes that

have been applied to *T. reesei*. The realization of that potential will require a broad understanding of how CLR-2 orthologs function across the filamentous ascomycete fungi. An appropriate first step in understanding the broader scope of CLR-2 function in fungi was to expand my work in *N. crassa* into another model fungus, *A. nidulans*.

In nature, *N. crassa* is associated with plant material that has been killed by fire [178]. *A. nidulans* is a more cosmopolitan species and has been isolated from a variety of substrates including dead plant matter [58]. Thus, both species commonly degrade plant cell wall material, but in differing contexts. Both *N. crassa* and *A. nidulans* have a wide array of genomic, genetic and biochemical tools that make these two species an ideal platform for comparative analyses of processes associated with plant cell wall deconstruction. Their relative phylogenetic distance (Figure 2) also makes them useful for assessment of functional conservation of gene function and regulation over 300 million years of evolutionary divergence [57].

In this chapter, I perform an in depth comparative analyses of genes induced by exposure of *N. crassa* and *A. nidulans* to crystalline cellulose, and identify conserved and divergent genes regulated by the *A. nidulans* transcription factors ClrA and ClrB, orthologs to *N. crassa* CLR-1 and CLR-2 respectively. I also show that while ClrB is critical for cellulase expression in *A. nidulans*, transcription of *clrB* is not sufficient for cellulase expression and secretion.

Conservation of CLR-1 and CLR-2 protein sequences in filamentous ascomycete fungi.

I performed a phylogenetic analysis of CLR-1 and CLR-2 across a custom built database of over 40 fungal genomes. The predicted protein sequences showed broad conservation across filamentous ascomycete fungi. Maximum likelihood phylogenetic trees of CLR-1 (Figure 21) and CLR-2 (Figure 22) homologs are largely concordant with an approximate species tree constructed with the highly conserved DNA licensing factor MCM7, previously used for phylogenetic studies of fungal evolution [179]. CLR-1 and CLR-2 were well conserved in the filamentous ascomycete fungi, with orthologs retained in most species included in this analysis. That conservation was limited the Pezizomycotina as likely orthologs were absent from the Saccharomycotina and more distant subphyla. One notable exception to this congruence of species and gene trees was the most conserved CLR-1 homolog from the industrially important species *T. reesei* (protein ID 121121). This gene clustered as an outgroup and was clearly not orthologous to CLR-1. I concluded that *T. reesei* has lost its true CLR-1 ortholog. CLR-1 orthologs were also missing from the animal fungal pathogens *Coccidioides immitis* and *Coccidioides posadasii*. CLR-2 appeared to be better conserved than CLR-1 across the species analyzed with a likely ortholog present in all Pezizomycotina examined. CLR-2 orthologs have undergone more duplication events, with paralogous genes in *A. fumigatus*, *A. nidulans* and *P. chrysogenum*. The *A. nidulans* orthologs ClrA and ClrB as well as the *A. oryzae* CLR-2 ortholog ManR (characterized as a regulator of mannanases [112] and cellulases [180] in response to β -mannobiose) are also highlighted in Figure 22.

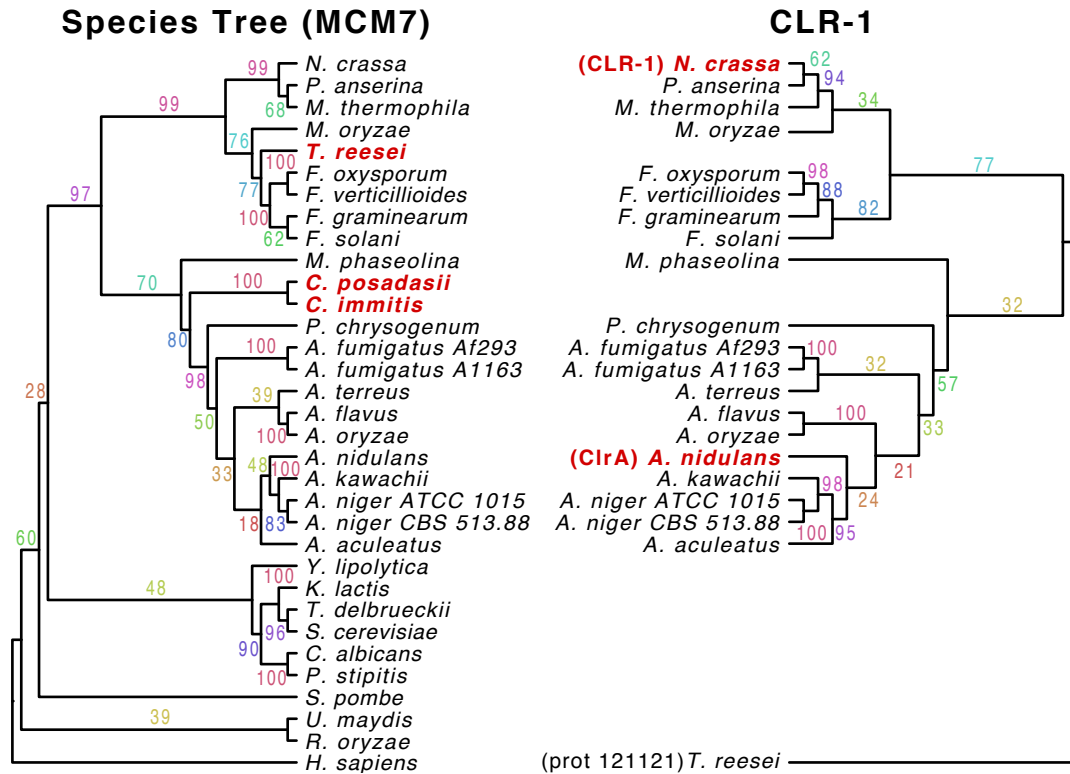


Figure 21. Phylogenetic tree of CLR-1 homologs in filamentous fungi.

Reference species tree computed with the DNA licensing factor *mcm7*. Branches labeled with bootstrap support from 100 random samples. The most similar protein from the *T. reesei* genome (Protein #121121) was used as an outgroup and is not orthologous to CLR-1. No orthologs to *clr-1* were found in the genomes of *C. posadasii* or *C. immitis*.

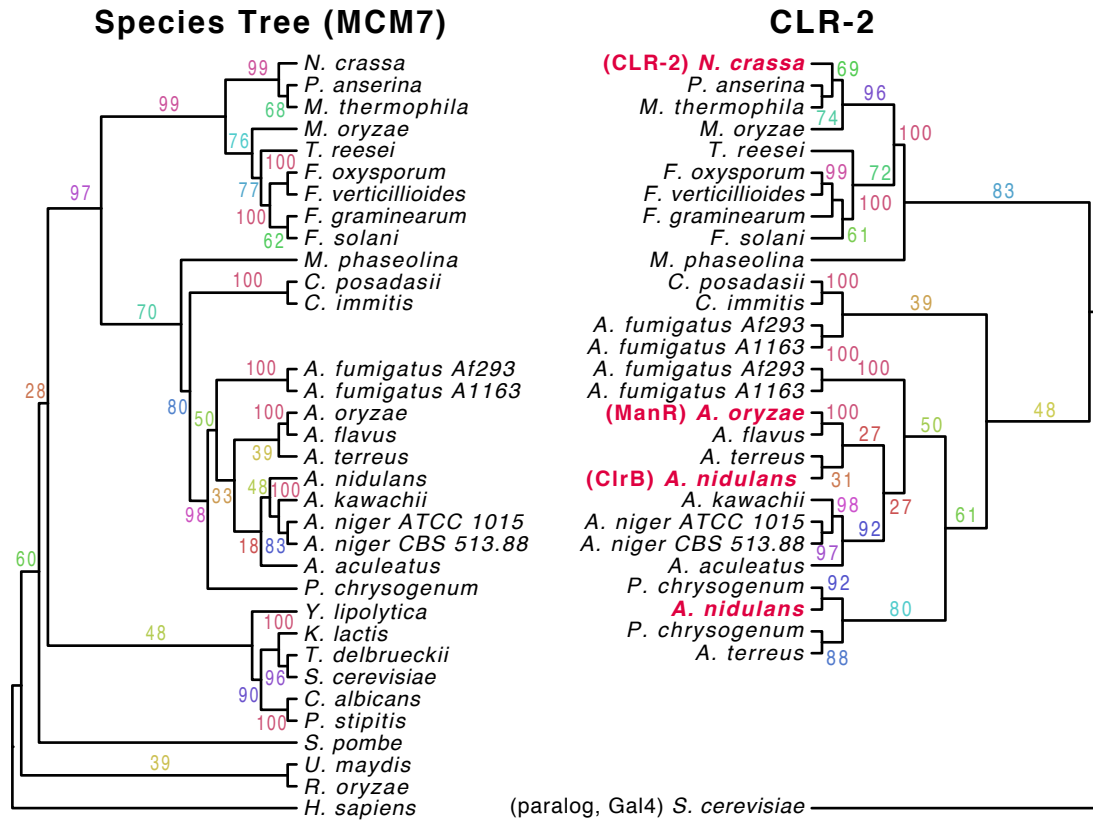


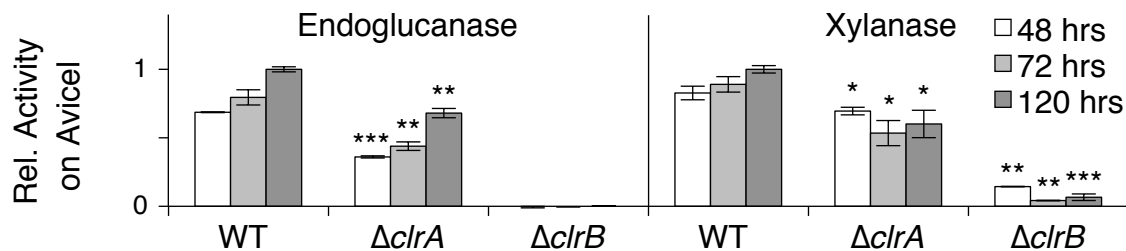
Figure 22. Phylogenetic tree of CLR-2 homologs in filamentous fungi.

Reference species tree computed with the DNA licensing factor *mcm7*. Branches labeled with bootstrap support from 100 random samples. The most similar protein from the *S. cerevisiae* genome (Gal4) was used as an outgroup and is not orthologous to CLR-2. Orthologs of note include the *A. nidulans* ClrB and *A. oryzae* mannobiose response regulator ManR.

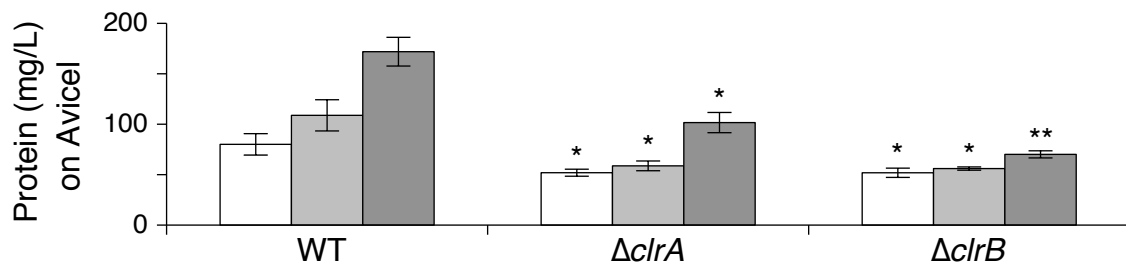
To test if ClrA and ClrB function to regulate genes involved in plant cell wall deconstruction *A. nidulans*, I collaborated with lab member Yi Xiong to assess cellulose utilization and cellulase gene expression in *clrA* and *clrB* deletion mutants. Yi created deletion strains for the loci AN5808 (*clrA*) and AN3369 (*clrB*). I then subjected those deletion strains to similar growth and enzyme activity test as I subjected $\Delta clr-1$ and $\Delta clr-2$ strains to in Chapter 4.

The *A. nidulans* $\Delta clrA$ and $\Delta clrB$ deletion strains were deficient for cellulase and xylanase activity as well as total protein secretion when pre-grown glucose cultures were transferred to Avicel (Figure 23). Enzyme activity was abolished in the $\Delta clrB$ mutant, but the $\Delta clrA$ mutant showed ~ 50% of WT activity (Figure 23A-B). Both deletion mutants were deficient for growth on cellobiose, although $\Delta clrB$ was more strongly affected (Figure 23C). In RT-PCR experiments carried out by Yi Xiong, the induction pattern of major cellulase genes in the $\Delta clrB$ mutant was several thousand-fold less than WT (Figure 23D), while in the $\Delta clrA$ mutant, the average induction was lower, but for 3 of 4 cellulases, this difference failed to meet the oft-used statistical criterion of $P < 0.05$ in a T-test. From these data, we concluded that *clrA* had a less important role in cellulase induction in *A. nidulans* as compared to *clr-1* in *N. crassa*. However, the function of CLR-2/ClrB as an essential activator for cellulase gene expression and activity was conserved.

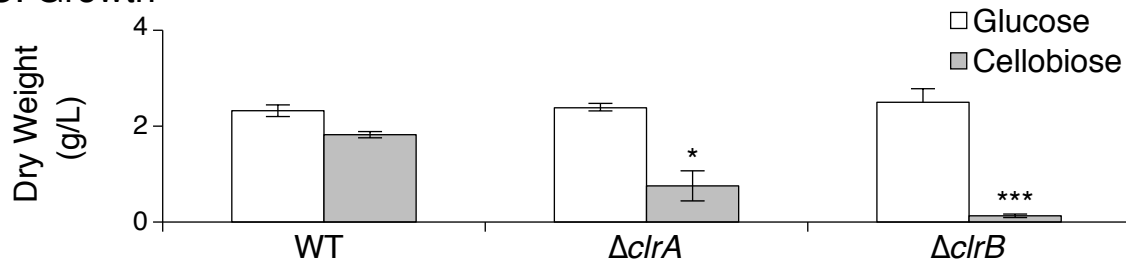
A. Enzyme Activity



B. Total Protein



C. Growth



D. Induction

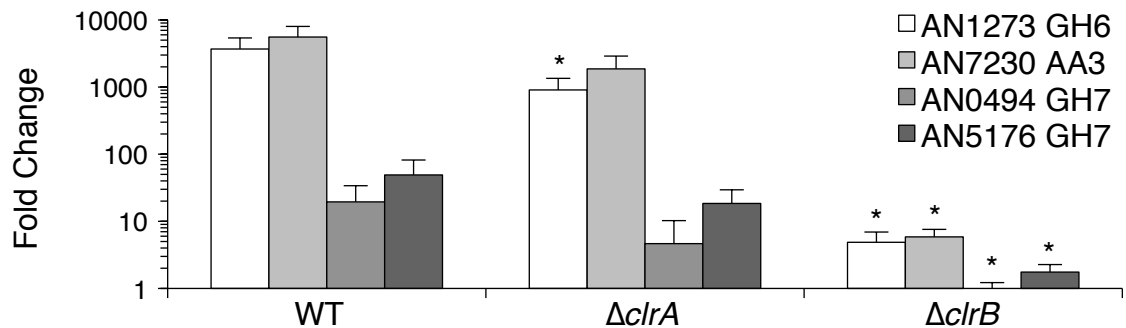


Figure 23. Phenotype of *Aspergillus nidulans* *clr* deletion strains.

(A) Enzyme activity of culture supernatants from $\Delta clrA$ and $\Delta clrB$ mutants grown on glucose and then shifted to Avicel medium for 48, 72 and 120 hrs.

(B) Total protein in supernatants of cultures grown on Avicel.

(C) Mycelial dry weights from WT and *clr* mutants from carbon-limited cultures on glucose and cellobiose (0.5% w/v).

(D) Induction of selected cellulase genes in WT and the *clr* mutants following an 8 hr shift to Avicel from glucose-grown cultures, by RT-PCR.

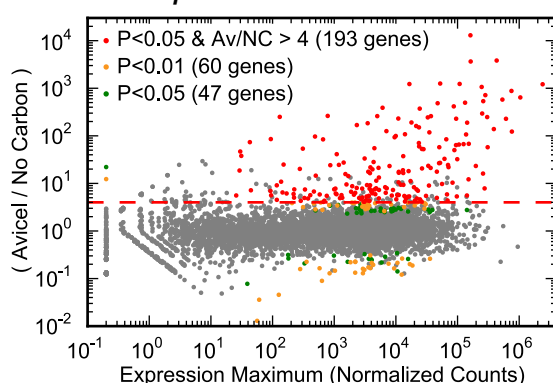
Statistical significance by one tailed, unequal variance T-test.

A comparison of the Avicel regulons of *A. nidulans* and *N. crassa* reveals conserved clusters of genes

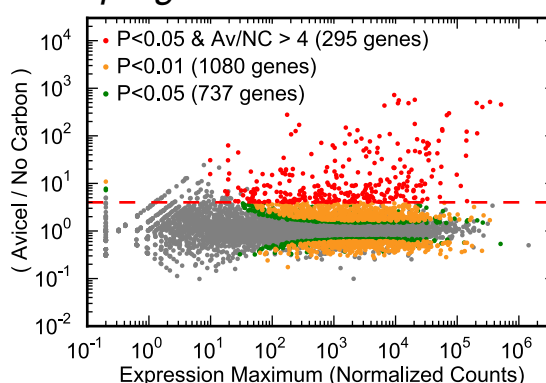
To comprehensively compare the CLR-2/ClrB regulons in these two evolutionarily divergent species, we first characterized the transcriptional response of *A. nidulans* to Avicel. To compare the response of *A. nidulans* to Avicel with that of *N. crassa*, Yi Xiong transferred WT *A. nidulans* strains from a simple carbon source, glucose, to Avicel or no carbon source media and subjected them to RNAseq. I then analyzed transcriptional profiles of the switched cultures in comparison to my previous experiments with WT *N. crassa*. To establish a comparable Avicel regulon, genes strongly induced on Avicel (>4-fold transcript abundance with a statistical significance beyond a 5% false discovery rate, as calculated with DESeq [135]) were compared between the two species (Figure 24). This analysis differed somewhat from my previous differential expression analysis (see Chapter 2), but was found to be more robust to differing levels of biological variation among independent samples than the earlier analysis. Note the large number *A. nidulans* genes for which DEseq assigned differential expression with a 1% false discovery rate but which had relatively small effect sizes (Figure 24B). By limiting the comparison to genes with relatively strong induction, I thereby avoided comparing genes with induction of dubious biological significance.

N. crassa and *A. nidulans* exhibited a broadly similar transcriptional response to Avicel with several hundred genes meeting the conservative criteria for specific induction (Figure 24A-B). The magnitude of this induction was greater in *N. crassa*, particularly for enzyme classes with predicted function in the degradation of crystalline cellulose, such as genes encoding enzymes in the glycoside hydrolase (GH; [16]) 7 family (cellobiohydrolase I), GH6 family (cellobiohydrolase II), polysaccharide monooxygenases (AA9) and GH5 family, subclass 5 (endoglucanases) (Figure 24C). Also enzyme classes GH74 (xyloglucanase and endoglucanase) and GH5 subclass 7 (endomannanase) were more highly expressed in *N. crassa*. Also more highly expressed in *N. crassa* was the CE5 (acetyl xylan esterase) enzyme class. Finally while the GH51 class was more highly expressed in *N. crassa*, this difference was driven by high expression of one alpha-arabinofuranosidase (NCU01900), which was balanced in *A. nidulans* by the cumulative expression of two alpha-arabinofuranosidases of the GH62 class (AN2632 and AN7908).

A. *Neurospora crassa*



B. *Aspergillus nidulans*



C. CAZy Class Expression

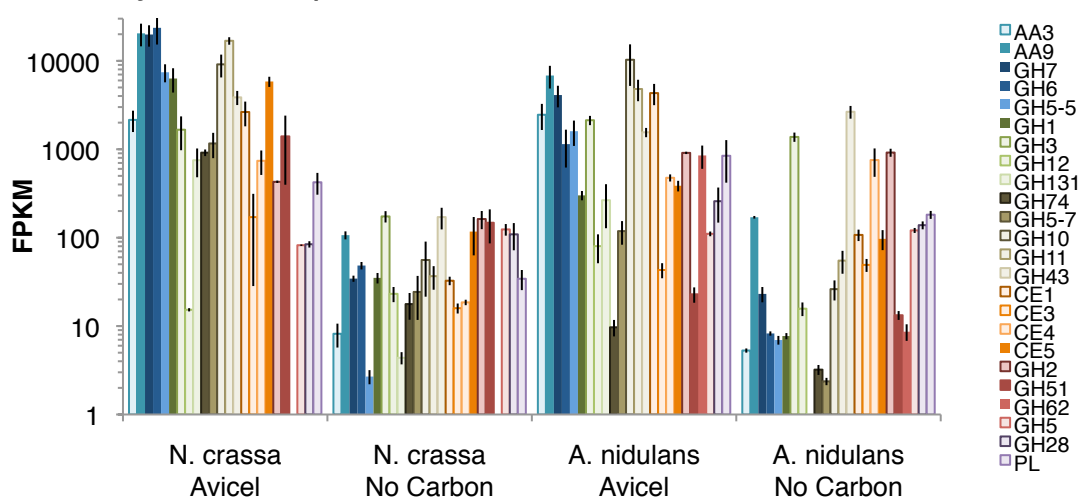


Figure 24. The Avicel regulons of *N. crassa* and *A. nidulans*.

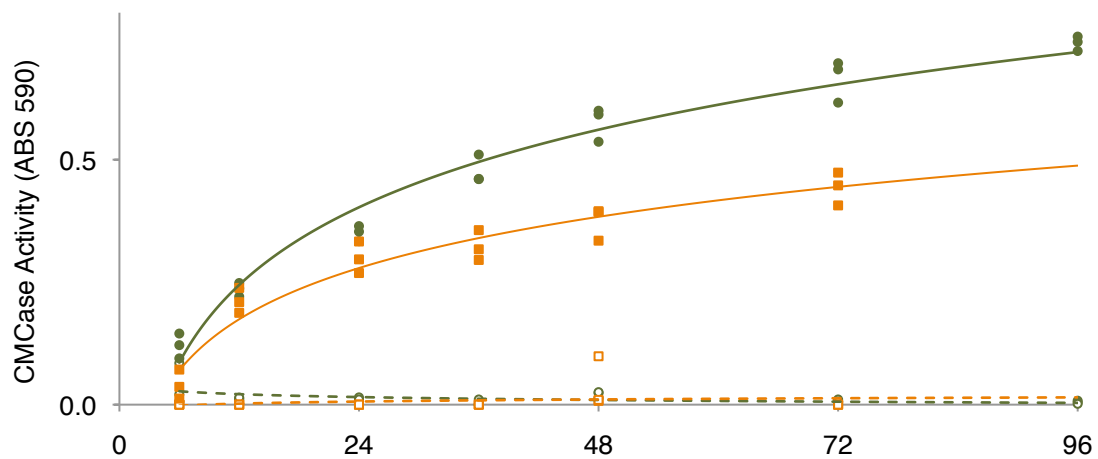
(A) Relative mRNA abundance of *N. crassa* exposed to Avicel vs no carbon conditions. Genes showing significant differential expression are colored by adjusted P-value.

(B) Relative mRNA abundance of *A. nidulans* exposed to Avicel versus no carbon source conditions. Genes showing significant differential expression are colored by adjusted P-value.

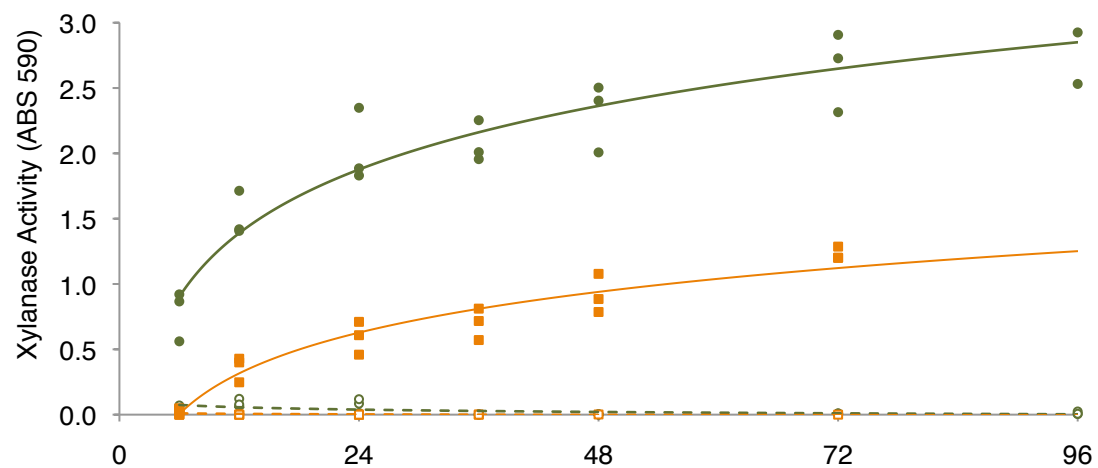
(C) Total expression from RNAseq data of genes encoding the major classes of plant cell wall carbohydrate active enzymes (CAZy; [16]) from a WT *N. crassa* culture shifted to media with no carbon source or to Avicel and from a WT *A. nidulans* culture shifted to media with no carbon source or to Avicel. FPKM (Fragments Per Kilobase per exon per Megabase mapped) for individual genes were averaged between biological replicates and pooled by CAZy class. Classes acting primarily on crystalline cellulose are shown in blue; classes acting primarily on hemicellulose are shown in brown and orange; and classes acting primarily on pectin are colored in purple. Classes with total expression varying by an order of magnitude or more between species are shown with solid bars.

To compare enzyme activity and protein secretion, I evaluated supernatants from *N. crassa* and *A. nidulans* cultures shifted to either Avicel or sucrose. Enzyme activity on amorphous cellulose (carboxymethyl cellulose) and xylan were significantly higher from *A. nidulans* Avicel-grown supernatants as compared to *N. crassa* (Figure 25A-B). However, enzyme activity on crystalline cellulose was significantly higher for Avicel supernatants from *N. crassa* (Figure 25D-F). *N. crassa* also secreted more protein, consumed more cellulose and accumulated more biomass on Avicel than *A. nidulans* (Figure 25C-E). These results reflected the observed transcriptional differences between *N. crassa* and *A. nidulans* in genes encoding enzymes with predicted activity on crystalline cellulose (Figure 24C).

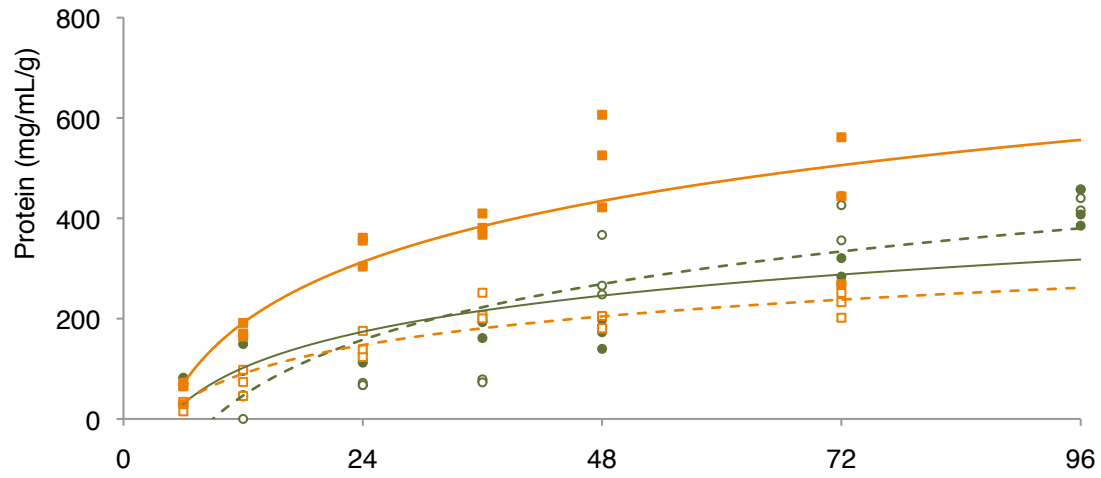
A. Endoglucanase Activity



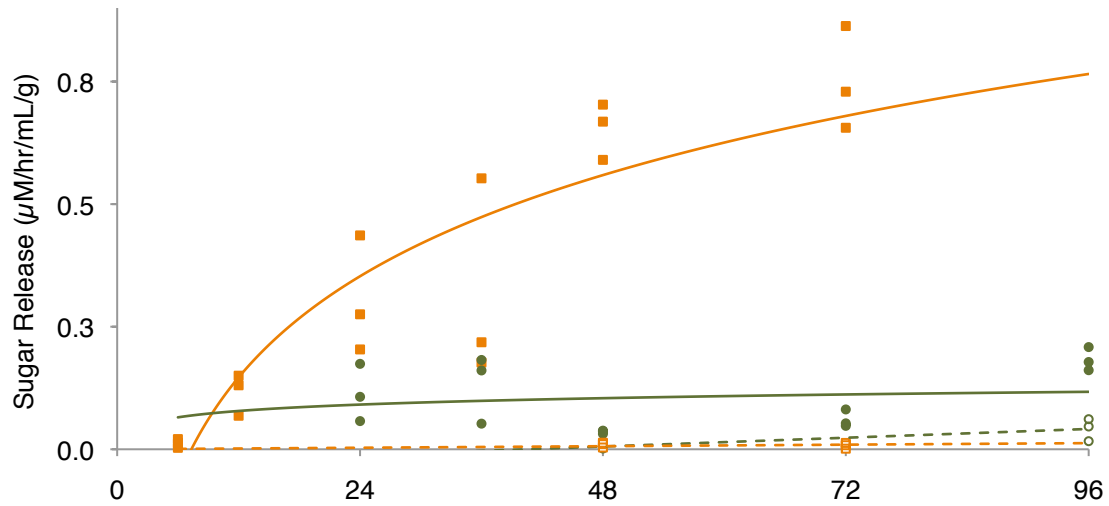
B. Xylanase Activity



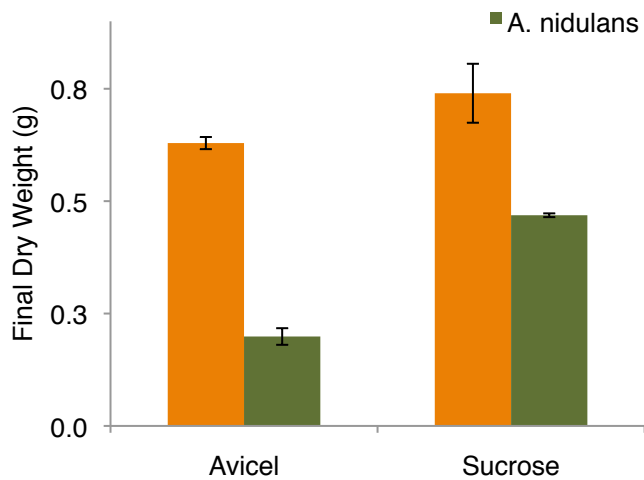
C. Secreted Protein



D. Avicelase Activity



E. Final Dry Weights



F. Avicel Digestion

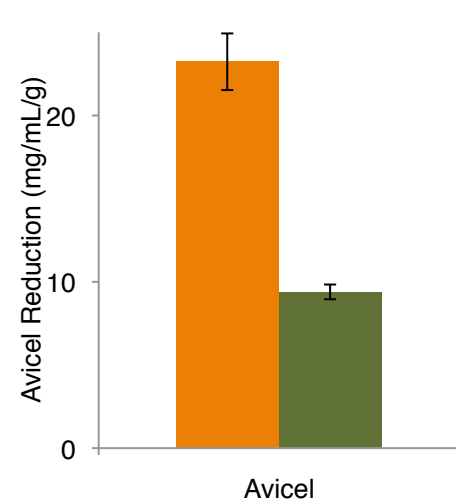


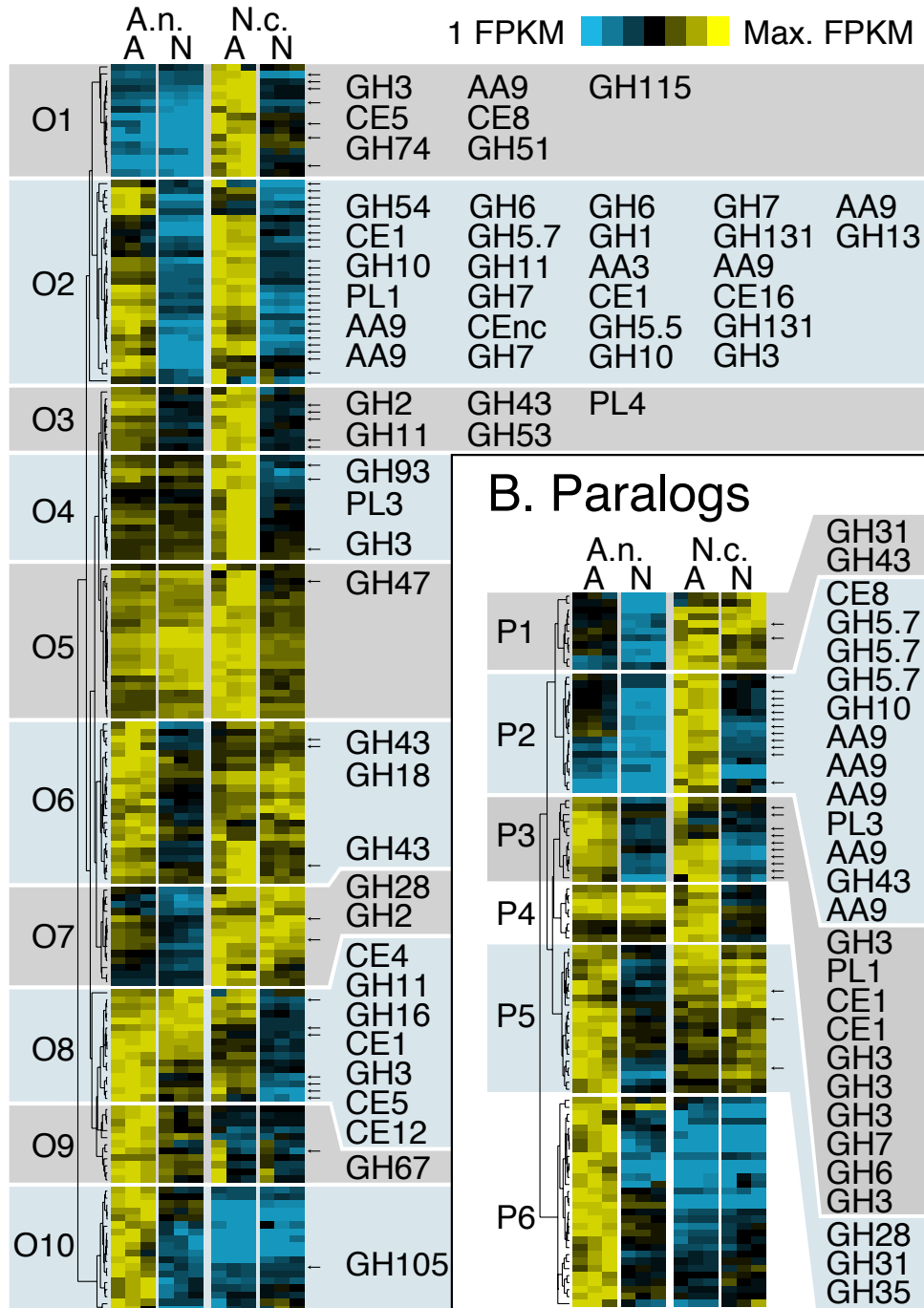
Figure 25. Time course of enzyme activity from *N. crassa* and *A. nidulans* culture supernatants.

(A) CMCase activity of culture supernatants from *N. crassa* or *A. nidulans* cultures shifted to either Avicel or sucrose. (B) Xylanase activity of culture supernatants from *N. crassa* or *A. nidulans* cultures to either Avicel or sucrose. (C) Secreted protein in *N. crassa* or *A. nidulans* supernatants from cultures shifted to either Avicel or sucrose, as determined by Bradford assay. (D) Cellulase activity, as determined by sugar release from Avicel, of *N. crassa* or *A. nidulans* supernatants from cultures shifted to either Avicel or sucrose. (E) Final mycelia dry weights 72 and 96 hrs after shift of *N. crassa* and *A. nidulans* cultures to sucrose or Avicel. Biomass in Avicel cultures was estimated by degrading mycelia and weighing the recalcitrant fraction [181]. (F) Cumulative breakdown of Avicel normalized to final dry weight of cultures. (A-D) All data points normalized to final culture biomass as in (E).

To assess the conservation of the Avicel regulon between *N. crassa* and *A. nidulans* on a gene-by-gene basis, expression of orthologous and paralogous gene pairs was sorted with hierarchical clustering. Among genes with likely orthologs in each species, 10 gene expression clusters were identified (Figure 26). Cluster O1 consisted of genes strongly induced in *N. crassa* under Avicel conditions, but with very low expression in *A. nidulans*. This cluster included genes encoding two putative transporters (NCU01633/AN4277 and *cdr4* NCU05591/AN0771), a β -glucosidase *gh3-6/bglB* (NCU07487/AN0712), as well genes encoding six additional secreted carbohydrate-active enzymes (CAZy), an acetylcholine esterase (NCU08752/AN7407) and a monooxygenase (NCU7055/AN8199).

Cluster O2 consisted of 29 genes strongly and specifically induced on Avicel in both species, 25 of which encoded secreted CAZy proteins. These included the most highly expressed and well-characterized cellulases: *cbh-1* (NCU07340, orthologous to both *cbhA* AN5176 and *cbhB* AN0494), *gh6-2* (NCU09680 orthologous to both *cbhC* AN5282 and *cbhD* AN1273), *gh5-1/eglA* (NCU00762/AN1285) and multiple polysaccharide monooxygenases (AA9). Cluster O3 consisted of 9 genes strongly induced on Avicel, but with higher expression under starvation conditions than genes in cluster O2 (Dataset S1). Most of these genes are implicated in pectin or hemicellulose deconstruction. Also included were two transporters: cellodextrin transporter *cdt-1* (NCU00801/AN2814) and a less strongly expressed putative sugar transporter NCU06384/AN3498.

A. Orthologs



B. Paralogs

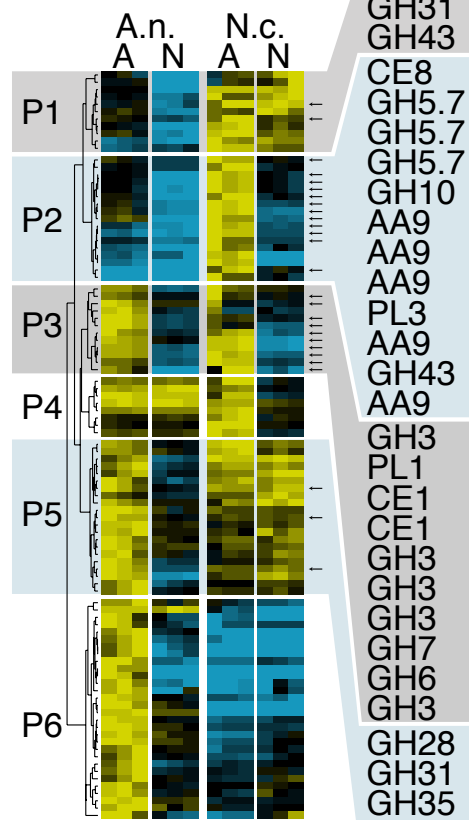


Figure 26. Hierarchical clustering of orthologous and paralogous genes induced by exposure to Avicel in *N. crassa* and *A. nidulans*.

(A) Hierarchical clusters of FPKM for ortholog pairs (O1-O10) for which at least one member passed criteria for specific induction on Avicel. (B) Hierarchical clusters of FPKM for paralogous pairs (P1-P6) for which at least one member passed criteria for specific induction on Avicel.

A.n.= *Aspergillus nidulans*; N.c.=*Neurospora crassa*. Avicel [A] or no carbon source [N] conditions. Heat maps display log (FPKM) scaled from 1 (bright blue) to the maximal value for each pair (bright yellow). CAZy genes in each cluster are marked with small arrows and listed in order to the right of each cluster.

Clusters O4 and O5 consisted of genes with specific induction under Avicel conditions in *N. crassa*, but relatively constant expression at an intermediate level in *A. nidulans*. These clusters were differentiated mostly by their relative transcript abundances in *N. crassa* under starvation conditions. They included several genes that function in xylose utilization: xylose reductase NCU08384/AN0423, xylitol dehydrogenase NCU00891/AN9064, xylulose kinase NCU11353/AN8790, ribose-5-phosphate isomerase NCU10107/AN5907, and mitochondrial triose phosphate isomerase NCU10106/AN5908. Six genes with known or suspected roles in protein secretion and trafficking also clustered tightly in O5: *hsp70-6* NCU09485/AN0847, calreticulum/calnexin homolog *clxA* NCU09265/AN3592, *sec61-1* NCU08897/AN7721, *grp78/bipA* NCU03982/AN2062, *SEC62* homolog NCU06333/AN6269, and protein disulfide isomerase A *pdiA* NCU09223/AN7436. Also prominent in these clusters were several genes involved in aspartate/asparagine and aromatic amino acid biosynthesis. The cellulolytic regulators *clr-1/clrA* (NCU07705/AN5808) and *clr-2/clrB* also clustered in O5 and O4, respectively. They showed specific induction in both species, but the magnitude of the change was more dramatic in *N. crassa*.

Clusters O6 and O7 consisted of 23 and 14 genes, respectively, with specific induction under Avicel conditions in *A. nidulans*, but with relatively constitutive expression in *N. crassa*. They included a few secreted pectinases (arabinosidase *gh43-7* NCU05965/AN7781, exopolysaccharuronase *gh28-2* NCU06961/AN8891, GDSE lipase/acylhydrolase NCU06364/AN1792) and intracellular enzymes with activity on pectin breakdown products (L-galactonate dehydratase NCU07064/AN6035, β -xylosidase *gh43-5/bxlD* NCU09652/AN7864, β -galactosidase *gh2-3/lacD* NCU00810/AN3201 and UDP-glucose 4-epimerase NCU05133/AN2951).

Cluster O8 had generally higher expression in *A. nidulans*, but more specific induction under Avicel conditions in *N. crassa*. It was most dominated by genes encoding predicted pectinases and hemicellulases (rhamnogalacturonan acetyltransferase *ce12-1* NCU09976/AN2528, *faeC* NCU08785/AN5267, acetylxylan esterase *ce5-3* NCU09664/AN5309, *gh11-1/xlnA* NCU02855/AN3613, mixed-linked glucanase NCU01353/AN0245) and hypothetical proteins. It also included the major secreted β -glucosidase *gh3-4/bglL* NCU04952/AN2828. Clusters O9 and O10 were induced in *A. nidulans*, but not in *N. crassa*. Cluster O9 included several stress response genes (*hsp30*-like proteins AN3555 and AN2530 orthologous to *hsp30* NCU09364, *hsp78* NCU02630/AN1163, a MFS multidrug transporter NCU00306/AN0732 and a mitochondrial hypoxia-responsive gene NCU02623/AN1066).

Genes with uncertain orthology relations were matched as pairs of cross-species paralogs and clustered by FPKM with the same parameters as the orthologous pairs, yielding 6 major clusters (Figure 3B). Most of the genes encoding CAZy proteins in this set clustered together in P2 and P3. Genes in P2 showed some induction in both species, but with greater magnitude in *N. crassa*. Most notable in this group were an expansion of highly expressed AA9s in *N. crassa* and an expansion of endo- β -mannases (GH5.7) in *A. nidulans*. Genes in cluster P3 were comparably expressed in both species. Most prominent in P3 were several β -

glucosidases and β -xylosidases and the highly expressed CBHII homolog *gh6-3* NCU07190, most similar to *cbhD* AN1273. Cluster P5 genes were induced on Avicel in *A. nidulans*, and included 3 paralogous genes encoding enzymes associated with hemicellulose and/or pectin breakdown. Cluster P6 included *A. nidulans* paralogs that were induced under Avicel conditions, but not in *N. crassa*.

Lytic polysaccharide monooxygenases (LPMOs, CAZy class AA9) oxidatively cleave polysaccharides. As of this writing, cleavage of cellulose [182,183], starch [184] and xyloglucan [185] have been demonstrated. Based on phylogenetic analyses, four AA9 enzyme subclasses have been identified in *N. crassa* with distinct stereochemical properties [186]; transcript abundances of these subclasses were particularly different between *A. nidulans* and *N. crassa*. In *N. crassa*, 10 of 14 predicted AA9 genes met our criterion for specific induction on Avicel, with no particular gene accounting for more than 27% of total FPKM among this class. In *A. nidulans*, 5 of 9 predicted AA9 genes were specifically induced, but just one type 3 AA9, AN3860, accounted for more than 93% of total FPKM on Avicel.

N. crassa genes without clear homologs in *A. nidulans* that were induced on Avicel were dominated by hypothetical proteins (34 of 39); 12 were predicted to be secreted and may have novel functions in plant cell wall deconstruction. *A. nidulans* genes without clear homologs in *N. crassa* that were induced on Avicel included genes encoding predicted pectinases and enzymes with roles in secondary metabolism (sterigmatocystin/aflatoxin, terrequinone A and penicillin biosynthesis).

Conservation of gene regulation by CLR-2 and ClrB

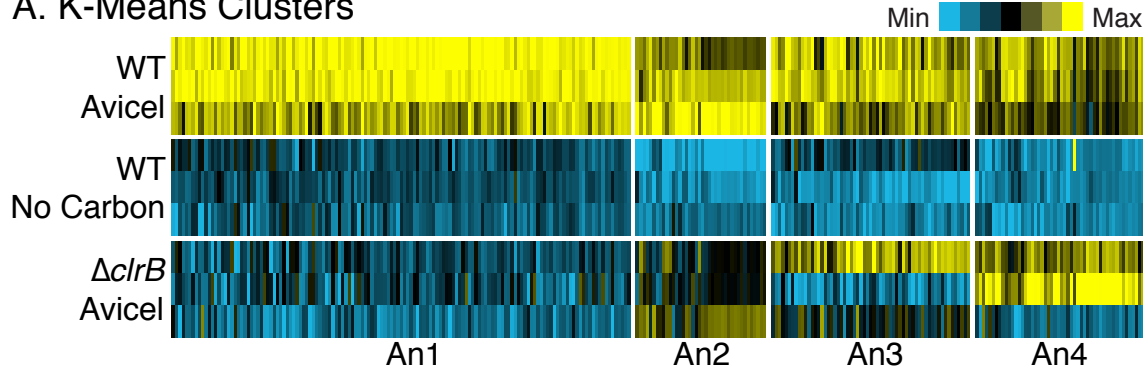
To assess conservation of function and respective target genes of CLR-2 and ClrB, I analyzed expression patterns of the Avicel regulon (295 genes total) in WT *A. nidulans* transferred to Avicel or no carbon source media versus the $\Delta clrB$ mutant transferred to Avicel. Once again these data were from RNAseq on libraries prepared in collaboration with Yi Xiong. I identified 4 clusters of genes with respect to ClrB dependence in *A. nidulans* (An1-An4; Figure 27). Cluster An1 was comprised of 141 genes strictly dependent on *clrB*. This cluster was dominated by highly expressed genes encoding the core enzymes associated with crystalline cellulose degradation: the AA3, AA9, GH7, GH6 and GH5.5 families. Although genes encoding cellulases were not induced in $\Delta clrB$ on Avicel, expression of some genes associated with hemicellulose degradation showed an increased expression level in the $\Delta clrB$ mutant on Avicel over WT under no carbon conditions (Figure 27).

Cluster An2 was comprised of 40 genes with intermediate transcript abundance in $\Delta clrB$ strains on Avicel as compared to WT, indicating a partial or indirect requirement for functional ClrB for induction on Avicel. This cluster included several genes with predicted function in secondary metabolite biosynthesis, including several robustly expressed genes of the sterigmatocystin biosynthesis gene cluster. Genes implicated in monodictyphenone and penicillin biosynthesis and a putative polyketide synthase (AN12331) also showed *clrB*-dependent induction, but were expressed at lower levels.

Genes in clusters An3 and An4 (61 and 53 genes) reached full induction in Avicel in the $\Delta clrB$ mutant, although variance in expression was high between

biological replicates (Figure 4; Dataset S1). This cluster was dominated by expression of the GH10, GH11, GH43 and GH62 gene families, which encode the major enzyme classes dedicated to hemicellulose degradation. Several of these genes or their orthologs have been identified as targets of the transcriptional regulator XlnR in *Aspergilli* [80], and also had highly variable expression in *A. nidulans* when induced by xylose [175].

A. K-Means Clusters



B. CAZy Class Expression

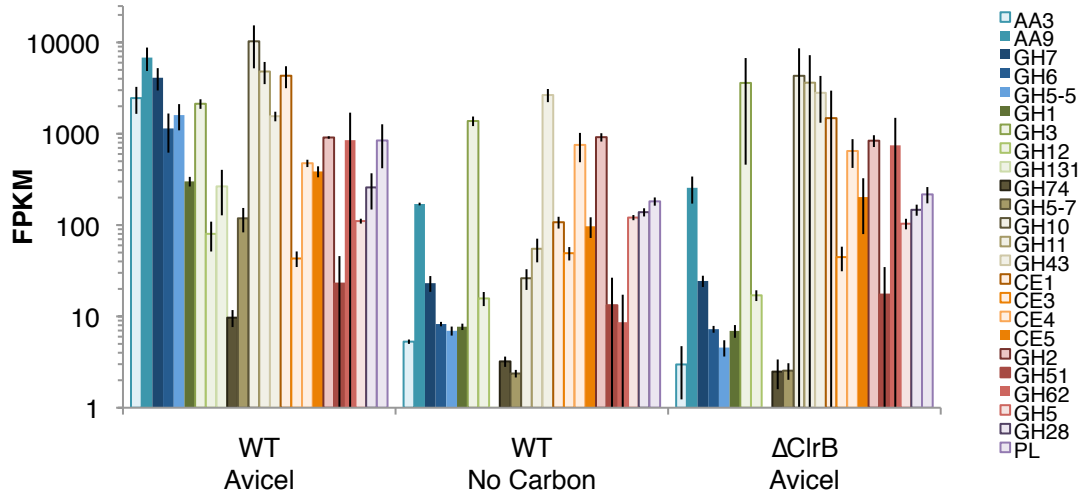


Figure 27. Dependence of Avicel-induced genes on functional ClrB in *A. nidulans*.

(A) K-means clusters of Avicel-induced genes in an *A. nidulans* $\Delta clrB$ mutant strain on Avicel as compared to a WT strain on Avicel or media with no carbon source. Heat map displays log (FPKM) scaled from minimal expression (bright blue) to the maximal expression (bright yellow).

(B) Total expression of genes encoding major CAZy classes in the $\Delta clrB$ mutant on Avicel as compared to a WT strain on Avicel or media with no carbon source. Note the absence of expression of genes encoding predicted cellulases in the $\Delta clrB$ mutant (AA9, GH6, GH7 and GH5.5 classes, shown in blue).

Table 3. Conserved orthologous genes in the CLR-2 and ClrB regulons

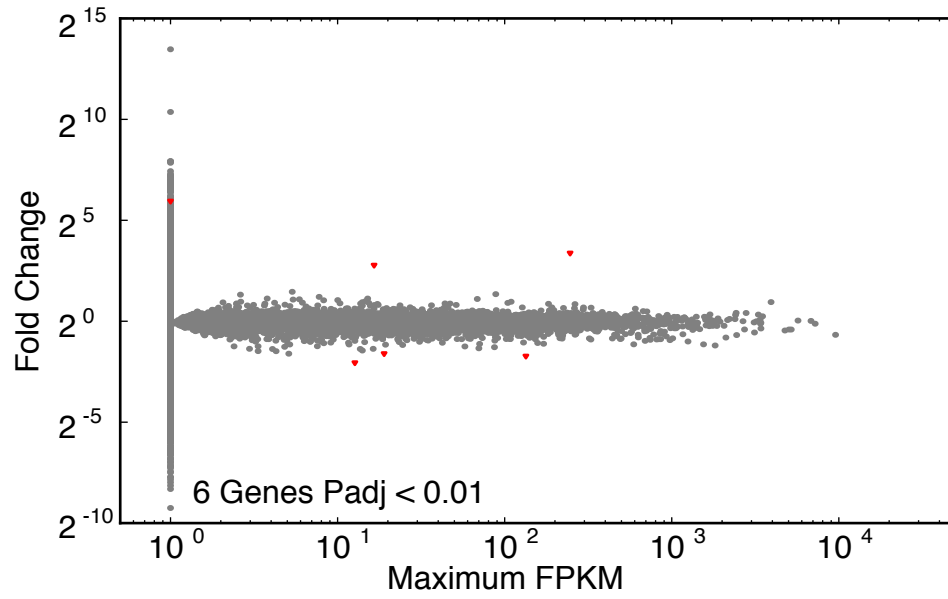
NCU#	Gene and Annotation	CAZy	AN#	Gene
Orthologs with conserved dependence on <i>clr-2/clrB</i>				
NCU07340	<i>cbh-1</i> cellobiohydrolase 1	GH7	AN5176	<i>cbhA</i>
			AN0494	<i>cbhB</i>
			AN1273	<i>cbhD</i>
NCU09680	<i>gh6-2</i> cellobiohydrolase 2	GH6	AN5282	<i>cbhC</i>
NCU00762	<i>gh5-1</i> endoglucanase 2	CGH5.5	AN1285	<i>eglA</i>
NCU07760	<i>gh61-2</i> LPMO	AA9	AN3860	
NCU00206	<i>cdh-1</i> cellobiose dehydrogenase	AA3	AN7230	
NCU07143	<i>lac-2</i> aldonolactonase		AN2112	
NCU08412	endo- β -1,4-mannanase	GH5.7	AN6427	<i>manC</i>
NCU03328	<i>gh61-6</i> LPMO	AA9	AN6428	
NCU02344	<i>gh61-12</i> LPMO	AA9		
NCU04952	<i>gh3-4</i> secreted β -glucosidase	GH3	AN2828	<i>bglL</i>
NCU09664	<i>ce5-3</i> acetylxylan esterase	CE5	AN5309	
NCU00449	broad specificity endoglucanase	GH131	AN7598	
NCU09764	broad specificity endoglucanase	GH131		
NCU06326	pectate lyase 1	PL1	AN0741	<i>pelA</i>
NCU03181	acetylxylan esterase	CEnc	AN5320	
NCU00972	arabinogalactan endo-1,4- β -galactosidase	GH53	AN5727	<i>gala</i>
NCU04997	<i>gh10-3</i> xylanase	GH10	AN7401	<i>xlnE</i>
NCU07705	<i>clr-1</i> cellulose degradation		AN5808	<i>clrA</i>
NCU05598	<i>asd-1</i> rhamnogalacturonase B	PL4	AN7135	<i>rglA</i>
NCU09976	<i>ce12-1</i> rhamnogalacturonan acetylcetase	CE12	AN2528	
NCU08176	<i>ply-2</i> pectate lyase A	PL3	AN6106	<i>plyF</i>
Orthologs with conserved modulation by <i>clr-2/clrB</i>				
NCU00130	<i>gh1-1</i> cytoplasmic β -glucosidase	GH1	AN10124	
NCU00801	<i>cdt-1</i> cellodextrin transporter		AN2814	
NCU04854	<i>gh7-2</i> endoglucanase	GH7	AN3418	<i>eglB</i>
NCU00890	<i>gh2-1</i> β -mannosidase	GH2	AN3368	<i>mndB</i>
NCU09416	cellulose-binding GDSE lipase/acylhydrolase	CE16	AN6422	

The genes identified as ClrB dependent in *A. nidulans* were compared to those identified as CLR-2 dependent in *N. crassa* in response to Avicel [39]. Twenty-eight orthologous pairs of genes with conserved regulation by CLR-2/ClrB were identified (Table 1). All of these genes encode proteins with some functional annotation (no hypothetical), though some are based solely on homology to enzymes with limited experimental data. The majority encoded secreted proteins that represent the conserved core of enzymes for cellulose degradation, including well-studied endoglucanases, exoglucanases and β -glucosidases, as well as the newly characterized LPMOs. The conserved list also included xylan and pectin degrading enzymes, including NCU00449 and NCU09764, orthologs to AN7598 (GH131; broad specificity glucanase), and which have sequence similarity to a characterized glucuronan lyase from *T. reesei* [187]. A conserved CLR-2/ClrB target NCU07143/AN2112 is homologous to a recently characterized extracellular aldonolactonase from *M. thermophila* [49] and was identified in *N. crassa* secretome studies on Avicel [127]. The expression of genes encoding several intracellular enzymes had conserved modulation by CLR-2/ClrB, including an intracellular β -glucosidase (NCU00130) and a β -mannosidase (NCU00890), as well as the cellodextrin transporter, *cdt-1* (NCU00801) [163]. Finally, one transcriptional regulator was conserved, *clr-1/clrA*.

Limited Conservation of gene regulation by ClrA

In parallel with our experiments on *clrB*, we also sequenced RNAseq libraries from a $\Delta clrA$ strain of *A. nidulans* transferred to Avicel. I analyzed differential expression between the $\Delta clrB$ and WT control libraries on Avicel and found evidence for a drastically reduced regulon. Genes with differential expression in the $\Delta clrA$ and $\Delta clrB$ mutants are shown in Figure 28. Only two genes and *clrA* itself were significantly down regulated in the $\Delta clrA$ mutant, with three genes showing slightly higher abundance of borderline statistical significance. The downregulated genes were AN11077 (hypothetical protein with a predicted acetyltransferase domain) and AN5282, cellobiohydrolase *cbhC*, a major cellulase and one of two *A. nidulans* hydrolases orthologous to *N. crassa cbh-2*. Thus it would appear the mild growth and enzyme activity defect that the $\Delta clrA$ mutant displayed on Avicel was due to the reduced secretion of a single key enzyme.

A. WT *A. nidulans* vs $\Delta clrA$



B. WT *A. nidulans* vs $\Delta clrB$

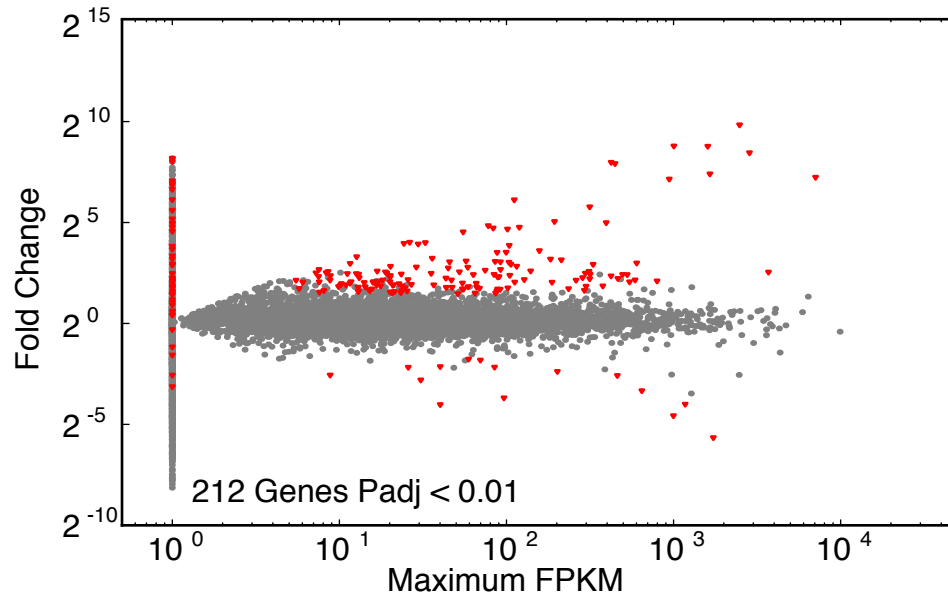


Figure 28. Differential expression in *clrA* and *clrB* deletion strains.

Transcript abundance in the WT strain vs mutant strain (fold change) is plotted against maximum abundance in either condition. Genes with an adjusted P value lower than 0.01 highlighted in red.

Constitutive expression of *clrB* and *clr-2* in *A. nidulans* is not sufficient for cellulase secretion

My earlier experiments in *N. crassa* showed it was possible to produce and secrete cellulases under non-inducing conditions by heterologous regulation of the transcriptional regulator, CLR-2 (Chapter 4). To test whether constitutive of *clr-2* homologs was sufficient for inducer-free cellulase production and secretion in other ascomycete fungi, Yi Xiong inserted the *clrB* and *clr-2* coding sequences into the *pyrG* locus of the *A. nidulans* Δ *clrB* strain under the regulation of the constitutive *gpdA* promoter (strains *Pgpd:clrB* and *Pgpd:clr-2*) and under the inducible *alcaA* promoter (strains *Palca:clrB* and *Palca:clr-2*). GFP tagged versions of these constructs expressed well and localized to the nucleus (Figure 29A). Strong expression of *N. crassa* CLR-2 was detrimental to cell growth (Figure 29B). Neither Yi Xiong nor I was ever able measure any substantial induction of cellulases in non-inducing conditions with any of these *clrB* or *clr-2* constitutive or inducible strains. Representative data for these experiments are shown in Figure 30. In Figure 30A, Yi Xiong performed RT-PCR on *Pgpd:clrB* and *Pgpd:clr-2* after a shift to glucose, no carbon and Avicel media, but saw no substantial induction of major cellulase genes. In Figure 30B, I measured endoglucanase activity in WT, Δ *clrB*, and *Pgpd:clrB* strains shifted to Avicel or sucrose media and measured no substantive enzyme activity from the *Pgpd:clrB* strain in non-inducing conditions.

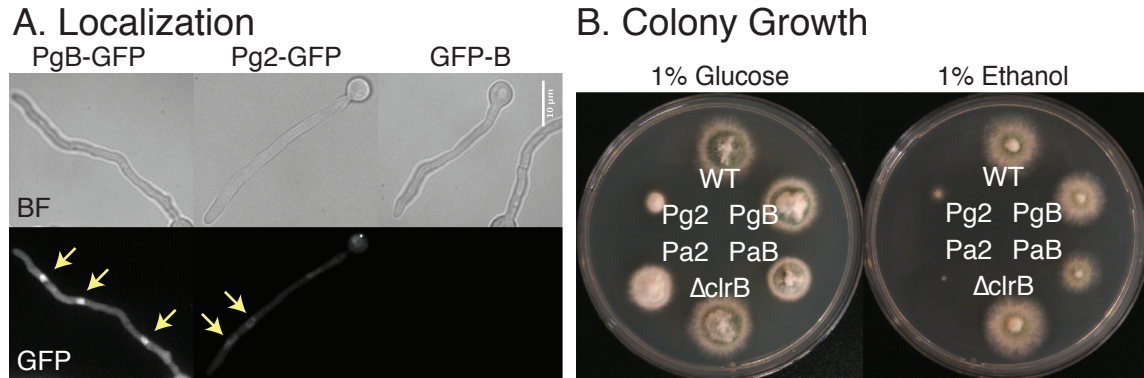
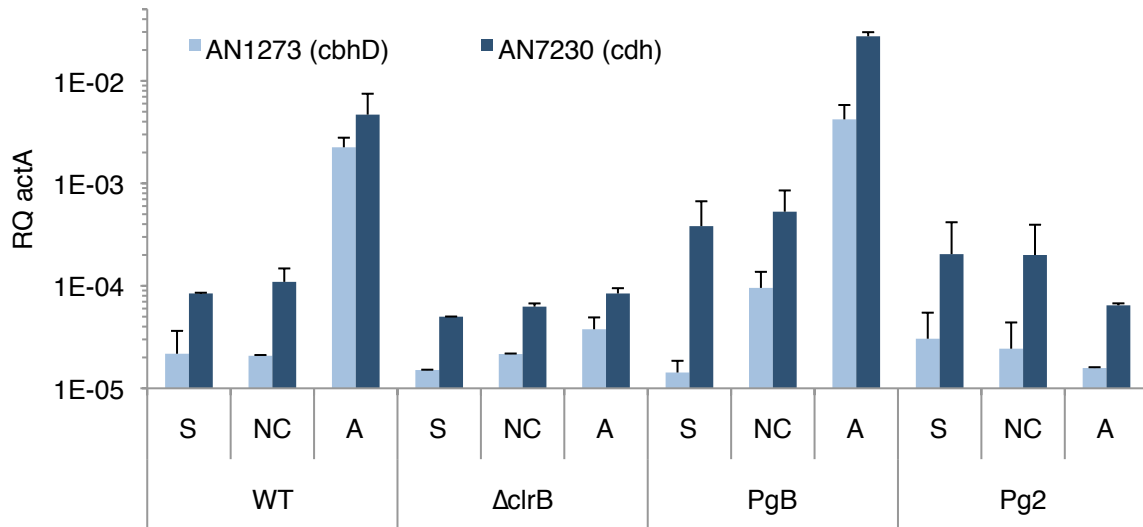


Figure 29. Growth phenotypes of *A. nidulans* *clrB* and *clr-2* misexpression mutants.

(A) Localization of GFP tagged constructs in cells germinated in liquid glucose medium. GFP labeled *ClrB* under native promoter (strain GFP-B) had no detectable fluorescence. GFP labeled *ClrB* or *CLR-2* driven by the *gpdA* promoter (strain GFP-PgB and GFP-Pg2) accumulated in nuclei, which were clearly distinguishable by their shape and prominent nucleoli. Arrowheads indicate nuclei.

(B) WT and mis-expression strains grown 3 days on 1% glucose or 1% ethanol plates at 30°C. PgB and Pg2: constitutive expression of *clrB* and *clr-2*, respectively, by the *gpdA* promoter in Δ *clrB* background. PaB and Pa2: ethanol inducible expression of *clrB* and *clr-2*, respectively, by the *alcaA* promoter in Δ *clrB* background.

A. RT-PCR



B. Endoglucanase Activity

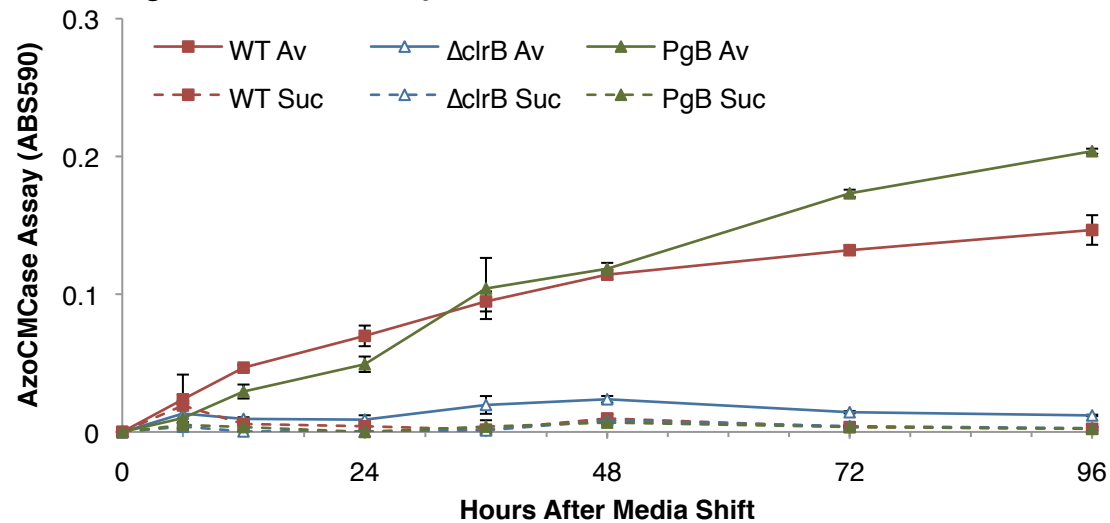
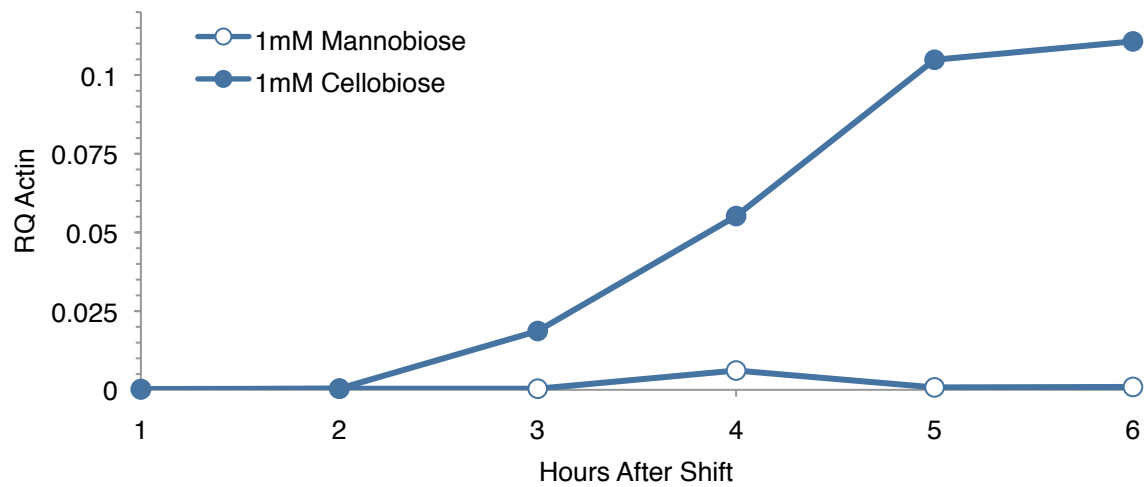


Figure 30. Cellulase induction in *A. nidulans* *clrB* and *clr-2* misexpression mutants.

(A) Transcript abundance of two major cellulases (*cbhD* and *cdh*) in *A. nidulans* WT versus strains with constitutive expression of *clrB* (PgB) or *clr-2* (Pg2) when shifted to glucose [G], no carbon media [NC] or Avicel [A] for 6 hrs. Transcript abundance relative to *actA* was measured with real-time quantitative RT-PCR.

(B) A time course of CMCase activity in WT, Δ *clrB* and PgB strains following a shift to either Avicel or sucrose.

A. cbhC AN5282 Transcript



B. cbhD AN1273 Transcript

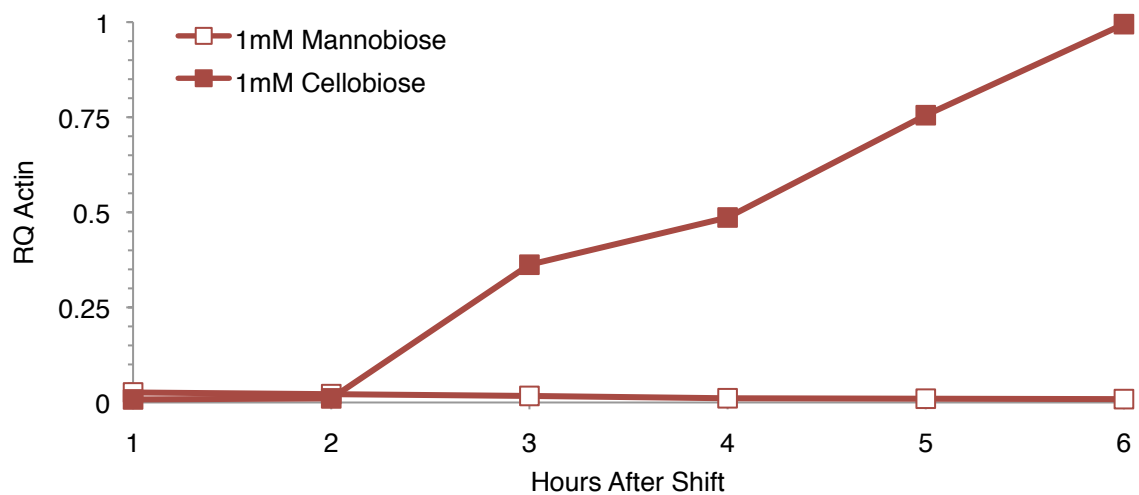


Figure 31. Timecourse of cellulase induction in *A. nidulans* FGSC A4 shifted to 1mM Cellobiose or 1mM Mannobiose.

n=1

***A. nidulans* cellulases are induced by cellobiose but not by mannobiose**

Coincident with the above experiments, Ogawa et. al. characterized ManR, a ZnBNC regulator of mannan utilization in *Aspergillus oryzae* that also regulates cellulases and at least one xylanase in response to mannobiose, but not cellobiose [112,180]. Although initially unrecognized by Ogawa et al as such, ManR is in fact an ortholog of CLR-2/clrB (Figure 22). Their results together with my own findings that Avicel retains traces of plant cell wall polysaccharides besides cellulose, and that endomannanases are co-regulated with cellulases by ClrB, raised the possibility that *A. nidulans* may also regulate its cellulases in response to mannobiose as opposed to cellobiose. To establish which oligosaccharide was the primary inducer of cellulases in *A. nidulans*, I transferred glucose grown cultures of WT *A. nidulans* to 1mM 1-4,β-D-cellobiose or 1mM 1-4,β-D-mannobiose and tracked induction of two major cellulases over a 6 hour time course. Cellobiose induced both *cbhC* AN5282 (*clrA* and *clrB* dependent) and *cbhD* (*clrB* dependent only) but mannobiose did not (Figure 31). A second transfer experiment with WT, Δ *clrB* and *PgpdA:clrB* strains showed that this induction by cellobiose was *clrB* dependent (Figure 32), suggesting that my results on Avicel were reflective of an inducing system responding primarily to released cellobiose and not mannobiose. Unexpectedly, in the *PgpdA:clrB* strain, *cbhD* was strongly induced on mannobiose as well as cellobiose, suggesting at least some cross-reactivity of mannobiose in the cellobiose sensing pathway.

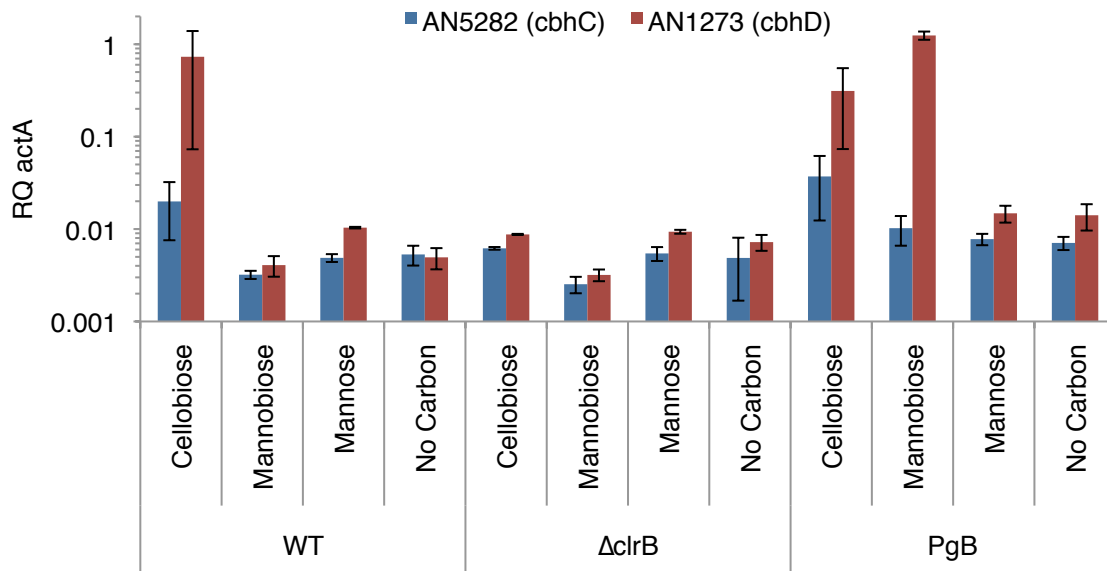


Figure 32. Induction of *A. nidulans* cellulases in response to 1,4-β-D-cellobiose and 1,4-β-D-mannobiose.

Transcript abundance 6 hours after a shift to 2mM sugars or fresh minimal media with no added carbon. Individual columns are independent biological replicates.

Discussion

Homologs of *clr-1* and *clr-2* are present in genomes of filamentous ascomycete fungi, particularly those in association with plants, suggesting conserved regulatory mechanisms associated with genes within the Avicel regulon. Induction of most cellulase genes required *clrB* but not *clrA*. However, Δ *clrA* was deficient for cellobiose utilization and in the induction of *cbhC* a major cellobiohydrolase with confirmed activity on cellodextrins [188]. Whereas in *N. crassa* *clr-2* expression was dependent upon *clr-1*, in *A. nidulans* *clrA* expression was dependent upon *clrB*. Thus, ClrA would appear to have a conserved role in cellulose and cellobiose utilization through regulation of *cbhC* expression but a dramatically smaller role in general cellulase regulation and cellobiose sensing as compared to *N. crassa*.

ClrB was essential for significant induction of cellulases in response to Avicel. Dozens of other enzymes involved in plant cell wall degradation and alternative sugar utilization were also dependent on functional ClrB, as was the case in with CLR-2 in *N. crassa*. Given the intricate associations of these polymers in nature, a simple explanation for the adaptive advantage of maintaining such a large and biochemically diverse regulon under control of cellobiose induction might be that cellulose sensing makes an effective proxy for hemicellulose and pectin sensing. Yet for the most part, the genes necessary to utilize those polymers (e.g the pentose utilization pathway) were not regulated by CLR-2/ClrB.

I hypothesize that the diversity of secreted enzymes in the CLR-2/ClrB regulon reflects the synergistic effects that hemicellulases and pectinases have on the deconstruction of cellulose complexed with other polysaccharides. This effect has been shown in simplified systems, but for many enzymes such synergism was highly substrate dependent [189]. Yet in this study, I observed twelve orthologous sets of secreted enzymes with predicted function on non-cellulose polymers with a conserved co-regulation with cellulases in two evolutionarily divergent species with very different ecological niches. Such conservation in co-regulation among orthologs might then suggest that these enzymes have a general synergistic value in cellulose deconstruction across a broad spectrum of substrates. Thus these enzymes and closely related paralogs may make excellent candidates for inclusion in customized cellulase mixtures for generalized applications. Conversely, the enzymes under CLR-2 regulation by not ClrB regulation, or vice-versa, may make excellent candidates for specialized enzyme blends for more targeted applications.

Constitutive expression of *clr-2* in *N. crassa* resulted in inducer independent production of plant cell wall degrading enzymes. In *A. nidulans*, constitutive expression of *clrB* produced no detectable enzyme activity under non-inducing conditions. RT-PCR analyses revealed a measurable, but small increase in transcript abundance for a cellobiose dehydrogenase (AN7230) and no increase for cellobiohydrolase *cbhD* (AN1273). I therefore infer that there is a fundamental difference in cellulase gene induction between *Neurospora* and *Aspergillus* species. Whereas CLR-2 is fully competent to drive transcription of its targets in non-inducing conditions in *N. crassa*, ClrB requires other factors for any significant effect.

An unexplained result of these experiments was the finding that over-expression of ClrB with the strong constitutive *gpdA* promoter sensitized *A. nidulans* to mannobiose, allowing the induction of cellulases by mannobiose at concentrations at which only cellobiose induced cellulases in the WT strain. Recently, the *A. oryzae* CLR-2 ortholog (now named ManR) was identified as transcriptional regulator of β -mannan utilization [112]. Ogawa et al performed transcriptional profiling of ManR mutants on glucomannan and reported that it regulated several mannanases and at least one endoglucanase. RT-PCR experiments on various oligosaccharides and sugars suggested that mannobiose and not cellobiose stimulated ManR dependent induction of these mannanases. Further examination of their profiling data on glucomannan (GEO accession GSE40590) reveals that ManR also regulated cellobiohydrolases and at least one xylanase though this was not reported in the initial publication. A brief follow-up paper [180] reported that ManR regulates cellulases in response to cellulose (Avicel), but this result is somewhat at odds with their earlier experiments showing no substantial induction of ManR targets by cellobiose. They also reported that carboxy-methyl-cellulose (CMC) failed to induce significant cellulase expression, though they showed no data from those experiments. In my opinion, the simplest explanation for these results would be that contaminating mannans in Avicel are sufficient to induce cellulase expression via ManR in *A. oryzae*.

Regardless of whether cellulose can or cannot induce cellulases in *A. oryzae*, there is clearly a divergence of function for ManR and ClrB with regards to inducing molecules. Given the structural similarity of mannobiose and cellobiose, and the propensity for both to be released from hydrolyzed plant cell wall material, some evolutionary plasticity between mannobiose and cellobiose sensing is perhaps unsurprising. A more surprising finding was that at least in the primacy of cellobiose in cellulase sensing, *A. nidulans* more closely resembles *N. crassa* than the much more closely related species, *A. oryzae*. Likely, induction of cellulases in response to cellobiose was the ancestral signaling pathway, but this has shifted to induction by mannobiose in *A. oryzae*, though more experiments in other species would be required to confirm that hypothesis.

Taken together, these data demonstrate both conserved and divergent aspects of CLR-1 and CLR-2 function across these most divergent filamentous ascomycetes. The finding that both ClrA and ClrB participate in regulation of cellulases in response to cellulose, demonstrates that similar functions are possible if not likely for other orthologs in diverse species. The radically reduced role of ClrA, the failure of ClrB to stimulate cellulase secretion in inducer free conditions, and the finding that ClrB and ManR have opposing responses to induction by cellobiose versus mannobiose, imply extensive rewiring of regulatory circuits controlling cellulase expression between filamentous fungal species. Reliably transferring knowledge between species will thus be impossible until we can understand how such functional changes occurred. I submit that the most likely route to that understanding is to gain deeper knowledge of the molecular mechanisms by which inducing molecules stimulate active transcription by CLR-1, ClrB and other fungal

cellulose response regulators. Experiments to elucidate those mechanisms are the topic of Chapter 6 and Chapter 7.

Acknowledgements

The experiments in this chapter were only made possible through a terrific collaboration with lab member Yi Xiong. Yi (with some help from Darae Jun under Yi's direct supervision) constructed all of the *A. nidulans* strains used here including the $\Delta clrA$ and $\Delta clrB$ strains, deletions for additional paralogous genes (which showed no phenotype and thus were not discussed here) and strains bearing both *clrB* and *clr-2* both tagged and untagged under multiple promoters. Yi performed the RT-PCR experiments in Figure 23 and Figure 30 and the microscopy in Figure 30. Yi and I constructed the *A. nidulans* RNAseq libraries discussed in this chapter together.

James Craig performed a phylogenetic analysis for *clr-1* and *clr-2* for the original publication on CLR-1/ClrA and CLR-2/ClrB [39]. The phylogenetic trees I created for Figure 21 and Figure 22 are an update on that analysis built from a local BLAST database constructed by me for another project with Areejit Samal. For that project Areejit has conducted extensive literature reviews and bioinformatic analysis on characterized enzymes in filamentous fungi and my phylogenetic analysis has benefited from the context he has provided.

Published material

Figure 23 was originally published in:

Coradetti ST, Craig JP, Xiong Y, Shock T, Tian CG, et al. (2012) Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *Proc Natl Acad Sci U S A* 109: 7397-7402.

Figure 1, Figure 26, Figure 29, and Figure 30 were originally published in and Figure 24, Figure 25, and Figure 27 were revised or reconstructed from data originally published in:

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.

Methods

Strains

The wild type *N. crassa* reference strain and background for mutant strains was OR74A (FGSC 2489) [161]. The deletion strains for *clr-2* (FGSC 15835) were obtained from the Fungal Genetics Stock Center (<http://www.fgsc.net/>) [190].

The *A. nidulans* reference strain was FGSC 4A. Gene deletions in *A. nidulans* were carried out by transforming FGSC A1149 (*pyrG89*; *pyroA4*; *nkuA::argB*) by established protocols [191]. Transformants were crossed to L01496 (*fwA1*, *pyrG89*, *nicA2*, *pabaA1*, from Berl R. Oakley) to remove *nkuA::argB* and *pyroA4*. The *clrB* mis-expression strains were constructed by transforming a $\Delta clrB$ mutant ($\Delta clrB::pyrG$, *pyroA4*, *pyrG89*) with a DNA fragment consisting of the *Aspergillus fumigatus* *pyroA*

gene, the *gpdA* [192] or the *alcA* [193] promoter and the coding region and 3' UTR of either *clrB* or *clr-2* inserted into the *pyrG* locus. Transformants were selected for pyrodoxine prototrophy [194]. The DNA fragments were generated by fusion PCR and transformation was carried out as described previously [195].

Culture conditions

N. crassa cultures were grown on Vogel's minimal medium (VMM) [119]; carbon sources were 2% w/v unless otherwise noted. Strains were inoculated into 3 mL VMM slants and grown at 30°C in the dark for 48 hrs, then at 25°C in constant light for 4-10 days to stimulate conidia production. Conidia were inoculated into 100 mL of liquid media at 10⁶ conidia/mL and grown at 25°C in constant light and shaking (200 rpm).

A. nidulans cultures were grown on *A. nidulans* minimal medium (MM) [196]. Carbon sources were 2% w/v unless otherwise noted. Conidia were grown on sucrose MM slants, then inoculated into liquid media at 3x10⁶ conidia/mL and grown at 37°C in constant light and shaking (200 rpm). All liquid cultures were 100mL in 250mL shakeflasks, except for experiments on mannobiose and cellobiose, for which 3mL cultures in 24-well plates were used due to the relative scarcity of purified mannobiose.

These culture conditions were optimal for cellulose utilization by each species in pilot experiments and are in line with previous studies on each species. The higher temperature and inoculum used for *A. nidulans* allowed for comparable biomass accumulation to *N. crassa* cultured over the same period, which was chosen as the best practical basis comparison between the two species.

Media shift experiments

Cultures were grown 16 hrs on sucrose (or 1 % glucose for the *A. nidulans* RNAseq experiments), centrifuged at 2000g for 10 min and washed in VMM or MM [119] without a carbon source. Mycelia were re-suspended in 100 mL VMM or MM with 2% carbon source (sucrose, cellulose (Avicel® PH-101, Sigma Aldrich, MO) or with no carbon source added. For 3mL cultures, strains were grown 24 hours on 1% glucose, then media was aspirated from the culture and replaced with 3mL MM with no added carbon. After a 10-minute incubation this no carbon wash was aspirated and the final media was added. Total RNA was extracted as described in [143] at 4 hrs after shift for *N. crassa* cultures and 6 hours after shift for *A. nidulans* cultures. These time points were selected as the earliest points of full cellulase gene induction by preliminary RT-PCR time course experiments. For enzyme activity assays, culture supernatants were sampled over a 96 hour period, centrifuged at 3000g twice to remove mycelia and stored at 4°C for prior to analysis. At the end of the time course experiments, fungal biomass was harvested, dried and weighed. The relative proportion of mycelia to residual Avicel was calculated by solubilizing fungal biomass in boiling acetic nitric reagent as in [181], washing the residual mass thoroughly, then drying and reweighing.

One mutant strain (*PgpdA:clr-2*) required alternate media switch conditions due to low growth and conidiation. *PgpdA:clr-2* conidia were inoculated in 3 mL 1% glucose MM and grown 3 days at 37°C in constant light and shaking (200 rpm).

Cultures were then washed in MM with no carbon and switched to 3 mL MM with 2% Avicel, 2% glucose or no carbon source. Other *A. nidulans* strains used in RT-PCR experiments were cultured similarly, but with only 24 hours pre-growth in 1% glucose to achieve similar biomass to Pg2.

RT-PCR

RT-PCR was performed with the EXPRESS One-Step SYBR® GreenER™ Kit, with premixed ROX (Invitrogen #11790-200) using the standard cycling protocol in a 20µl reaction with 5ng total RNA as the input. Amplification was carried out on a CFX Connect™ Real-Time PCR Detection System (Biorad). Primers were tested for similar efficiencies and specificities were checked with melt-curve analysis. The ΔCQ method was used for transcript abundance with actin (NCU04713 or AN6542) as the reference gene. Primer sequences are listed in Appendix B.

RNA sequencing and transcript abundance

Libraries were prepared with standard protocols from Illumina Inc (San Diego, CA) and sequenced on the Illumina Genome Analyzer IIx and HiSeq 2000 platforms. Sequenced libraries were mapped against predicted transcripts from the *N. crassa* OR74A genome [118] or the *A. nidulans* FGSC A4 genome [197] with Tophat v2.0.4 [133]. Transcript abundance (FPKM) was estimated with Cufflinks v2.0.2 [171] mapping against reference isoforms. Profiling data are available at the NCBI gene expression omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/_SeriesRecordGSE44100). Raw counts for reads mapping to unique exons were tallied with HTSeq [135] and used as inputs for the DESeq software packages. Genes with a multiple-hypothesis-adjusted p-value below 0.05 and at least 4-fold induction on Avicel were selected as belonging to the Avicel regulon and used for further analyses.

Orthologous gene pairs were taken from the fungal orthogroups repository (<http://www.broadinstitute.org/regev/orthogroups/>) and supplemented by a best reciprocal blast search for recently added or revised gene models in the *N. crassa* and *A. nidulans* genomes. In a few cases, Avicel induced genes had an ortholog listed in the repository with very low expression and a second close homolog with similar sequence conservation which showed specific induction. In those cases the induced homolog was selected as the likely ortholog. For each species, specifically induced genes without a clear ortholog were used as BLAST queries against the other genome. The most significant alignment (E-value < 10⁻⁵⁰) was taken as a possible paralog. In some cases this procedure matched several paralogs in one species to a single gene in the other. These pairings were treated as distinct entries for clustering analysis. Genes with very low expression (FPKM < 5 on Avicel) were filtered from the cross-species clustering analysis. For clustering of orthologous and paralogous pairs of genes, FPKM were log transformed and normalized on a per gene basis such that the highest value recorded in any sample for either member of the pairing was set to 1. The normalized values were then scaled and centered to set 1 FPKM to -1 and the maximal FPKM to 1 for clustering. Average linkage clustering was performed with Cluster 3.0 [159] using Pearson correlation as the similarity metric.

For clustering of WT and mutant *clr-2/clrB* libraries in each species, FPKM were log transformed then normalized and centered on per gene basis with Cluster 3.0 standard settings such that values from each gene ranged from -1 (minimum/uninduced value) to 1 (maximum/induced value) and clustered with K-means clustering with Euclidean distance as the similarity metric.

Enzyme activity assays

Endoglucanase and xylanase activity of culture supernatants was assayed with Remazolbrilliant Blue R conjugated CMC and xylan (birchwood) kits from Megazyme (Wicklow, Ireland). Total protein was determined with the Bradford assay (BioRad). Cellulase activity was measured as previously reported [143], with minor adjustments. A 200µl aliquot of culture supernatant was added to a final volume of 1 mL 1% Avicel in 50mM sodium acetate and incubated with shaking at 37C overnight. Released glucose and cellobiose levels were measured by HPLC.

Phylogenetic trees

Phylogenetic trees of *clr-1*, *clr-2* and *mcm-7* homologs were built by searching a local database of the most complete and well-annotated fungal genomes available for all matches with expectation values below 1. All gathered homologs were aligned with MAFFT [198] and a first draft tree was built with Fasttree [199]. The draft tree was then pruned for likely orthologs and a suitable outgroup. The subtree was then re-aligned with MAFFT and a final tree was built with RaxML using the JTT substitution model and building 20 trees and bootstrapping with 100 resampled trees.

Variation and statistical significance

All error bars on charts and figures represent standard deviation from biological triplicate samples unless otherwise noted.

Chapter 6

A Selective Screen for Genes Involved in Cellobiose Sensing

Abstract

Despite the identification and characterization of multiple transcription factors regulating the expression of cellulases and other plant cell wall degrading enzymes in response to environmental sugars, functional conservation of these transcription factors across species remains enigmatic. Characterization of the sugar sensing networks these transcription factors participate in may bring the deeper understanding of enzyme regulation how fungi interact with their environment. Such understanding will be required to rationally engineer expression of cellulases and other plant cell wall degrading enzymes across diverse species for diverse applications. To investigate the cellobiose-sensing network in *N. crassa*, I conducted a selective screen for constitutive cellulase producing mutants with an antibiotic under cellulase regulation. Bulk segregant analysis of an isolated mutant identified an uncharacterized gene with a role cellulase regulation.

Introduction

To date the isolation and characterization of genes involved in plant cell wall deconstruction has been relatively laborious, either involving long running mutant screens in which strains of interest must be isolated by hand [39,79,108,112] or through identification of likely regulators through a combination of similarity to known regulators and expression analysis [105,109]. The former approach is limited by the laborious phenotyping of large numbers of strains, which are themselves laborious to create and/or identify and is largely limited to identification of loss of function mutations. The latter approach will only identify regulating components with similar sequence and expression patterns to those that have already been identified. Aro et al [86] took an innovative high throughput approach to this problem, using a yeast one-hybrid system for an efficient search of *T. reesei* genomic libraries for transcriptional regulators in *S. cerevisiae*, selecting for constructs that could induce transcription of selectable marker under control of the *T. reesei cbh1* promoter. That approach was limited to regulators that were well expressed and functional in a heterologous system.

Furthermore, many of the above studies were targeted to transcription factors only. As we have seen in previous chapters, a general focus on transcription factors has lead to the appreciation that dozens of transcription factors, mostly in the ZnBNC superfamily, are involved in regulation of plant cell wall degrading enzymes. No mechanism of induction has been described for any of these newly described transcription factors. The underlying molecular interactions that mediate the sensing of plant cell wall carbohydrates remain completely unknown.

At the same time our ability to transfer findings in one filamentous fungal species to another is poor, with known regulators showing wildly varying effects in even closely related species [88,93,94,200] even as they have some conserved

function across remarkably distant species [87,169]. A deeper understanding of the mechanisms of induction may improve our ability to predict functional conservation by assessing sequence conservation of a complete multi-protein system as opposed to individual transcription factors.

In this chapter I describe the construction and implementation of a selectable strain for cellulase induction. I demonstrate a mutagenesis, selection, verification, and identification pipeline for novel gain-of-function mutants with respect to cellulase induction and identify new genes with previously unknown function in cellulase induction that are likely involved in cellobiose sensing.

A selectable strain for cellulase induction

To build an optimal selectable strain for cellulase induction, I began with the RNAseq data that my coworkers and I have collected on diverse carbon sources. A selectable marker for cellulase induction should be expressed well on inducing media, but have extremely low expression in repressing and non-inducing media. Of particular interest is expression on xylan, identified in Chapter 2 as the non-inducing media with the best combination of robust growth and weak repression of cellulases. The promoter regions for *cbh-1* NCU07340 orthologs have been used as the prototypical cellulase promoter in other studies [86]. As shown in Figure 33, however, NCU07340 has substantial expression in non-inducing conditions. So does NCU00762, the major endoglucanase and all the major β -glucosidases such as NCU08755. NCU08042, *clr-2*, has a switch-like expression pattern across potentially useful media conditions, but disrupting or replacing *clr-2* would severely disrupt cellulase induction. The most promising gene for use as a selectable marker was NCU01050, a highly inducible LMPO with extremely low expression in non-inducing conditions. A possible secondary locus for a selectable maker was NCU00206, cellobiose dehydrogenase, which had a similar expression profile to *clr-2* but with higher expression in inducing conditions.

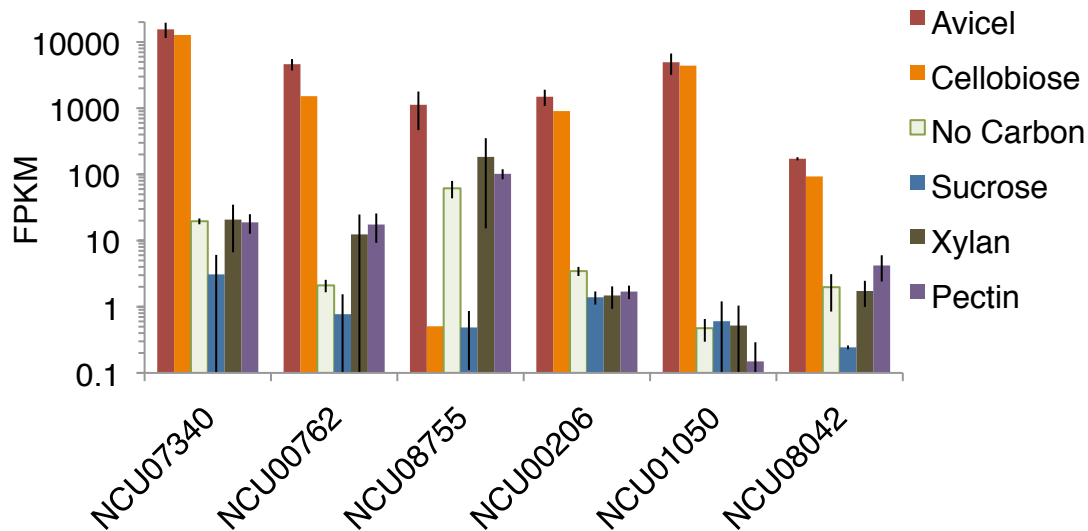


Figure 33. Expression profiles of candidate genes for selectable markers

FPKM on various carbon sources. Avicel, no carbon and sucrose data from [169]. Xylan and Pectin data from [201]. Cellobiose ($\Delta\beta$ G mutant) from [128]

In order to most fully and faithfully capture all aspects of cellulase regulation, including promoter sequence, untranslated mRNA and local chromosomal context, I replaced the NCU01050 open reading frame at the native locus with the open reading frame of hygromycin B phosphotransferase from *E. coli* [202]. As standard *N. crassa* transformation procedures employ media containing repressing carbon sources [161], this selectable marker under cellulase regulation was not usable to select for transformed conidia. Thus, I also included the resistance-encoding gene (*ble*) from the bacterial transposon Tn5 for resistance against phleomycin [203] downstream of the NCU01050 UTR (Figure 34).

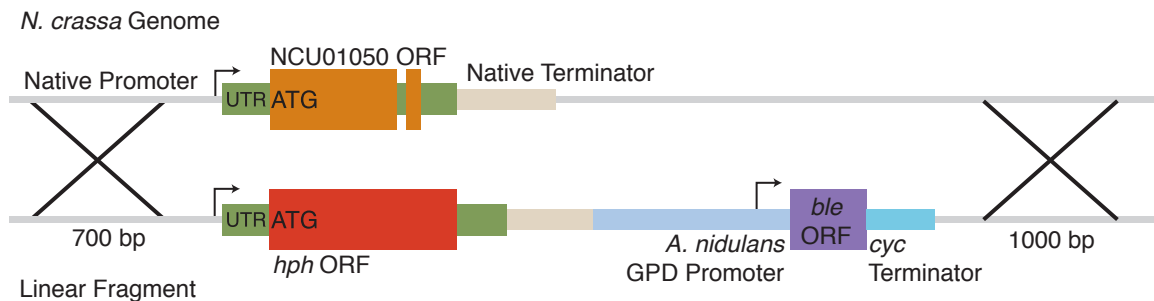
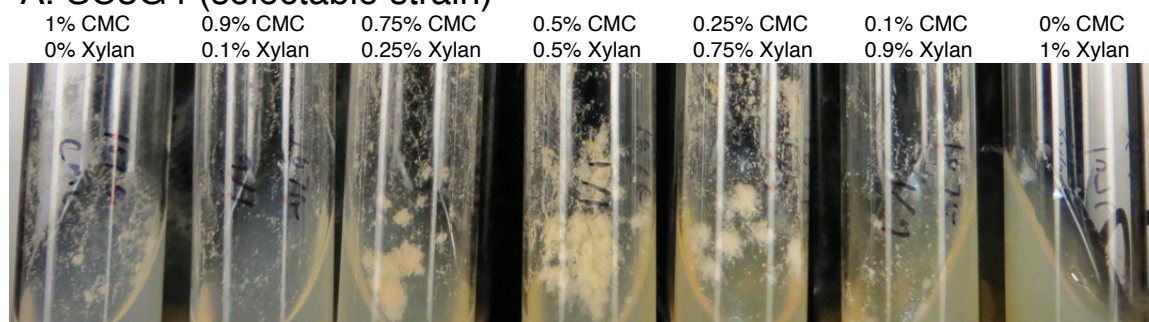


Figure 34. Selectable construct for cellulase induction.

I transformed a $\Delta mus-52$ mutant strain of *N. crassa* with the selectable construct and selected on phleomycin media. Twelve transformants were confirmed for construct integration by PCR and tested for positive growth on CMC (inducing) hygromycin media and negative growth on xylan (noninducing) hygromycin media. Several transformants showed the desired selectable behavior. One mutant was crossed back to WT FGSC 2489 and progeny were selected for glufosinate sensitivity (WT for *mus-52*) and phleomycin resistance. A *mat a* strain (SC3G4) and a *mat A* strain (SC3G5) were then selected for mutagenesis and backcrossing. Growth of the *mat a* Pncu01050:*hph* strain SC3G4 on a gradient of CMC/xylan concentrations with hygromycin is shown in Figure 35. A growth condition of 0.5% xylan was selected a neutral (non-repressing, non-inducing) growth condition.

A. SC3G4 (selectable strain)



B. Hyg+ Control

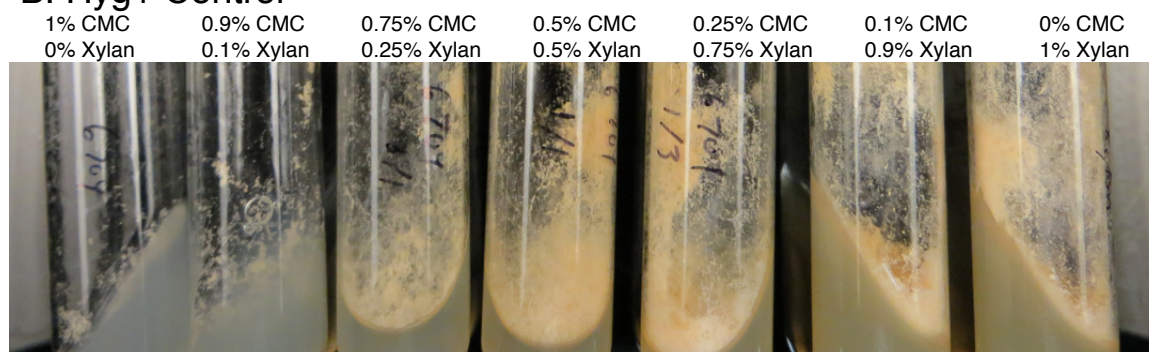


Figure 35. Growth of the cellulose inducible selection strain on CMC/xylan hygromycin media
Equal inocula of conidia of (A) SC3G4, the selectable strain with *hph* under control of the NCU01050 promoter and (B) a constitutively hygromycin resistant strain were grown in slants with 200µg/mL hygromycin B and differing concentrations of CMC (inducing cellulose analog) or xylan (neutral substrate).

Mutagenesis and mutant selection

I subjected selectable strain SC3G4 to mutagenesis with N-methyl-N-nitro-N-nitrosoguanidine (NTG) with protocols adapted from Mallng et al [204]. In a pilot experiment, 1×10^8 conidia were exposed to mutagenesis and plated on xylan hygromycin media. Approximately 200 hygromycin resistant strains were picked from the selection plates (Figure 36) and transferred to xylan hygromycin slants to confirm growth. Approximately 100 strains were viable in this confirmation round. Those strains were then grown on xylan for three days and assayed for endoglucanase activity. Out of the pilot round of mutagenesis, only one strain, M13, had verified constitutive cellulase activity. The other mutants are apparent hygromycin resistant mutants independent of induction of the engineered *hph* construct.

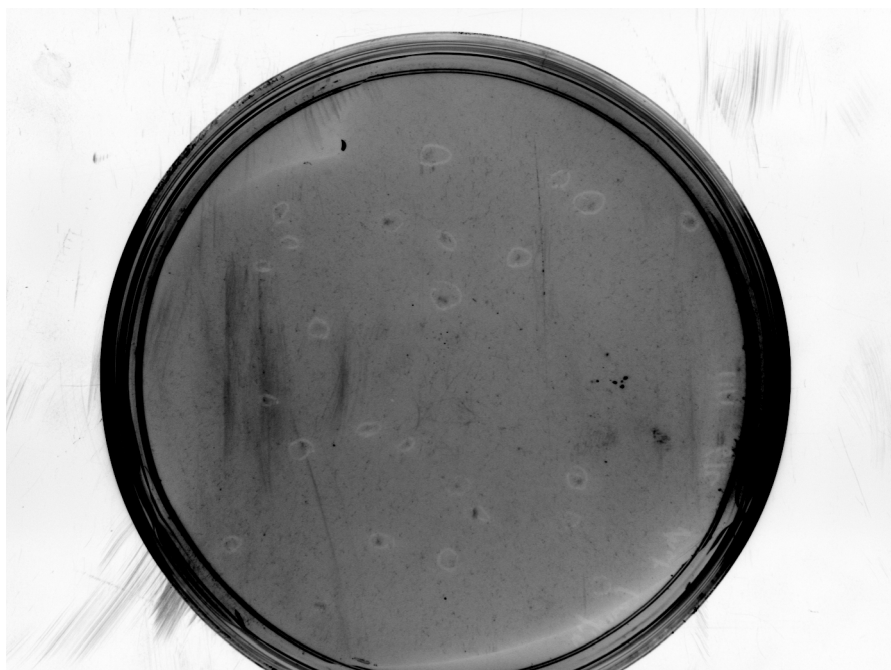


Figure 36. Hygromycin resistant colonies from NTG exposed conidia growing on 0.5% xylan with 175 $\mu\text{g}/\text{mL}$ hygromycin.

Endoglucanase Activity on Xylan

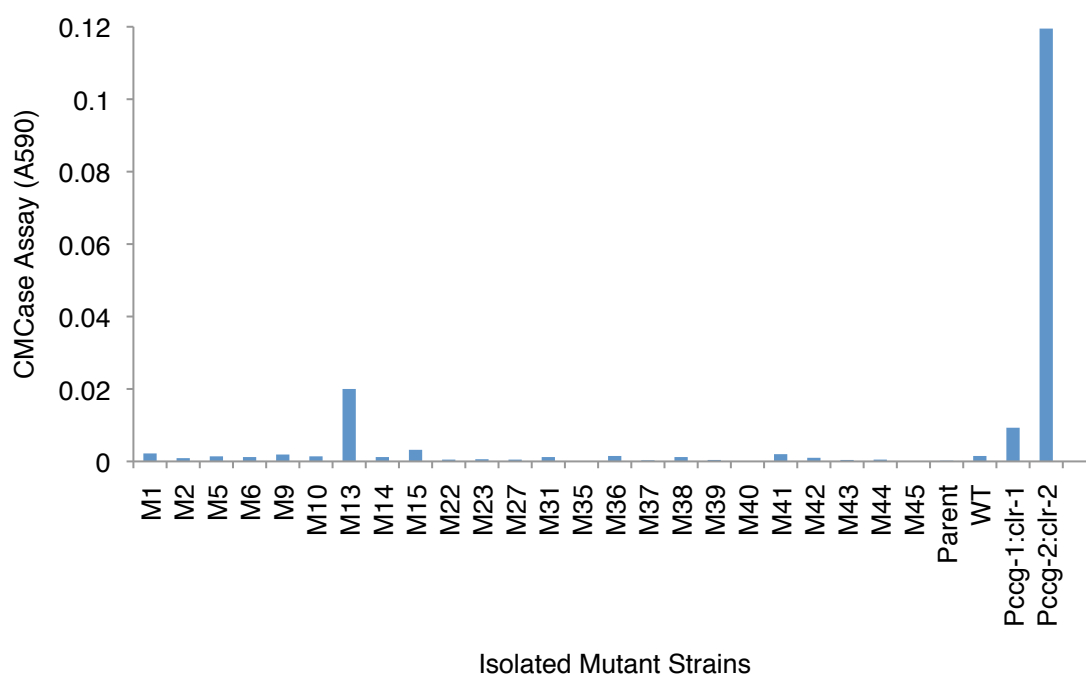


Figure 37. Cellulase activity of xylan cultures from mutant strains. Endoglucanase activity in xylan grown culture supernatants of NTG mutant hygromycin resistant strains from parent SC3G4.

Mapping mutations with bulk segregant analysis

I employed bulk segregant analysis to identify the mutation in strain M13 conferring constitutive cellulase activity. First M13 was crossed back to the complement *mat A* parent, SC3G5. Progeny were then tested for hygromycin resistance or sensitivity and binned *hyg*+/- strain were tested for endoglucanase activity on xylan media. Representative data from 23 progeny (1 24-well plate with a blank well control) are shown in Figure 38. Hygromycin resistance segregated with constitutive cellulase activity, though activity levels were highly variable. The weakness of the activity phenotype (compare M13 to *Pccg-1:clr-2*, Figure 37) made mistaken binning of false-negative progeny a concern, as strains with weak cellulase expression might also have weak *hph* expression and thus appear as hygromycin sensitive as well. Nevertheless, total segregation approached one to one with 58 positive progeny and 53 negative progeny isolated. Of these, genomic DNA was extracted from 50 positive and 50 negative progeny.

Endoglucanase Activity

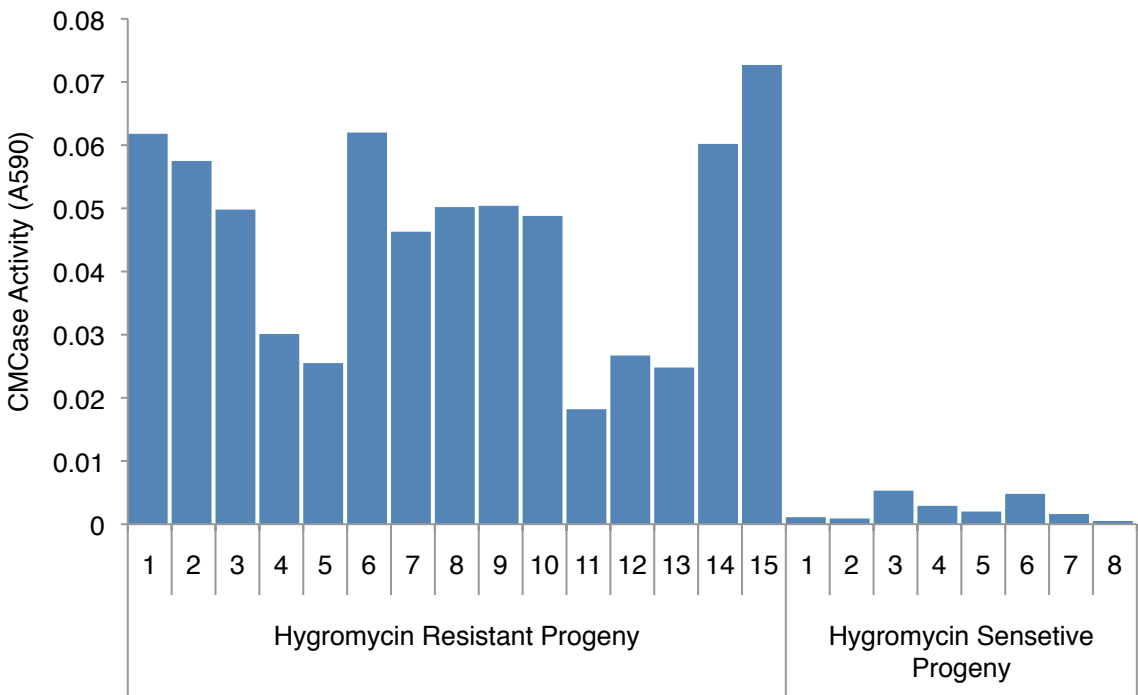


Figure 38. Segregation of hygromycin resistance and constitutive cellulase activity in M13 cross progeny.

Genomic DNA from pooled positive progeny, pooled negative progeny and the initial parent strains SC3G4 and SC3G5 was submitted to the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley for resequencing. Reads were mapped to the reference *N. crassa* sequence and mutations were analyzed with the Genome Analysis Toolkit (see methods). A total of six high confidence point mutations not found in the parental strains were identified in the mutant pools (Table 4). Of those mutations, only one was highly enriched in the positive

(constitutive cellulase producing) pool. The mutation leads to a single amino acid substitution in a conserved protein and is the subject of Chapter 7.

I then replicated the pilot protocol with a larger number of conidia. Given the low mutation rate found in M13, I employed a range of mutagenesis conditions in the second round to sample different mutation rates for effectiveness at isolating cellulase induction mutants. Conidia were exposed to mutagenesis in the following combinations: 1×10^8 conidia, 50 μ M NTG (same as pilot), 1×10^8 conidia, 100 μ M NTG, 1×10^9 conidia 50 μ M NTG, and 1×10^9 conidia 500 μ M NTG. Six more constitutive cellulase mutants were isolated with this procedure. As of this writing segregant pools of these additional mutants await sequencing at the Vincent J. Coates Genomics Sequencing Laboratory.

Table 4. Bulk segregation of mutations identified in strain M13

Supercontig	Position	Mutation		Percentage Reads with Mutation			
		Ref	New	M13 Progeny		Parent Strains	
				Pos	Neg	SC3G4	SC3G5
12.7	4037101	T	C	99	9	0	0
12.12	12731	C	A	52	28	0	0
12.3	2549781	A	G	51	28	0	0
12.5	1486592	C	T	47	7	0	0
12.4	4505270	T	C	43	67	0	0
12.6	752449	T	C	30	100	0	0

Discussion

The unique morphological features that make filamentous fungi successful colonizers of plant detritus and thus have helped enabled them to evolve complex and effective strategies for plant cell wall deconstruction can also make them difficult to engineer. Selective strategies for mutant isolation are thwarted by macroscopic multinucleate fungal colonies. The fact that cellulases are secreted outside the colony also removes the benefits of hypersecretion from an individual in a mixed culture. By tying antibiotic resistance to cellulase induction and working with conidia, I have circumvented some of these issues and built a system suitable for selection of new mutants from sample populations orders of magnitude greater than can be sampled by laborious screens based on quantification of colony growth or cellulase clearance assays.

A limitation of the selective system in its current form is the propensity to isolate mutants with increased hygromycin resistance unrelated cellulase secretion. With results collected thus far, it would appear with a single antibiotic resistance reporter, general mutations conferring resistance outnumber mutations in cellulase regulation by at least one hundred to one. These false positives slow mutant identification but the yield of true positives is still tractable for identification with rapid enzyme activity assays on cultures in a medium throughput, 24-well plate format. The false positive rate would likely be greatly reduced by the addition of a second selectable marker at another suitable locus, such as NCU00206.

This selective system is also limited to queries of nutrient sensing and cellulase transcriptional induction (as opposed to genes involved in protein

secretion). If the selection of this system could be tied to protein secretion, perhaps with a secreted marker that remained attached to the cell wall, then it would have even more potential for discovery of new factors enabling efficient cellulase production. Even with its current limitations this selectable system may unlock the as yet unknown molecular mechanism underlying cellobiose sensing.

Acknowledgements

Though based on published methods, the mutagenesis protocol I employed in this chapter was also based on one first used by lab member Jason Liu for another project, who first researched the method, gathered the necessary reagents and prepared a refined protocol. His efforts helped me rapidly and easily deploy the same protocol later. Reagents and methods from lab members Timo Shurg and Willfried Jonkers were helpful in establishing the effective use of phleomycin selection. Jens Heller provided some reagents and helpful tips to improve the quality of my genomic extractions for best sequencing results. Finally, the computational pipeline I employed here was first implemented for *N. crassa* by combination of publically available tools by lab member David Kowbel. His notes and well-tested software installation provided me a clear path to quickly analyze my own data.

Published material

The data in Figure 33 were taken from previously published data by myself [169], Elizabeth Znameroski [128] and J. Phillip Benz [201]. All other data in this chapter are original and unpublished.

Methods

Strains

The starting strains for this study were FGSC 9719 (Δ mus-52::bar⁺ mat a) and FGSC 2489 OR74A mat A. Selectable strains SC3G4 (Pncu01050:*hph* mat a) and SC3G5 (Pncu01050:*hph* mat A) were created by transforming FGSC 9719 with pSC107 (Appendix C: Sequences) using established protocols [161] with minor adjustments. After electroporation, conidia were resuspended in 1 mL chilled sorbitol, then diluted in 10mL Vogel's medium [119] with 2% sucrose and incubated 4 hours at 30C. Conidia were then centrifuged 2 minutes at 1000 rcf and the supernatant removed. The pellet was resuspended in 20 mL 1% top agar with vogels salts and fructose-glucose-sorbose carbon source (10x FGS autoclaved separately and added at 1/10 final volume), adjusted to pH 8 with 0.1 M Tris buffer (1M stock at pH 8.1 autoclaved separately, then added at 1/10 final volume). Top agar was then spread over 4 plates with 10 mL each of the same media with 1.5% agar and 10 μ g/mL phleomycin. Phleomycin resistant strains were then crossed back to FGSC 2489 and glufosinate sensitive, phleomycin resistant progeny of each mating type were isolated.

Mutagenesis

Established protocols were used for NTG mutagenesis [204] with slight alterations. 1×10^8 or 1×10^9 conidia were harvested and suspended in 25mL PBS with 50 μ M, 100 μ M or 500 μ M N-methyl-N'-nitro-N-nitrosoguanidine and incubated at room temperature for 2 hours with gentle rotation. Conidia were then centrifuged 10 minutes at 2000 rcf and washed with 50 mL Vogel's salts with 0.1% thiosulfate three times.

Conidia were then resuspended in 20mL molten top agar with 1% agar, Vogel's salts without added calcium and with 0.11 M Tris, pH 8 (added as a 1/9 dilution of 1M Tris pH 8.1 autoclaved separately) and 0.5% beechwood xylan (Sigma Aldrich). The low calcium concentrations may discourage germling fusion [205] and the 0.11 M Tris pH 8 induces psuedo colonial growth in *N. crassa* through an unknown mechanism. Both these alterations in media were undertaken because the typical method of stimulating colonial growth, addition of 0.05% glucose, 0.05% fructose and 2% sorbose would inhibit cellulase induction through carbon catabolite repression and without any alterations to morphology and fusion, resistant colonies would spread rapidly and fuse. This would likely frustrate analysis and possibly even resulting in complementation of resistance phenotypes and colony death. Top agar was plated on 4 plates 10mL with the same composition as above except with 1.5% agar and 262 μ g/mL hygromycin B (final concentration 175 μ g/mL). Resistant colonies were then picked to slants with Vogel's 0.5% xylan and 200 μ g/mL hygromycin B (Invitrogen)

Enzyme activity assays

Individual mutant progeny were inoculated at 5×10^5 conidia per ml in Vogel's medium with 2% xylan in 3mL cultures in 24-well plates (GE Healthcare UK Limited) and grown 48-72 hours at 25C constant light with constant shaking at 200 rpm. Cellulase activity was detected by endoglucanase activity of culture supernatants, assayed with the Remazolbrilliant Blue R conjugated CMC kit from Megazyme (Wicklow, Ireland).

Bulk segregant analysis

Individual mutant progeny were inoculated at 5×10^5 conidia per ml in Vogel's media with 2% sucrose in 3mL cultures and grown as above 48 hours. Mycelia were harvested by filtering on Whatman paper #1, frozen at -80C overnight and lyophilized. Lyophilized pads were ground by beadbeating dry with silica beads. DNA was extracted by incubating ground mycelia in 400 μ L Lysis buffer (0.05M NaOH, 1mM EDTA, 1% TritonX) for 30min at 65C. 80 μ L of 1M Tris HCl, pH 7.5 was added to neutralize the buffer, then 400 μ L Phenol/Chloroform/Isoamylalcohol 25:24:1. Extracted DNA (400 μ L) was treated with 1 μ L RNase A (Thermoscientific) for 30 min at 37C. Quality of genomic DNA was then checked on 1% agarose gels. If a single clear high molecular weight band was present, then abundance was estimated from the gel and sample pools were combined at approximately equal DNA per progeny. Pooled or single (parent) samples were then further purified with Qiagen DNeasy Blood and Tissue kits and submitted to the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley for library preparation an sequencing.

The libraries were sequenced on the Illumina HiSeq2000 platform with 100bp paired reads of 500bp fragments. Reads with quality scores below 20 were then filtered out with the FastX Toolkit (Fastq Quality Filter). Filtered reads were realigned with the cmpfastq.pl script from NIHR, London. Concordant reads were then mapped to the *N. crassa* reference genome (Version 12, from the Broad Institute) with Bowtie version 2.0.0-beta6 [133]. Read groups were added with Picard (<http://picard.sourceforge.net>) and mutations were called as SNPs with the Genome Analysis Toolkit [206].

Chapter 7

NCU05846: a Repressor of CLR-1 Activation

Abstract

A mutant strain of *N. crassa* with constitutive cellulase activity (M13) was recently identified and shown to bear a non-synonymous mutation in NCU05846, a conserved protein with a domain of unknown function. Here I show that M13 is a loss of function mutation for NCU05846 and characterize its effects on cellulase regulation. The Δ NCU05846 strain had constitutive cellulase activity in culture supernatants, but activity was lower in neutral xylan media than inducing Avicel media, suggesting that NCU05846 acts as a repressor of CLR-1 activation and other activating factors remain to be discovered. The constitutive activity phenotype of the Δ NCU05846 mutant was CLR-1 dependent and transcriptional profiles of Δ NCU05846 mutants on no carbon media supported an effect through direct activation of CLR-1 as opposed to induction of CLR-2.

Introduction

In Chapter 6, I introduced a selective screen for mutants with constitutive cellulase induction with a goal of identifying new genes involved in cellobiose sensing and elucidating the molecular mechanism of CLR-1 activation in the presence of cellobiose. As an initial pilot run of that system, I identified a point mutation in a conserved protein (NCU05846) that led to weak constitutive cellulase secretion. In this chapter I explore the nature of that mutation and present evidence that NCU05856 acts as a repressor of CLR-1 activation in non-inducing conditions

The M13 mutation is a loss of function mutation in a conserved domain of unknown function.

A single mutation in strain M13 that segregated with constitutive cellulase secretion mapped to a conserved region of a hypothetical protein NCU05846. NCU05846 includes a conserved domain of unknown function, PFAM DUF1479. As shown in Figure 39, this mutation results in an E to G mutation in a highly conserved central region of DUF1479. This residue is conserved in all the ascomycete homologs of this protein analyzed and is likely required for proper function.

Alignment of NCU05846 Homologs

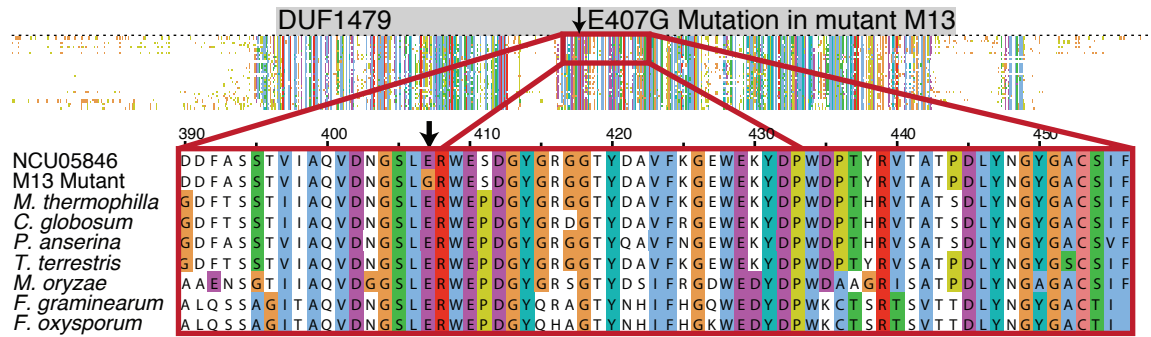


Figure 39. The M13 mutation is a non-synonymous mutation in a conserved domain of unknown function.

(Top) an alignment of the amino acid sequence of NCU05846 with homologs from other filamentous ascomycete species. Coloration of amino acid positions by conservation highlights conserved positions coincident with a predicted domain of unknown function.

(Bottom) The local region of the M13 mutation aligned with orthologous genes and the WT NCU05846 sequence.

DUF1479 is widely distributed among fungi and bacteria, but reports on functions of this domain are limited. In one study on pH tolerance in *Escherichia coli*, increased expression of a DUF1479 containing protein (YbiU) conferred higher tolerance to low pH [207]. In a metabolic modeling study, also in *E. coli*, YbiU was proposed as part of an uncharacterized pathway for thymidylate synthesis [208]. A proteomic affinity study on the *E. coli* protein deacetylase CobB identified YbiU as possible interacting protein [209]. So while the domain is of interest, no strong data on its function in bacteria have been reported.

In fungi, DUF1479 genes have been associated with pathogenesis of animals. One study found several DUF1479 genes were more highly expressed in *Cryptococcus grubii* than the less virulent *Cryptococcus neoformans*. [210]. In *Candida albicans*, mutants for the DUF1479 domain protein GlG2 had reduced virulence in mouse models and increase sensitivity to hydrogen peroxide [211]. That study noted that GlG2 was transcriptionally induced by N-acetylglucosamine (GlcNAc) and reported evidence that GlG2 bound GlcNAc or a chemically similar compound and suggested a role for GlG2 in GlcNAc metabolism.

A phylogenetic analysis of NCU05846 and likely homologs suggests that it belongs to a small family of conserved proteins limited to filamentous ascomycete species. The family consists of three major clades (Figure 40). One clade is present in both Sordariomycetes and Eurotiomycetes with orthologs present in several *Aspergillus* species; the major plant pathogens *F. graminearum* and *M. oryzae*; the industrially important *T. reesei*; and several species closely related *N. crassa* such as *P. anserina*. *N. crassa* has lost this relatively conserved gene. Two other clades are limited to the Sordariomycetes and lack members in *T. reesei*, but *N. crassa* has a gene in each. One clade includes NCU05846 and likely orthologs limited to the Sordariomycetes. A third clade includes a paralogous *N. crassa* gene, NCU02600 and genes in *Fusarium* species, *M. thermophila* and *P. anserina* with unclear orthology as these genes did not cluster in a monophyletic group and several branches in this

part of the tree had poor bootstrap support. None of these homologs have any reported functions. NCU05846 and NCU02600 are the most similar *N. crassa* proteins to *C. albicans* GlG2, but presence of more similar proteins in *Fusarium* species suggests the relationship of NCU05846 and GlG2 is one of paralogy rather than orthology.

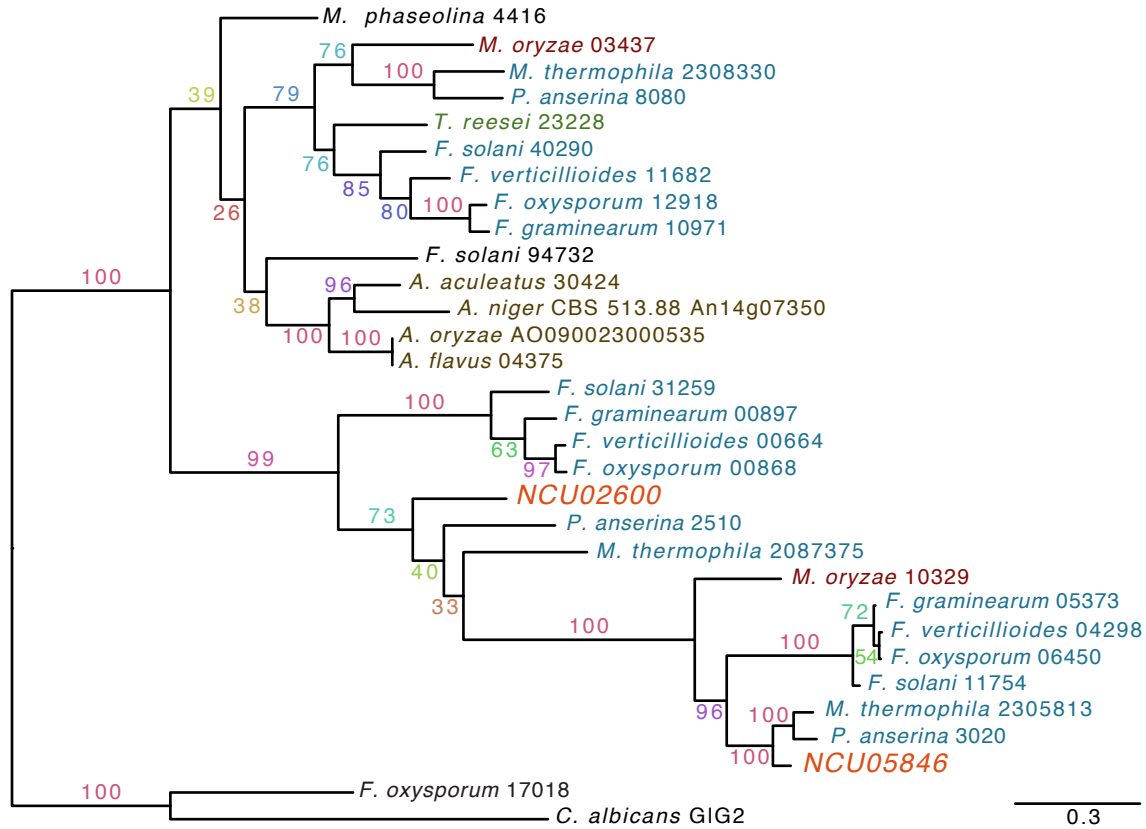


Figure 40. Phylogenetic Tree of NCU05846 and Homologs.

Three major clades representing at least three distinct paralogs are present. Sordariomycetes with all three paralogs in blue. Branch labels are bootstrap scores for 100 simulated trees.

To confirm the causality of the identified M13 mutation in NCU05846 and investigate any role for the paralog NCU02600 in cellobiose signaling, I cultured deletion strains for both genes from the deletion collection and measured endoglucanase activity after a shift to several carbon sources. I cultured both mating types of the NCU05846 available in the deletion collection to serve as biological duplicates and to test for any mating-type specific defects or any inconsistent phenotype between different cross progeny bearing the same deletion. Only the *mat a* mating type is present in the deletion collection for NCU02600, so I grew that strain in duplicate. Both NCU05846 deletion mutants and the NCU02600 deletion mutant had WT endoglucanase activity on 2% Avicel and 2% sucrose, as did two independently isolated cross progeny bearing the M13 mutation, demonstrating that positive activation of CLR-1 and carbon catabolite regulation of cellulases remained intact in all these strains. Only the Δ NCU05846 mutant and the M13 progeny had elevated endoglucanase activity on 2% xylan and slightly more

repressive 1% sodium acetate. The NCU05846 deletion mutant and strain M13 each had similar levels of activity in all conditions. The Δ NCU02600 strain was fully WT in all tested conditions. From this data I concluded that the M13 mutation was a simple loss of function mutation and thus that NCU05846 plays a role in preventing illegitimate induction of cellulases in non-inducing conditions.

Endoglucanase Activity

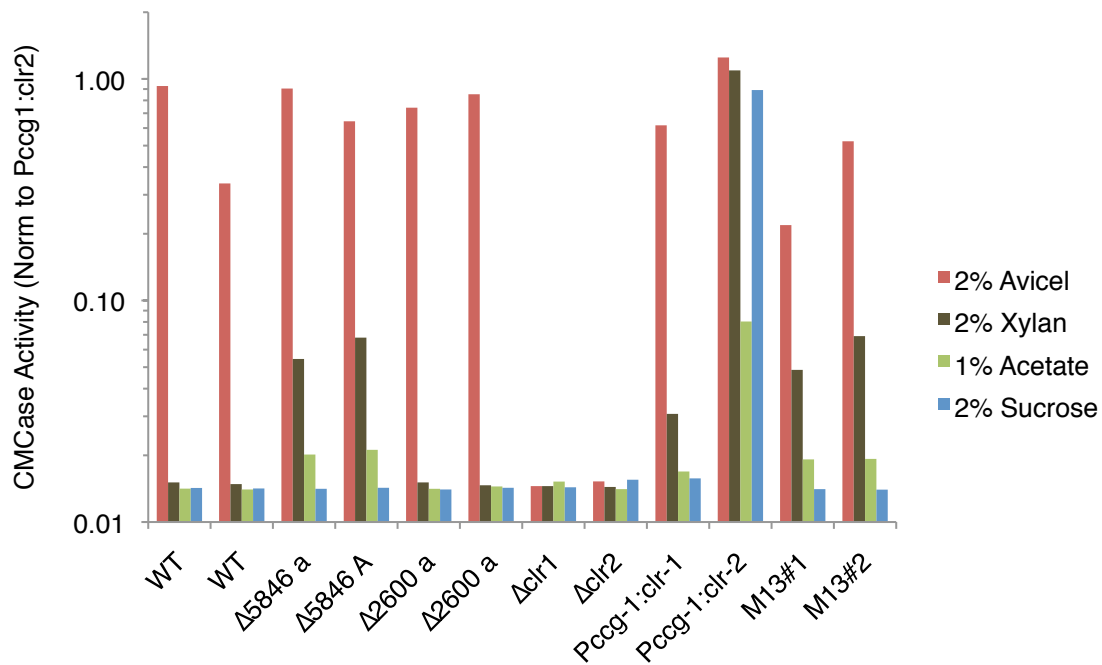


Figure 41. Endoglucanase activity in supernatants from NCU05846 deletion mutants 48 hours after a shift to inducing and noninducing conditions.

Endoglucanase activity by AzoCMCase activity, normalized to the *Pccg-1:clr-2* strain on Avicel.

NCU02600, a Paralog of NCU05846 Has No Clear Role in Cellulase Induction.

To test if NCU02600 had any overlapping or redundant functions with NCU05846, or had a weak effect that might be more detectable in a Δ NCU05846 background, I crossed the two deletion mutants and isolated a Δ NCU02600 Δ NCU05846 double mutant. This mutant had indistinguishable cellulase activity from the Δ NCU05846 mutant after a transfer to xylan media (Figure 42). Thus far, no phenotype of any kind has been recorded for an NCU02600 mutant on any media.

Endoglucanase Activity On Xylan

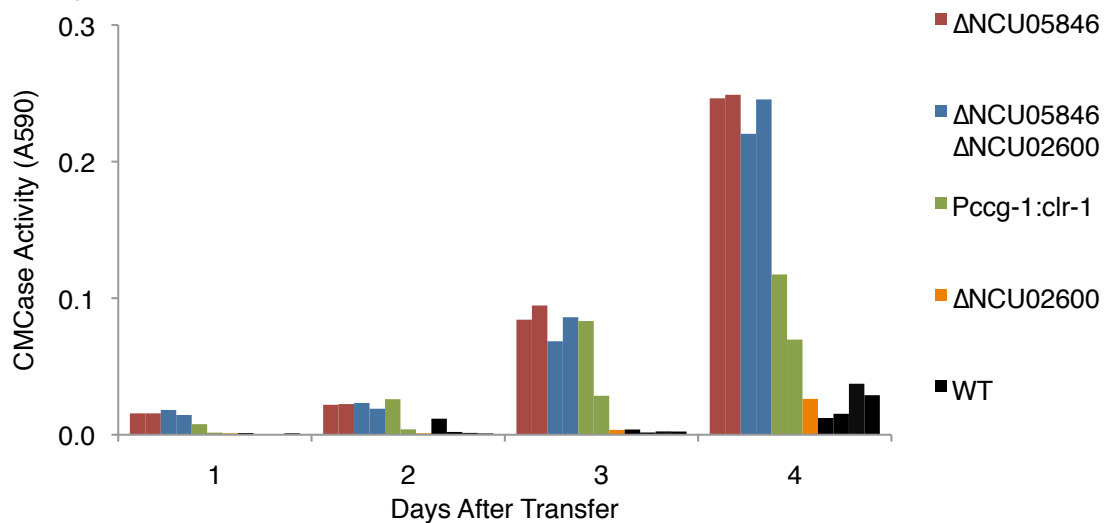


Figure 42. Endoglucanase activity on xylan for NCU02600 deletion mutants. Time course of endoglucanase activity in cultures germinated on xylan.

The Δ NCU05846 constitutive activation phenotype requires CLR-1

To test if the Δ NCU05846 phenotype was manifested through an activation of CLR-1 or if in this genotype another, independent activator could be implicated, I crossed the Δ NCU05846 mutant to the *Pccg-1:clr-1* strain (constitutive expression construct at the *his-3* locus, deletion cassette at the *clr-1* locus) and isolated both *Pccg-1:clr-1* mutants and Δ *clr-1* mutants in the Δ NCU05846 background. The constitutive cellulase activity was abolished in four independently isolated Δ *clr-1* Δ NCU05846 progeny and was slightly heightened in four independently isolated *Pccg-1:clr-1* Δ NCU05846 progeny switched to xylan media (Figure 43). I interpreted these data as confirmation that the Δ NCU05846 phenotype requires CLR-1 and likely is mediated by illegitimate, though incomplete, CLR-1 activation.

Endoglucanase Activity On Xylan

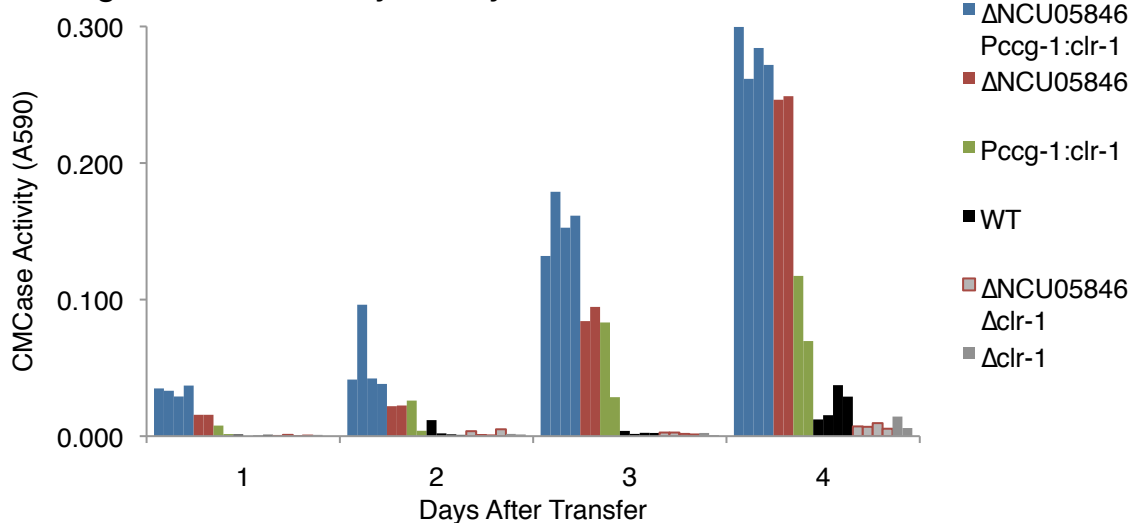
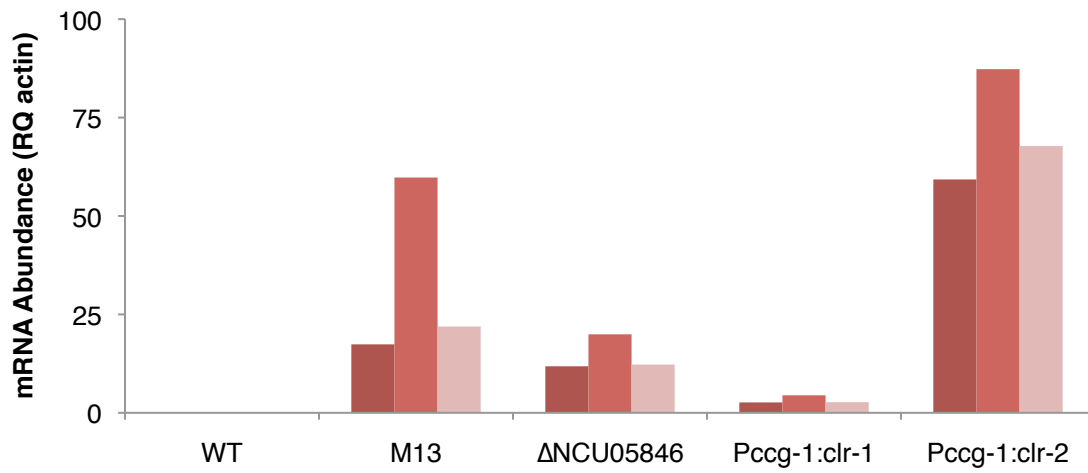


Figure 43. Constitutive cellulase activity in the Δ NCU05846 mutant requires *clr-1*
Time course of endoglucanase activity in cultures germinated on xylan.

A direct CLR-1 target is strongly induced in Δ NCU05846

Previous results demonstrated that while *cbh-1* (NCU073400) induction by cellobiose requires both CLR-1 and CLR-2, the induction of *cdt-2* (NCU08114) by cellobiose only requires CLR-1, and is likely a direct target of CLR-1 (Figure 18). I examined induction of these two genes in the M13 and Δ NCU05846 mutants and compared that expression to WT and *clr-1* and *clr-2* constitutive expression strains (Figure 44). While *cbh-1* transcript was clearly increased over the WT and *Pccg-1:clr-1* strains, it was still significantly lower than in the *Pccg-1:clr-2* strain, reflecting the relative cellulase activity in these strains. However, *cdt-2* abundance (Figure 44B) is highest in the M13 and Δ NCU05846 strains, likely reflecting a relatively strong induction of direct CLR-1 targets.

A. *cbh-1* Transcript



B. *cdt-2* (NCU08114) Abundance

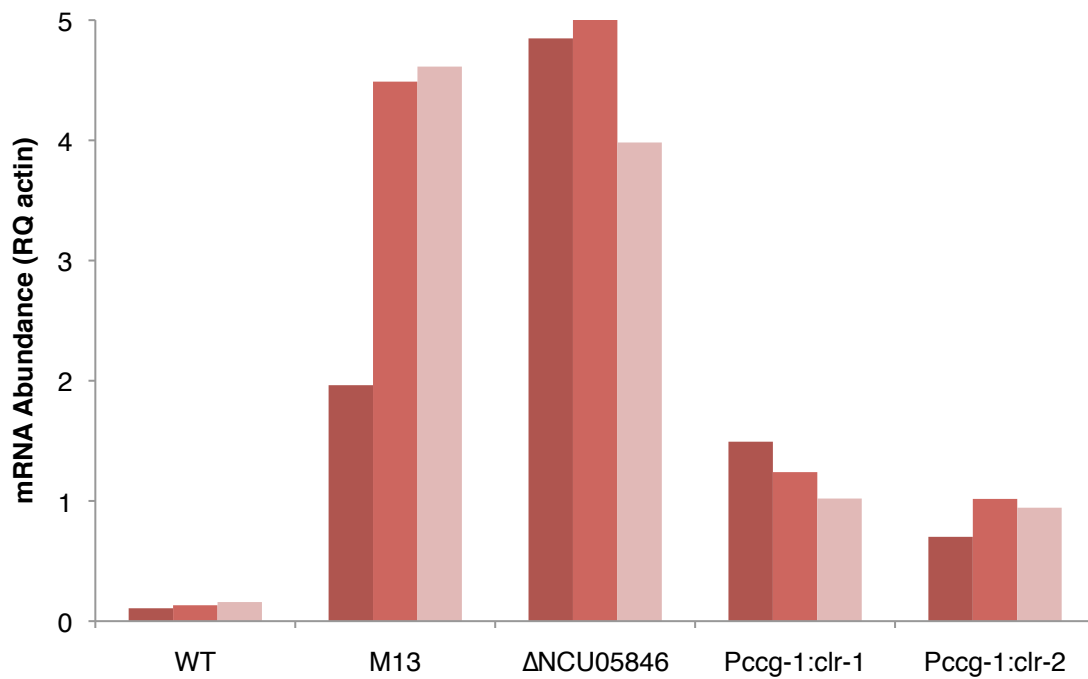


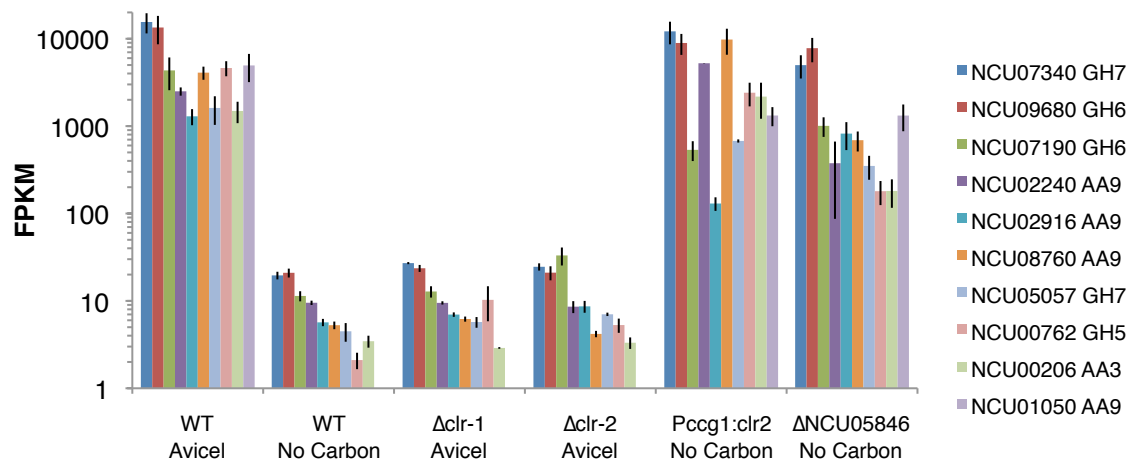
Figure 44. Induction of CLR-2 independent genes in the Δ NCU05846 mutant.

RT-PCR of (A) *cbh-1* NCU07340 and (B) *cdt-2* NCU08114 after a shift to no carbon media. The three bars for each genotype represent individual biological replicates.

CLR-1 regulated genes are induced in Δ NCU05846

To broadly assess the induction of the Avicel regulon in the Δ NCU05846 mutant, I subjected Δ NCU05846 mutants shifted to no carbon media to transcriptional profiling with RNAseq. The RNAseq libraries largely reflected the earlier RT-PCR results, with all major cellulases clearly induced, but generally lower than in *Pccg-1:clr-2* mutants in the same condition (Figure 45A). Conversely, cellobiose utilization transcripts with a stronger dependence on CLR-1 were generally more abundant in the Δ NCU05846 mutant than the *Pccg-1:clr-2* mutant (Figure 45B).

A. Cellulases



B. Cellobiose Utilization

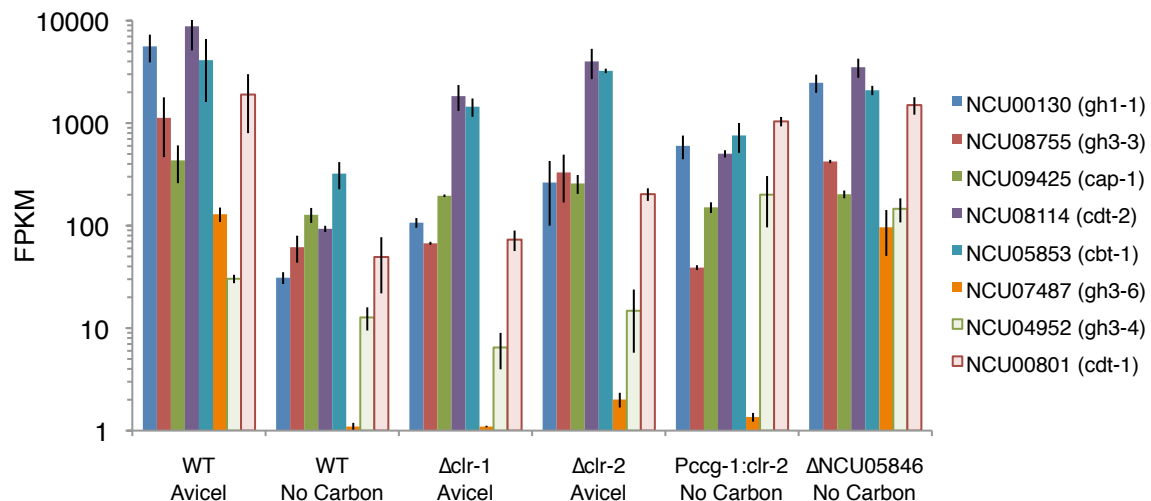
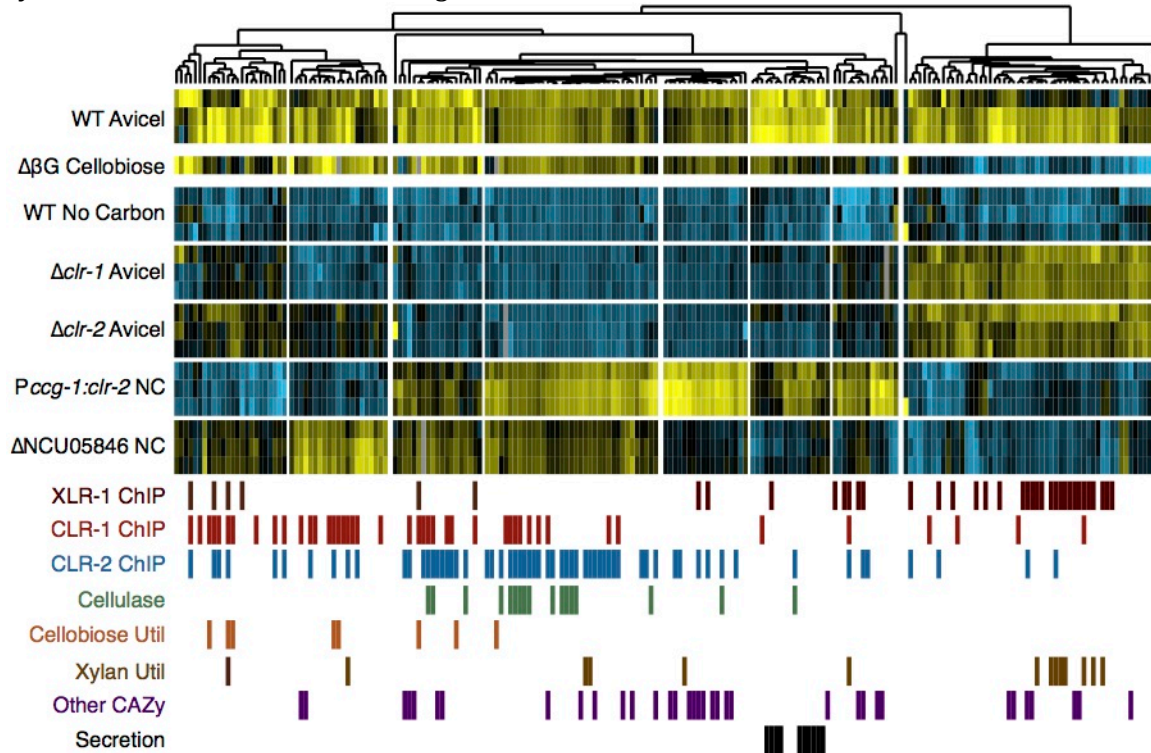


Figure 45. Expression of cellulase and cellobiose utilization genes in the Δ NCU05846 mutant. (A) FPKM of major cellulases. (B) FPKM of genes implicated in cellobiose/cellobionic acid transport or utilization

To get a general overview of gene induction in the Δ NCU05846 mutant I sorted genes in the Avicel regulon by hierarchical clustering of FPKM in inducing conditions (Avicel and cellobiose [128]), non inducing conditions (no carbon), induction deficient mutants (Δ *clr-1* and Δ *clr-2* on Avicel) and constitutive mutants (Δ NCU05836 and *Pccg-1:clr-2* on no carbon) (Figure 46). Broadly, genes that were dependent on *clr-1* but not *clr-2* for induction were induced in the Δ NCU05846 mutant. Approximately half of the genes dependent upon both *clr-1* and *clr-2* were induced in Δ NCU05846, whereas effectively all of these genes were induced *Pccg-1:clr-2*. Genes that were independent of *clr-1* and *clr-2* for induction on Avicel were not induced on cellobiose and not induced in either Δ NCU05846 or *Pccg-1:clr-2*. Direct binding of CLR-1, CLR-2 and XLR-1 by ChIPseq [102] was largely concordant with the RNAseq data, with XLR-1 binding mostly promoters of CLR-1/CLR-2 independent genes induced by hemicellulosic contamination of Avicel; CLR-2 binding promoters of genes induced in the *Pccg-1:clr-2* mutant, including all major cellulases; and CLR-1 binding mostly genes induced in the Δ NCU05846 mutant, including most major cellobiose utilization genes. More of the CAZy genes with likely activity on polymers and side chains aside from cellulose or xylan were bound by CLR-2 and induced in the *Pccg-1:clr-2* mutant and not the Δ NCU05846 mutant.



LOW RESOLUTION FIGURE. REPLACE

Figure 46. Transcription of Avicel induced genes in the Δ NCU05846 mutant on no carbon media

Hierarchical clustering of the genes in the Avicel regulon with added data from the Δ NCU05846 mutant on no carbon media. Heat map shows FPKM blue = low expression, yellow = high expression. Bars below show genes with evidence of direct binding by XLR-1, CLR-1 and CLR-2 from chromatin immunoprecipitation sequencing (ChIPseq) [102]. Additional bars indicate genes with predicted or characterized function in cellulose depolymerization; cellobiose utilization (genes from Figure 45);

xylan depolymerazation and xylose utilization; other carbohydrate active enzymes with likely involvement in plant cell wall utilization; and genes with predicted function in protein folding and secretion.

A gene set of note consisted of nine tightly clustered genes with known or suggested roles in protein secretion (Table 5) and a tenth co-regulated hypothetical protein, NCU03083. That gene is orthologous to an *A. niger* gene responsive to DTT and tunamycin [212], suggesting a possible role in the unfolded protein response. A priming of the unfolded protein response has been noted as part of the *N. crassa* response to cellulose [201]. In turn, repression of cellulases and xylanases has been noted in *T. reesei* under secretion stress [213].

The ten genes with implicated roles in protein secretion were induced in the WT on Avicel and in the *Pccg-1:clr-2* mutant on no carbon but not in the Δ NCU05846 mutant on no carbon. These data might suggest these genes are under the regulation of CLR-2, but they were not bound by CLR-2 in ChIPseq libraries constructed and analyzed by Craig et al [102]. Together these data suggest an indirect role for CLR-2 in regulation of these genes as an adaptation of the general protein secretory apparatus to the large throughput required for efficient cellulase secretion.

Table 5: CLR-2 Responsive genes involved in protein secretion

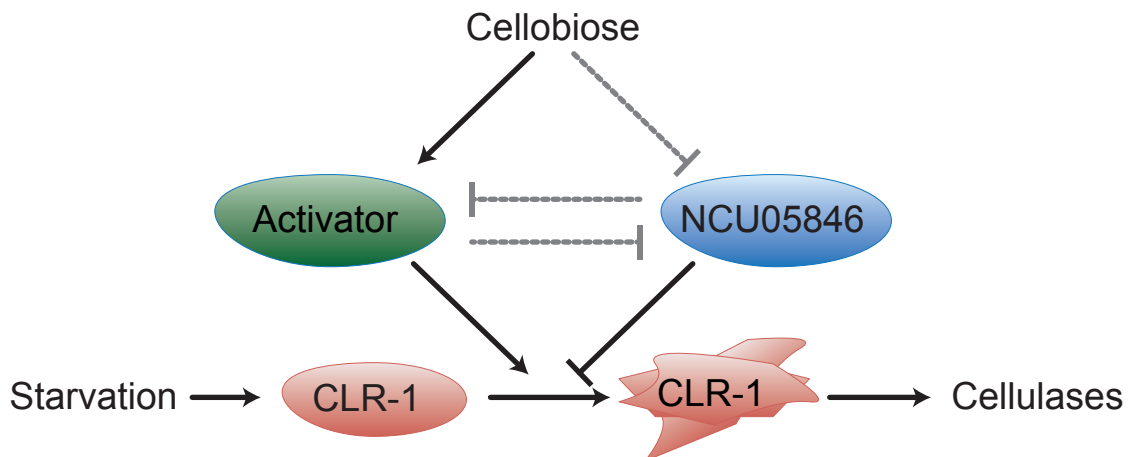
NCU	Annotation
NCU09485	chaperone dnaK
NCU00169	translocation complex component
NCU00813	disulfide isomerase
NCU06333	translocation protein SEC62
NCU09223	protein disulfide-isomerase
NCU03982	glucose regulated protein 78
NCU08897	protein transporter SEC61 subunit alpha
NCU03083	hypothetical protein
NCU09265	calnexin
NCU10721	solute carrier family 35 member B1 protein

Discussion

As of this writing, molecular mechanisms of oligosaccharide sensing by fungi remain unelucidated. The discovery the a DUF1479 containing protein plays an important role in cellulase regulation by repressing illegitimate cellulase expression in non-inducing conditions is the first hint at how this system may operate, but so far how this effect is achieved is unknown. Given that induction of cellulases was relatively weak in the Δ NCU05846 deletion mutant on xylan, but fully WT on Avicel, I conclude that another repressor or a positive activator of CLR-1 remains to be found. That activator/repressor may be another protein with altered binding or activity in the presence of cellobiose or it may be cellobiose itself in direct interaction with CLR-1.

Two general models for the role of NCU05846 in CLR-1 mediated induction of cellulases in response to cellobiose are depicted in Figure 47. In the first model, CLR-1 requires posttranslational modification for efficient cellulase induction. If this is the case, then NCU05846 must act as an antagonist against a cellobiose dependent activator of CLR-1. From available data, it is not clear if NCU05846's repressive effect on cellulase induction is modulated by cellobiose, but any activator would have to be responsive to cellobiose in order to drive robust induction of cellulases in the presence of Avicel in an NCU05846 deletion mutant. In the second model, CLR-1 is fully competent to bind cellulase promoters and initiate transcription without any further modification, but is subject to repression by NCU05846 and another unknown repressor protein. Such a repressor must be responsive to cellobiose, but not fully competent to block cellulase transcription in the absence of NCU05846.

A. Activation Model



B. Repression Model

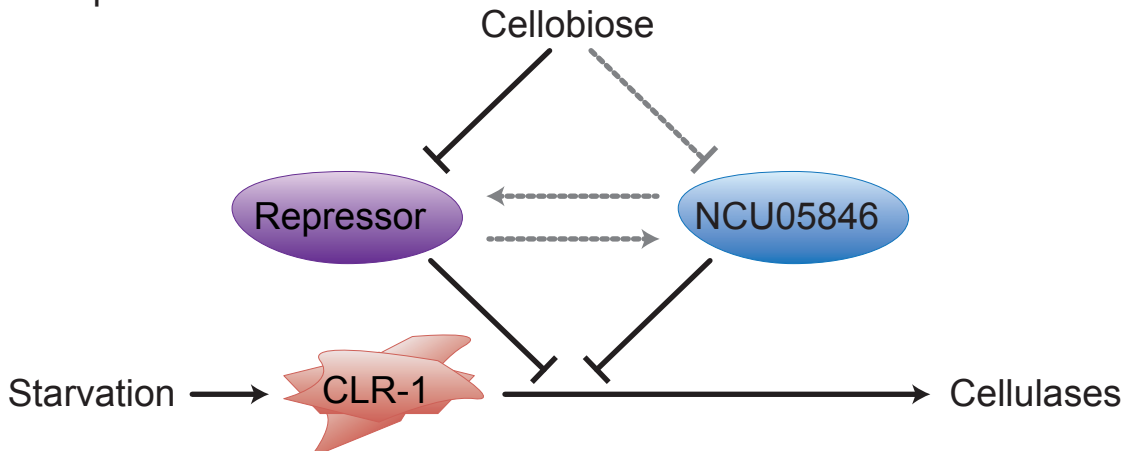


Figure 47. Alternative models for the *N. crassa* cellobiose-sensing network

Activation/repression relationships with dashed lines are possible, but not necessary to explain extant data.

There are several possible modes by which NCU05846 could repress cellulase expression. In ChIPseq experiments, James Craig detected nuclear-localized CLR-1 bound to cellulase promoters in cultures grown on sucrose [102] and I have observed nuclear localization of GFP tagged CLR-1 in non-inducing and repressing conditions (data not shown). Cellulase induction is then not likely a matter of nuclear import or DNA binding, but of modulation of CLR-1's competency to recruit general transcription factors and RNA polymerase. Such modulation could be achieved by an interaction between CLR-1, NCU05846, and a positive activator; an interaction that is altered in the presence of cellobiose, reminiscent of activation of Gal4 by altered interaction with Gal80 (repressor) when Gal3 (activator) binds galactose [91]. Alternatively, post-translational modifications to CLR-1 could be required for activation. Studies on NCU05846's localization, direct interaction with CLR-1, and possible enzymatic activities may differentiate between these models and further elucidate novel mechanisms of sugar sensing.

A larger goal of understanding mechanisms of cellulase induction was to gain sufficient insight into the larger system of proteins required to better predict functional conservation of cellulose induction systems across the filamentous ascomycetes. The conservation of NCU05846 orthologs (Figure 40) together with conservation of CLR-1, may imply a conserved sensing network. Indeed, NCU05846 is conserved in the closely related thermophilic species *M. thermophila* and heterologous expression of the CLR-1 ortholog of that species in a $\Delta clr-1$ *N. crassa* background restored growth and cellulase activity on Avicel (James Craig, personal communication) suggesting that in these two species at least, the cellobiose sensing pathway is well conserved. The identification of more proteins involved in this sensing network should improve the confidence and phylogenetic breadth with which we can predict such functional conservation.

Acknowledgements

The comparative RNAseq analysis in this chapter benefited from cellobiose induction data collected by Elizabeth Znameroski [128]. ChIPseq data included in Figure 46 were collected, sequenced and analyzed by lab member James Craig for a forthcoming publication [102] to which I also contributed and are reproduced here with his consent.

Published material

None of the new data presented in this chapter have been published previously.

Methods

Strains

The M13 strain was created as described in Chapter 6. The *Pccg-1:clr-2* strain was created as described in Chapter 4. The *Pccg-1:clr-1* strain was created crossing strain FGSC 11029 (*clr-1::HygR* mat a) with strain RLM 13-18 (*sad-1::HygR rid-1*, *his-3* mat A) to obtain strain SC2A6 (*clr-1::HygR sad-1::HygR rid-1*, *his-3* mat A) and transforming with linearized plasmid pSC65 (Appendix C), derived from

pMF272 with *clr-1* under control of the *ccg-1* promoter with a C-terminal 6xHis tag and V5 tag. The $\Delta clr-1$ mat a, $\Delta clr-2$ mat a, $\Delta NCU05846$ mat a, $\Delta NCU05856$ mat A, and $\Delta NCU02600$ mat a strains were taken from the *N. crassa* deletion collection from the Fungal Genetics Stock Center (FGSC) [157]; FGSC numbers 11029, 15835, 14349, 14350, and 19887, respectively.

Culture conditions

N. crassa cultures were grown on Vogel's minimal medium (VMM) [119]; carbon sources were 2% w/v unless otherwise noted. Strains were inoculated into 3 mL VMM slants and grown at 30°C in the dark for 48 hrs, then at 25°C in constant light for 4-10 days to stimulate conidia production. Liquid culture were grown at 25°C in constant light and shaking (200 rpm).

For enzyme activity experiments, conidia were inoculated into 3mL cultures in 24-well plates (GE Healthcare UK Limited) at 5×10^5 conidia per mL, grown 24 hours then washed in Vogel's media with no carbon and shifted to media with 2% sucrose, 2% xylan, 2% Avicel or 1% sodium acetate. Culture supernatants were then sampled from 4-96 hrs.

For transcriptional analysis, conidia were inoculated into 100 mL of liquid media at 10^6 conidia/mL and grown for 16 hours, then washed in Vogel's media with no carbon and transferred to fresh Vogel's media with no carbon or with 2% Avicel, 2% xylan or 0.2% cellobiose for $\Delta \beta$ -glucosidase cultures [128]. At 4 hours, cultures mycelia were harvested by filtering on Whatman paper #1 and flash frozen in liquid nitrogen.

Enzyme activity assays

Relative enzymatic activity of culture supernatants was assayed with Remazolbrilliant Blue R conjugated CMC as in Chapter 2

RNA extraction and transcript abundance

All RNA extraction, RT-PCR, sequencing, abundance calculation and hierarchical clustering as in Chapter 5.

Chapter 8

Discussion and Conclusions

Conserved strategies for plant cell wall deconstruction by filamentous fungi

My comparative analysis of CLR-1/CLR-2 function in *N. crassa* and *A. nidulans* was also an excellent opportunity to compare strategies for plant cell wall deconstruction between divergent species with different lifestyles.

Until recently, studies on regulation of plant cell wall deconstruction in *Aspergillus* species focused predominantly on induction by hemicellulosic sugars via the transcriptional regulator, XlnR [80,81,214]. *A. nidulans* is more proficient at saccharification of xylan and relatively less effective at degrading crystalline cellulose than *N. crassa*. This phenotypic distinction was consistent with both relative expression and diversity of CAZy families between species, with *A. nidulans* specifically inducing more xylanases, β -mannases and pectate lyases in particular. Together these data suggest that *A. nidulans* and perhaps other *Aspergillus* species have evolved to primarily utilize amorphous polymers over crystalline cellulose. In these species, canonical cellulases may therefore serve a more accessory role for hemicellulolytic activity by synergistic action on complex substrates.

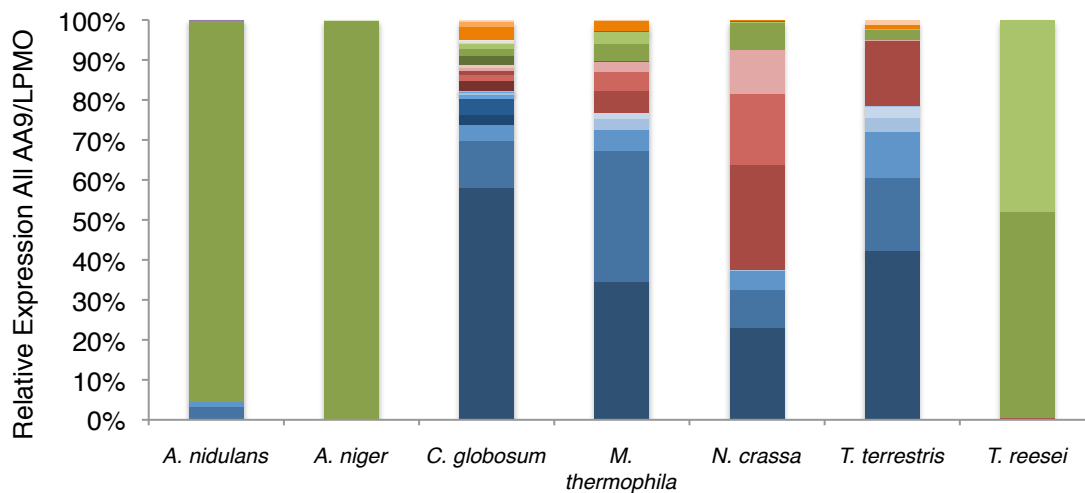
Regardless of any differing specialization in terms particular polysaccharides, one would expect that efficient access to those polymers requires coordinated deconstruction of chemically complex, interleaved and crosslinked assemblages of polysaccharides, including cellulose. Indeed, the unified regulation of hemicellulases, pectinases and carbohydrate esterases by CLR-2 in *N. crassa*, even as apparently biochemically redundant enzymes or maintained under the control of other regulators, suggests a strong adaptive advantage to their co-expression with cellulases. The conserved co-regulation of such a diverse group of carbohydrate active enzymes by ClrB in *A. nidulans* suggests that such a synergistic strategy is widespread among filamentous fungi.

Just as conserved co-regulation of poorly characterized secreted enzymes with cellulases may point to conserved activities on common polysaccharide complexes or linkages, enzymes with non-conserved co-regulation may have niche or substrate specific advantage in cell wall deconstruction. The lytic polysaccharide monooxygenase (LPMO, CAZy class AA9) enzyme class is a particularly interesting example of divergent regulation across *N. crassa* and *A. nidulans*. Transcripts of LPMOs in *A. nidulans* were dominated by expression of a single gene, while several *N. crassa* LPMOs had similarly high induction. Further, while *A. nidulans* and *N. crassa* are extreme examples of this trend, a re-analysis of published RNAseq data for other fungal species on cellulosic substrates reveals two common strategies for LPMO expression (Figure 48). One LPMO dominated expression in *A. niger* on wheat straw [214] and just two were highly expressed in *T. reesei* on wheat straw [215]. However, several LPMOs were highly expressed in *Chaetomium globosum*, *Myceliophthora thermophilla*, and *Thielavia terrestris* on barley straw [216]. The

differing transcript abundances in *N. crassa* and *A. nidulans* are partly reflective of a different polygenetic distribution of LPMOs in the two species (Figure 48B), with many fewer *A. nidulans* LPMOs falling in the proposed subclasses. However, none of the LPMOs without an identified subclass were well expressed either. What is the adaptive advantage for *N. crassa* to maintain and express so many LPMOs?

The number of well-expressed LPMOs in *N. crassa* and their divergent regulation of LPMOs between fungal species may hold clues to as yet unappreciated functional differences amongs them. Indeed, since I first published this observation of divergent LPMO regulation [169], four subclasses of cellulose active LPMOs have been described based on phylogeny and the chemistry of their oxidative attack on cellulose. Subclass 1 and subclass 3S LPMOs oxidize the C1 position on glucose, subclass 2 LPMOs oxidize the C6 position and subclass 3 LPMOs oxidize either position [217]. In the transcriptomes analyzed so far, if a small number of LPMOs dominate expression, those LPMOs were all of the type 3 subclass. Also LPMOs have recently been shown to be active on xyloglucan [185] and on starch [184]. Given that many fungal genomes encode dozens of LPMOs (e.g. 43 predicted *C. globosum* LPMOs are listed in the CAZy database [16]), it is likely that more diverse enzymatic mechanisms and preferred substrates for this class of enzymes remain to be identified. The diverse expression patterns for differing subclades of LPMOs across species may provide clues to uncover those differences.

A. Transcript Abundance



B. Phylogenetic Distribution

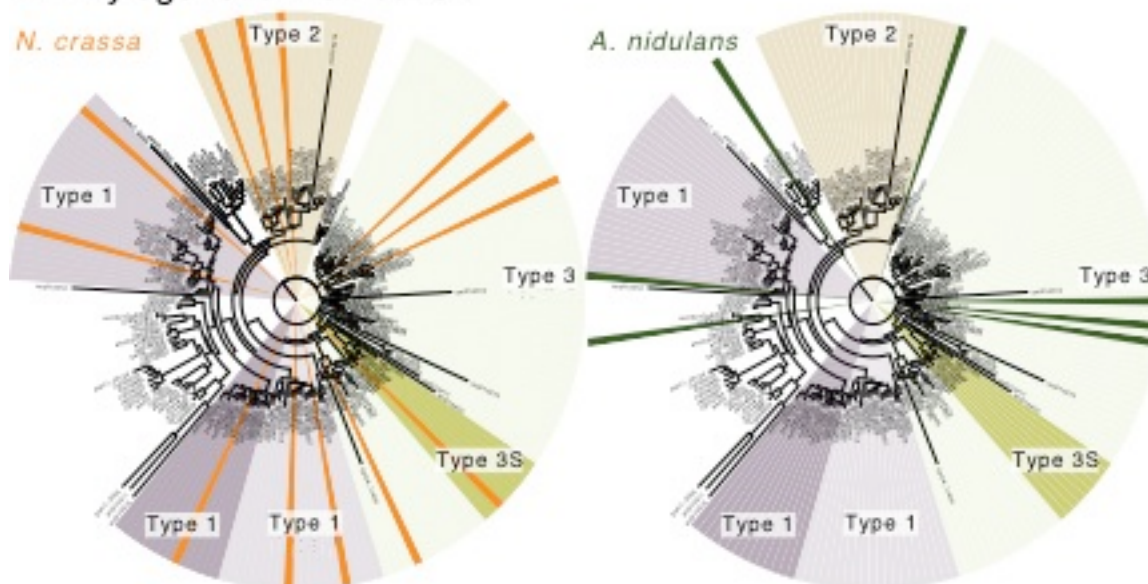


Figure 48. Relative expression and phylogenetic distribution of LPMOs (AA9 CAZy class) by subclass.

(A) Fraction of all AA9 FPKM by subclass. Blues = Class 1, Reds = Class 2, Greens = Class 3, Orange = Class 3S

(B) Phylogenetic tree of LPMOs in ascomycete fungi with distribution of *N. crassa* (orange) and *A. nidulans* (green). AA9 genes from *A. nidulans*, *A. niger*, *C. globosum*, *M. thermophilla*, *N. crassa*, *T. terrestris*, and *T. reesei* aligned with MAFFT [198], distance tree computed with FastTree [199].

Sugar sensing and secreted enzyme regulation by *N. crassa* and other filamentous ascomycetes

Over the course of this research, from 2009 to 2014, our knowledge of transcription factors regulating plant cell wall degradation has expanded from a single essential factor (XlnR/Xyr) with a few modulating factors to at least six primary response regulators and a similar number of additional modulating factors. My contributions to that list were the essential activators CLR-1 and CLR-2. Constitutive expression of CLR-2 was sufficient for cellulase secretion, raising the possibility that manipulation of a single conserved master regulator might control cellulase production in divergent production strains. That result was not directly translatable to the distantly related species *A. nidulans*. This discrepancy was just one in a cohort of experiments that suggest a diversity of function for cellulase regulators in ascomycete fungi.

Unpublished reports from colleagues suggest that a CLR-2 deletion in *T. reesei* shows no cellulolytic defect, and at least one published report found that over-expression of the CLR-2 ortholog had no effect on cellulase secretion on lactose [105]. Similar to *N. crassa* *clr-2*, in *A. nidulans*, *clrB* was absolutely required for cellulolytic activity in response to cellobiose. In *A. oryzae* ManR/ClrB regulated cellulases but in response to mannobiose, not cellobiose [180]. In *A. aculeatus*, ClbR, a ZnBNC transcription factor limited to closely related Aspergilli, was required for full induction of cellulases in response to cellobiose [108]. In *A. niger* and *T. reesei*,

the xylanolytic regulator XlnR/Xyr1 was also the primary regulator of cellulases in response to cellulose, although that function was not at all conserved in *N. crassa* [89], *A. aculeatus* [93], *A. nidulans* [94], *F. graminearum* [88], or *M. oryzae* [94]. Taken together, these data suggest diversity in the signaling pathways that activate cellulases through an assemblage of ZnBNC transcription factors. Perhaps the most surprising feature of this growing picture of diversity is that phylogenetic relationships among species and conservation of protein sequence have been a poor predictor of functional conservation amongst regulators. The last decade of research has enriched our understanding of this signaling diversity but left us with no workable framework to predict how orthologs of a cellulase regulator characterized in one species will function in another.

A common limitation in all of these studies has been a lack of mechanistic understanding of how these various regulators function. A greater understanding of the other proteins involved in the sensing of a various inducing sugars may yield understanding of these systems with greater predictive power. My attempt to address this critical gap took the form of a selection for mutants with constitutive expression of an antibiotic marker inserted at a cellulase locus. This selective screen has yielded several promising mutants that have begun reveal new aspects of *N. crassa*'s cellobiose sensing mechanism. The first new gene discovered from these mutants is NCU05846, a member of a small conserved family of genes limited to filamentous ascomycete species and encoding a domain of unknown function. NCU05846 acts to repress CLR-1 activation or effective transcriptional induction by a transiently/weakly active CLR-1 in non-inducing conditions.

Conservation of NCU05846 function in other species has yet to be shown, but its conservation profile is not inconsistent with apparent conservation of CLR-1 function. Our sparse sampling of regulator function across species and still preliminary understanding of the function of NCU05846 make any inferences low confidence indeed at this point. With the tools I have developed, more components in the cellobiose-sensing network will likely be identified soon. As our window into that network expands, so may be our ability to generalize to other systems. Experience to this point suggests, though, that cross-species, comparative studies will be required for quite some time if we are to fully understand cellulase regulation in its ecological context and for manipulation of industrial enzyme production.

References

1. (2014) U.S. Petroleum Products Supplied. U.S. Energy Information Administration.
2. (2014) Climate Change 2014 Synthesis Report. The Intergovernmental Panel on Climate Change (IPCC).
3. Maass P (2005) The Breaking Point. The New York Times Magazine. New York: Times Company.
4. Perlack RD, Stokes BJ (2011) Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry. . In: Agriculture Do, editor. ORNL/TM-2011/224 Oak Ridge National Laboratory, Oak Ridge, TN. pp. 227.
5. Beckman J, Borchers A, Jones C (2013) Agriculture's Supply and Demand for Energy and Energy Products. In: Service UER, editor: USDA.
6. Kominkova D, Kuehn KA, Busing N, Steiner D, Gessner MO (2000) Microbial biomass, growth, and respiration associated with submerged litter of *Phragmites australis* decomposing in a littoral reed stand of a large lake. *Aquatic Microbial Ecology* 22: 271-282.
7. Hieber M, Gessner MO (2002) Contribution of stream detritivores, fungi, and bacteria to leaf breakdown based on biomass estimates. *Ecology* 83: 1026-1038.
8. Heimann M, Reichstein M (2008) Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature* 451: 289-292.
9. Frey SD, Six J, Elliott ET (2003) Reciprocal transfer of carbon and nitrogen by decomposer fungi at the soil-litter interface. *Soil Biology & Biochemistry* 35: 1001-1004.
10. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, et al. (2012) The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. *Science* 336: 1715-1719.
11. Margeot A, Hahn-Hagerdal B, Edlund M, Slade R, Monot F (2009) New improvements for lignocellulosic ethanol. *Curr Opin Biotechnol* 20: 372-380.
12. Sainz MB (2011) Commercial Cellulosic Ethanol The Role of Plant-Expressed Enzymes. *Biofuels: Global Impact on Renewable Energy, Production Agriculture, and Technological Advancements*: 237-264.
13. Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, et al. (2004) Toward a systems approach to understanding plant cell walls. *Science* 306: 2206-2211.
14. Pauly M, Keegstra K (2010) Plant cell wall polymers as precursors for biofuels. *Curr Opin Plant Biol* 13: 305-312.
15. Popper ZA, Michel G, Herve C, Domozych DS, Willats WG, et al. (2011) Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu Rev Plant Biol* 62: 567-590.
16. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* 37: D233-D238.
17. Jordan DB, Bowman MJ, Braker JD, Dien BS, Hector RE, et al. (2012) Plant cell walls to ethanol. *Biochem J* 442: 241-252.

18. Gilbert HJ, Stalbrand H, Brumer H (2008) How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Curr Opin Plant Biol* 11: 338-348.
19. Dodd D, Cann IK (2009) Enzymatic deconstruction of xylan for biofuel production. *Glob Change Biol Bioenergy* 1: 2-17.
20. van den Brink J, de Vries RP (2011) Fungal enzyme sets for plant polysaccharide degradation. *Appl Microbiol Biotechnol* 91: 1477-1492.
21. Yadav S, Yadav PK, Yadav D, Yadav KDS (2009) Pectin lyase: A review. *Process Biochemistry* 44: 1-10.
22. Liu RG, Yu H, Huang Y (2005) Structure and morphology of cellulose in wheat straw. *Cellulose* 12: 25-34.
23. Adav SS, Chao LT, Sze SK (2012) Quantitative secretomic analysis of *Trichoderma reesei* strains reveals enzymatic composition for lignocellulosic biomass degradation. *Mol Cell Proteomics* 11: M111 012419.
24. Adav SS, Ravindran A, Sze SK (2012) Quantitative proteomic analysis of lignocellulolytic enzymes by *Phanerochaete chrysosporium* on different lignocellulosic biomass. *J Proteomics* 75: 1493-1504.
25. Bommarius AS, Sohn M, Kang Y, Lee JH, Realff MJ (2014) Protein engineering of cellulases. *Curr Opin Biotechnol* 29: 139-145.
26. Lambertz C, Garvey M, Klinger J, Heesel D, Klose H, et al. (2014) Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. *Biotechnol Biofuels* 7: 135.
27. Kubicek CP, Mikus M, Schuster A, Schmoll M, Seiboth B (2009) Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. *Biotechnol Biofuels* 2: 19.
28. Scheller HV, Ulvskov P (2010) Hemicelluloses. *Annu Rev Plant Biol* 61: 263-289.
29. Levy S, Maclachlan G, Staehelin LA (1997) Xyloglucan sidechains modulate binding to cellulose during in vitro binding assays as predicted by conformational dynamics simulations. *Plant J* 11: 373-386.
30. Popper ZA, Fry SC (2005) Widespread occurrence of a covalent linkage between xyloglucan and acidic polysaccharides in suspension-cultured angiosperm cells. *Ann Bot* 96: 91-99.
31. Benkő Z, Siika-aho M, Viikari L, Réczey K (2008) Evaluation of the role of xyloglucanase in the enzymatic hydrolysis of lignocellulosic substrates. *Enzyme and Microbial Technology* 43: 109-114.
32. van der Weijde T, Alvim Kamei CL, Torres AF, Vermerris W, Dolstra O, et al. (2013) The potential of C4 grasses for cellulosic biofuel production. *Front Plant Sci* 4: 107.
33. Grabber JH, Ralph J, Hatfield RD (2000) Cross-linking of maize walls by ferulate dimerization and incorporation into lignin. *J Agric Food Chem* 48: 6106-6113.
34. Atmodjo MA, Hao Z, Mohnen D (2013) Evolving views of pectin biosynthesis. *Annu Rev Plant Biol* 64: 747-779.
35. Zhong R, Ye ZH (2014) Secondary Cell Walls: Biosynthesis, Patterned Deposition and Transcriptional Regulation. *Plant Cell Physiol*.
36. Pandey MP, Kim CS (2011) Lignin Depolymerization and Conversion: A Review of Thermochemical Methods. *Chemical Engineering & Technology* 34: 29-41.

37. Yu Z, Gwak KS, Treasure T, Jameel H, Chang HM, et al. (2014) Effect of lignin chemistry on the enzymatic hydrolysis of woody biomass. *ChemSusChem* 7: 1942-1950.
38. Greetham D, Wimalasena T, Kerruish DW, Brindley S, Ibbett RN, et al. (2014) Development of a phenotypic assay for characterisation of ethanologenic yeast strain sensitivity to inhibitors released from lignocellulosic feedstocks. *J Ind Microbiol Biotechnol* 41: 931-945.
39. Coradetti ST, Craig JP, Xiong Y, Shock T, Tian CG, et al. (2012) Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *Proc Natl Acad Sci U S A* 109: 7397-7402.
40. Glass NL, Schmoll M, Cate JH, Coradetti S (2013) Plant cell wall deconstruction by ascomycete fungi. *Annu Rev Microbiol* 67: 477-498.
41. Turner BC, Perkins DD, Fairfield A (2001) *Neurospora* from natural populations: A global study. *Fungal Genet Biol* 32: 67-92.
42. Davis RH, Perkins DD (2002) *Neurospora*: a model of model microbes [Review]. *Nature Reviews Genetics* 3: 397-403.
43. Kuo HC, Hui S, Choi J, Asiegbu FO, Valkonen JP, et al. (2014) Secret lifestyles of *Neurospora crassa*. *Sci Rep* 4: 5135.
44. Tian C, Beeson WT, Iavarone AT, Sun J, Marletta MA, et al. (2009) Systems analysis of plant cell wall degradation by the model filamentous fungus *Neurospora crassa*. *Proc Nat Acad Sci USA* 106: 22157-22162.
45. Phillips CM, Iavarone AT, Marletta MA (2011) Quantitative proteomic approach for cellulose degradation by *Neurospora crassa*. *J Proteome Res* 10: 4177-4185.
46. Galazka JM, Tian C, Beeson WT, Martinez B, Glass NL, et al. (2010) Cellodextrin transport in yeast for improved biofuel production. *Science* 330: 84-86.
47. 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.
48. Beeson WT, Phillips CM, Cate JH, Marletta MA (2012) Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *J Am Chem Soc* 134: 890-892.
49. Beeson WT, Iavarone AT, Hausmann CD, Cate JH, Marletta MA (2011) Extracellular aldonolactonase from *Myceliophthora thermophila*. *Appl Environ Microbiol* 77: 650-656.
50. Benz JP, Protzko RJ, Andrich JM, Bauer S, Dueber JE, et al. (2014) Identification and characterization of a galacturonic acid transporter from *Neurospora crassa* and its application for *Saccharomyces cerevisiae* fermentation processes. *Biotechnol Biofuels* 7: 20.
51. Cai P, Gu R, Wang B, Li J, Wan L, et al. (2014) Evidence of a critical role for cellodextrin transporter 2 (CDT-2) in both cellulose and hemicellulose degradation and utilization in *Neurospora crassa*. *Plos One* 9: e89330.
52. Du J, Li S, Zhao H (2010) Discovery and characterization of novel d-xylose-specific transporters from *Neurospora crassa* and *Pichia stipitis*. *Mol Biosyst* 6: 2150-2156.

53. 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.
54. Nihira T, Saito Y, Nishimoto M, Kitaoka M, Igarashi K, et al. (2013) Discovery of cellobionic acid phosphorylase in cellulolytic bacteria and fungi. *FEBS Lett* 587: 3556-3561.
55. Davis RH, Perkins DD (2002) *Neurospora*: a model of model microbes. *Nature Reviews Genetics* 3: 397-403.
56. Timberlake W, Clutterbuck A, Martinelli S, Kinghorn J (1994) *Aspergillus*: 50 years on. *Aspergillus*: 50 years on.
57. Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B (2005) Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res* 15: 1620-1631.
58. Klich MA (2002) Biogeography of *Aspergillus* species in soil and litter. *Mycologia* 94: 21-27.
59. Kuhls K, Lieckfeldt E, Samuels GJ, Kovacs W, Meyer W, et al. (1996) Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina*. *Proceedings of the National Academy of Sciences* 93: 7755-7760.
60. Reese ET, Levinson HS (1952) A Comparative Study of the Breakdown of Cellulose by Microorganisms. *Physiologia Plantarum* 5: 345-366.
61. Le Crom S, Schackwitz W, Pennacchio L, Magnuson JK, Culley DE, et al. (2009) Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei* using massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 106: 16151-16156.
62. Kawamori M, Morikawa Y, Takasawa S (1986) Induction and production of cellulases by l-sorbose in *Trichoderma reesei*. *Appl Microbiol Biotechnol* 24: 449-453.
63. Derntl C, Gudynaite-Savitch L, Calixte S, White T, Mach RL, et al. (2013) Mutation of the Xylanase regulator 1 causes a glucose blind hydrolase expressing phenotype in industrially used *Trichoderma* strains. *Biotechnol Biofuels* 6: 62.
64. Roehr M, Zehentgruber O, Kubicek CP (1981) Kinetics of biomass formation and citric acid production by *Aspergillus niger* on pilot plant scale. *Biotechnol Bioeng* 23: 2433-2445.
65. Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, et al. (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat Biotechnol* 25: 221-231.
66. Machida M, Yamada O, Gomi K (2008) Genomics of *Aspergillus oryzae*: learning from the history of Koji mold and exploration of its future. *DNA Res* 15: 173-183.
67. Brown NA, Bass C, Baldwin TK, Chen H, Massot F, et al. (2011) Characterisation of the *Fusarium graminearum*-Wheat Floral Interaction. *J Pathog* 2011: 626345.
68. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, et al. (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434: 980-986.

69. Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, et al. (2012) The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology* 13: 414-430.
70. Mary Wanjiru W, Zhensheng K, Buchenauer H (2002) Importance of Cell Wall Degrading Enzymes Produced by *Fusarium graminearum* during Infection of Wheat Heads. *European Journal of Plant Pathology* 108: 803-810.
71. Mandels M, Reese ET (1957) Induction of Cellulase in *Trichoderma-Viride* as Influenced by Carbon Sources and Metals. *Journal of Bacteriology* 73: 269-278.
72. Mandels M, Reese ET (1960) Induction of cellulase in fungi by cellobiose. *J Bacteriol* 79: 816-826.
73. Eberhart BM (1961) Exogenous enzymes of *Neurospora conidia* and mycelia. *J Cell Comp Physiol* 58: 11-16.
74. Mandels M, Reese ET, Parrish FW (1962) Sophorose as an Inducer of Cellulase in *Trichoderma Viride*. *Journal of Bacteriology* 83: 400-&.
75. Myers MG, Eberhart B (1966) Regulation of cellulase and cellobiase in *Neurospora crassa*. *Biochem Biophys Res Commun* 24: 782-785.
76. Karaffa L, Fekete E, Gamauf C, Szentirmai A, Kubicek CP, et al. (2006) D-Galactose induces cellulase gene expression in *Hypocrea jecorina* at low growth rates. *Microbiology* 152: 1507-1514.
77. Ilmen M, Saloheimo A, Onnela ML, Penttila ME (1997) Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. *Appl Environ Microbiol* 63: 1298-1306.
78. MargollesClark E, Ilmen M, Penttila M (1997) Expression patterns of ten hemicellulase genes of the filamentous fungus *Trichoderma reesei* on various carbon sources. *Journal of Biotechnology* 57: 167-179.
79. van Peij NN, Visser J, de Graaff LH (1998) Isolation and analysis of xlnR, encoding a transcriptional activator co-ordinating xylanolytic expression in *Aspergillus niger*. *Mol Microbiol* 27: 131-142.
80. van Peij NNME, Gielkens MMC, de Vries RP, Visser J, de Graaff LH (1998) The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in *Aspergillus niger*. *Appl Environ Microbiol* 64: 3615-3619.
81. Gielkens MM, Dekkers E, Visser J, de Graaff LH (1999) Two cellobiohydrolase-encoding genes from *Aspergillus niger* require D-xylose and the xylanolytic transcriptional activator XlnR for their expression. *Appl Environ Microbiol* 65: 4340-4345.
82. de Vries RP, Visser J, de Graaff LH (1999) CreA modulates the XlnR-induced expression on xylose of *Aspergillus niger* genes involved in xylan degradation. *Res Microbiol* 150: 281-285.
83. Zeilinger S, Ebner A, Marosits T, Mach R, Kubicek CP (2001) The *Hypocrea jecorina* HAP 2/3/5 protein complex binds to the inverted CCAAT-box (ATTGG) within the cbh2 (cellobiohydrolase II-gene) activating element. *Mol Genet Genomics* 266: 56-63.
84. Zeilinger S, Haller M, Mach R, Kubicek CP (2000) Molecular characterization of a cellulase-negative mutant of *Hypocrea jecorina*. *Biochem Biophys Res Commun* 277: 581-588.

85. Aro N, Ilmen M, Saloheimo A, Penttilä M (2003) ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression. *Appl Environ Microbiol* 69: 56-65.
86. Aro N, Saloheimo A, Ilmen M, Penttilä M (2001) ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of *Trichoderma reesei*. *Journal of Biological Chemistry* 276: 24309-24314.
87. Stricker AR, Grosstessner-Hain K, Wurleitner E, Mach RL (2006) Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. *Eukaryotic Cell* 5: 2128-2137.
88. 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*. *Current Genetics* 52: 213-220.
89. Sun JP, Tian CG, Diamond S, Glass NL (2012) Deciphering Transcriptional Regulatory Mechanisms Associated with Hemicellulose Degradation in *Neurospora crassa*. *Eukaryotic Cell* 11: 482-493.
90. MacPherson S, Larochelle M, Turcotte B (2006) A fungal family of transcriptional regulators: The zinc cluster proteins. *Microbiology and Molecular Biology Reviews* 70: 583-+.
91. Egriboz O, Goswami S, Tao X, Dotts K, Schaeffer C, et al. (2013) Self-association of the Gal4 inhibitor protein Gal80 is impaired by Gal3: evidence for a new mechanism in the GAL gene switch. *Mol Cell Biol* 33: 3667-3674.
92. Battaglia E, Klaubauf S, Vallet J, Ribot C, Lebrun MH, et al. (2013) Xlr1 is involved in the transcriptional control of the pentose catabolic pathway, but not hemi-cellulolytic enzymes in *Magnaporthe oryzae*. *Fungal Genetics and Biology* 57: 76-84.
93. Tani S, Kanamasa S, Sumitani J, Arai M, Kawaguchi T (2012) XlnR-independent signaling pathway regulates both cellulase and xylanase genes in response to cellobiose in *Aspergillus aculeatus*. *Current Genetics* 58: 93-104.
94. Klaubauf S, Narang HM, Post H, Zhou M, Brunner K, et al. (2014) Similar is not the same: Differences in the function of the (hemi-)cellulolytic regulator XlnR (Xlr1/Xyr1) in filamentous fungi. *Fungal Genetics and Biology*.
95. Noguchi Y, Sano M, Kanamaru K, Ko T, Takeuchi M, et al. (2009) Genes regulated by AoXlnR, the xylanolytic and cellulolytic transcriptional regulator, in *Aspergillus oryzae*. *Appl Microbiol Biotechnol* 85: 141-154.
96. Mach-Aigner AR, Pucher ME, Steiger MG, Bauer GE, Preis SJ, et al. (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.
97. Calero-Nieto F, Di Pietro A, Roncero MI, 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.
98. Noguchi Y, Tanaka H, Kanamaru K, Kato M, Kobayashi T (2011) Xylose triggers reversible phosphorylation of XlnR, the fungal transcriptional activator of xylanolytic and cellulolytic genes in *Aspergillus oryzae*. *Biosci Biotechnol Biochem* 75: 953-959.

99. Sadowski I, Niedbala D, Wood K, Ptashne M (1991) GAL4 is phosphorylated as a consequence of transcriptional activation. *Proceedings of the National Academy of Sciences* 88: 10510-10514.
100. Hasper AA, Trindade LM, van der Veen D, van Ooyen AJ, de Graaff LH (2004) Functional analysis of the transcriptional activator XlnR from *Aspergillus niger*. *Microbiology* 150: 1367-1375.
101. Lichius A, Seidl - Seiboth V, Seiboth B, Kubicek CP (2014) Nucleo - cytoplasmic shuttling dynamics of the transcriptional regulators XYR1 and CRE1 under conditions of cellulase and xylanase gene expression in *Trichoderma reesei*. *Mol Microbiol* 94: 1162-1178.
102. Craig JP, Coradetti ST, Glass NL ((In preparation)) Invivo DNA binding of plant cell wall degradation regulators XLR-1, CLR-1 and CLR-2 in *Neurospora crassa*.
103. Wang D, Hu Y, Zheng F, Zhou K, Kohlhaw GB (1997) Evidence that intramolecular interactions are involved in masking the activation domain of transcriptional activator Leu3p. *J Biol Chem* 272: 19383-19392.
104. Guo H, Kohlhaw GB (1996) Regulation of transcription in mammalian cells by yeast Leu3p and externally supplied inducer. *FEBS Lett* 390: 191-195.
105. Hakkinen M, Valkonen MJ, Westerholm-Parvinen A, Aro N, Arvas M, et al. (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.
106. Nitta M, Furukawa T, Shida Y, Mori K, Kuhara S, et al. (2012) A new Zn(II)(2)Cys(6)-type transcription factor BglR regulates beta-glucosidase expression in *Trichoderma reesei*. *Fungal Genetics and Biology* 49: 388-397.
107. 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.
108. Kunitake E, Tani S, Sumitani JI, Kawaguchi T (2012) 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.
109. Battaglia E, Visser L, Nijssen A, van Veluw GJ, Wosten HA, et al. (2011) Analysis of regulation of pentose utilisation in *Aspergillus niger* reveals evolutionary adaptations in Eurotiales. *Stud Mycol* 69: 31-38.
110. Gruben BS, Zhou M, Wiebenga A, Ballering J, Overkamp KM, et al. (2014) *Aspergillus niger* RhaR, a regulator involved in L-rhamnose release and catabolism. *Appl Microbiol Biotechnol* 98: 5531-5540.
111. Pardo E, Orejas M (2014) The *Aspergillus nidulans* Zn(II) 2 Cys 6 transcription factor AN5673/RhaR mediates L-rhamnose utilization and the production of inverted question mark-L-rhamnosidases. *Microb Cell Fact* 13: 161.
112. Ogawa M, Kobayashi T, Koyama Y (2012) ManR, a novel Zn(II)2Cys6 transcriptional activator, controls the beta-mannan utilization system in *Aspergillus oryzae*. *Fungal Genetics and Biology* 49: 987-995.

113. Kulmburg P, Mathieu M, Dowzer C, Kelly J, Felenbok B (1993) Specific binding sites in the *alcR* and *alcA* promoters of the ethanol regulon for the CREA repressor mediating carbon catabolite repression in *Aspergillus nidulans*. *Mol Microbiol* 7: 847-857.
114. Sun JP, Glass NL (2011) Identification of the CRE-1 Cellulolytic Regulon in *Neurospora crassa*. *Plos One* 6: e25654.
115. Brown NA, Ries LN, Goldman GH (2014) How nutritional status signalling coordinates metabolism and lignocellulolytic enzyme secretion. *Fungal Genetics and Biology*.
116. Riederer A, Takasuka TE, Makino S, Stevenson DM, Bukhman YV, et al. (2011) Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. *Appl Environ Microbiol* 77: 1243-1253.
117. Arvas M, Pakula T, Smit B, Rautio J, Koivistoinen H, et al. (2011) Correlation of gene expression and protein production rate - a system wide study. *Bmc Genomics* 12: 616.
118. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, et al. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422: 859-868.
119. Vogel H (1956) A convenient growth medium for *Neurospora* (medium N). *Microb Genet Bull* 13: 2-43.
120. Charlang GW (1979) An improved glycerol minimal medium. *Neurospora Newsletter* 26: 20-21.
121. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, et al. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422: 859-868.
122. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 37: D233-238.
123. Ruepp A, Zollner A, Maier D, Albermann K, Hani J, et al. (2004) The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res* 32: 5539-5545.
124. Stricker AR, Grosstessner-Hain K, Wurleitner 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.
125. Portnoy T, Margeot A, Linke R, Atanasova L, Fekete E, et al. (2011) The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation. *Bmc Genomics* 12.
126. 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 Genetics and Biology* 40: 166-175.
127. Phillips CM, Iavarone AT, Marletta MA (2011) Quantitative Proteomic Approach for Cellulose Degradation by *Neurospora crassa*. *Journal of Proteome Research* 10: 4177-4185.

128. Znameroski EA, Coradetti ST, Roche CM, Tsai JC, Iavarone AT, et al. (2012) Induction of lignocellulose-degrading enzymes in *Neurospora crassa* by cellodextrins. *Proc Natl Acad Sci U S A* 109: 6012-6017.
129. Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, et al. (2011) Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proc Natl Acad Sci U S A* 108: 2831-2836.
130. Coradetti ST, Craig JP, Xiong Y, Shock T, Tian C, et al. (2012) Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *Proc Natl Acad Sci U S A* 109: 7397-7402.
131. Colot HV, Park G, Turner GE, Ringelberg C, Crew CM, et al. (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proc Natl Acad Sci USA* 103: 10352-10357.
132. Vogel HJ (1956) A convenient growth medium for *Neurospora*. *Microbiol Genet Bull* 13: 42-46.
133. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25.
134. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 12: R22.
135. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11: R106.
136. Kasuga T, Townsend JP, Tian C, Gilbert LB, Mannhaupt G, et al. (2005) Long-oligomer microarray profiling in *Neurospora crassa* reveals the transcriptional program underlying biochemical and physiological events of conidial germination. *Nucleic Acids Res* 33: 6469-6485.
137. Noguchi Y, Sano M, Kanamaru K, Ko T, Takeuchi M, et al. (2009) Genes regulated by AoXlnR, the xylanolytic and cellulolytic transcriptional regulator, in *Aspergillus oryzae*. *Applied Microbiology and Biotechnology* 85: 141-154.
138. Stricker AR, Mach RL, de Graaff LH (2008) Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). *Applied Microbiology and Biotechnology* 78: 211-220.
139. van Peij NN, Gielkens MM, de Vries RP, Visser J, de Graaff LH (1998) The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in *Aspergillus niger*. *Appl Environ Microbiol* 64: 3615-3619.
140. 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.
141. Aro N, Saloheimo A, Ilmen 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.
142. Tian C, Kasuga T, Sachs MS, Glass NL (2007) Transcriptional profiling of cross pathway control in *Neurospora crassa* and comparative analysis of the Gcn4 and CPC1 regulons. *Eukaryotic Cell* 6: 1018-1029.

143. Tian C, Beeson WT, Iavarone AT, Sun J, Marletta MA, et al. (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.
144. 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.
145. Lockington RA, Rodbourn L, Barnett S, Carter CJ, Kelly JM (2002) Regulation by carbon and nitrogen sources of a family of cellulases in *Aspergillus nidulans*. *Fungal Genet Biol* 37: 190-196.
146. Tilburn J, Sarkar S, Widdick DA, Espeso EA, Orejas M, et al. (1995) The *Aspergillus* PacC zinc finger transcription factor mediates regulation of both acid- and alkaline-expressed genes by ambient pH. *EMBO J* 14: 779-790.
147. Sun J, Glass NL (2011) Identification of the CRE-1 cellulolytic regulon in *Neurospora crassa*. *PLoS ONE* 6: e25654.
148. Koutelou E, Hirsch CL, Dent SY (2010) Multiple faces of the SAGA complex. *Curr Opin Cell Biol* 22: 374-382.
149. Luthra R, Kerr SC, Harreman MT, Apponi LH, Fasken MB, et al. (2007) Actively transcribed GAL genes can be physically linked to the nuclear pore by the SAGA chromatin modifying complex. *Journal of Biological Chemistry* 282: 3042-3049.
150. Green EM, Jiang Y, Joyner R, Weis K (2012) A negative feedback loop at the nuclear periphery regulates GAL gene expression. *Mol Biol Cell* 23: 1367-1375.
151. Winston F, Carlson M (1992) Yeast Snf Swi Transcriptional Activators and the Spt Sin Chromatin Connection. *Trends in Genetics* 8: 387-391.
152. Sudarsanam P, Winston F (2000) The Swi/Snf family - nucleosome-remodeling complexes and transcriptional control. *Trends in Genetics* 16: 345-351.
153. Xiang QJ, Glass NL (2002) Identification of *vib1*, a locus involved in vegetative incompatibility mediated by *het-c* in *Neurospora crassa*. *Genetics* 162: 89-101.
154. Campbell RN, Leverentz MK, Ryan LA, Reece RJ (2008) Metabolic control of transcription: paradigms and lessons from *Saccharomyces cerevisiae*. *Biochemical Journal* 414: 177-187.
155. Cang Y, Prelich G (2002) Direct stimulation of transcription by negative cofactor 2 (NC2) through TATA-binding protein (TBP). *Proc Natl Acad Sci U S A* 99: 12727-12732.
156. Dunlap JC, Borkovich KA, Henn MR, Turner GE, Sachs MS, et al. (2007) Enabling a Community to Dissect an Organism: Overview of the *Neurospora* Functional Genomics Project. *Advances in Genetics: Academic Press*. pp. 49-96.
157. McCluskey K, Wiest A, Plamann M (2010) The Fungal Genetics Stock Center: a repository for 50 years of fungal genetics research. *J Biosci* 35: 119-126.
158. Wood TM, Bhat KM (1988) Methods for Measuring Cellulase Activities. *Methods Enzymol* 160: 87-112.

159. de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* 20: 1453-1454.
160. Saldanha AJ (2004) Java Treeview--extensible visualization of microarray data. *Bioinformatics* 20: 3246-3248.
161. Colot HV, Park G, Turner GE, Ringelberg C, Crew CM, et al. (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proc Natl Acad Sci U S A* 103: 10352-10357.
162. Elvin M, Loros JJ, Dunlap JC, Heintzen C (2005) The PAS/LOV protein VIVID supports a rapidly dampened daytime oscillator that facilitates entrainment of the *Neurospora* circadian clock. *Genes & Development* 19: 2593-2605.
163. Galazka JM, Tian CG, Beeson WT, Martinez B, Glass NL, et al. (2010) Cellodextrin Transport in Yeast for Improved Biofuel Production. *Science* 330: 84-86.
164. Xiong Y, Coradetti ST, Li X, Gritsenko MA, Clauss T, et al. (2014) The proteome and phosphoproteome of *Neurospora crassa* in response to cellulose, sucrose and carbon starvation. *Fungal Genetics and Biology*.
165. Levine SE, Fox JM, Blanch HW, Clark DS (2010) A mechanistic model of the enzymatic hydrolysis of cellulose. *Biotechnol Bioeng* 107: 37-51.
166. Znameroski EA, Coradetti ST, Roche CM, Tsai JC, Iavarone AT, et al. ((in press)) Induction of lignocellulose degrading enzymes in *Neurospora crassa* by cellodextrins. *Proc Natl Acad Sci USA*.
167. Murugan R (2012) Theory on the dynamics of feedforward loops in the transcription factor networks. *Plos One* 7: e41027.
168. Benko Z, Siika-aho M, Viikari L, Reczey K (2008) Evaluation of the role of xyloglucanase in the enzymatic hydrolysis of lignocellulosic substrates. *Enzyme Microb Technol* 43: 109-114.
169. 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.
170. Freitag M, Hickey PC, Raju NB, Selker EU, Read ND (2004) GFP as a tool to analyze the organization, dynamics and function of nuclei and microtubules in *Neurospora crassa*. *Fungal Genetics and Biology* 41: 897-910.
171. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 12.
172. Stricker AR, Steiger MG, Mach RL (2007) Xyr1 receives the lactose induction signal and regulates lactose metabolism in *Hypocrea jecorina*. *FEBS Lett* 581: 3915-3920.
173. Hakkinen M, Arvas M, Oja M, Aro N, Penttila M, et al. (2012) Re-annotation of the CAZy genes of *Trichoderma reesei* and transcription in the presence of lignocellulosic substrates. *Microb Cell Fact* 11: 134.
174. Sternberg D, Mandels GR (1979) Induction of cellulolytic enzymes in *Trichoderma reesei* by sophorose. *J Bacteriol* 139: 761-769.
175. Andersen MR, Vongsangnak W, Panagiotou G, Salazar MP, Lehmann L, et al. (2008) A trispecies *Aspergillus* microarray: comparative transcriptomics of three *Aspergillus* species. *Proc Natl Acad Sci U S A* 105: 4387-4392.

176. Porciuncula Jde O, Furukawa T, Mori K, Shida Y, Hirakawa H, et al. (2013) Single nucleotide polymorphism analysis of a *Trichoderma reesei* hyper-cellulolytic mutant developed in Japan. *Biosci Biotechnol Biochem* 77: 534-543.
177. Liu G, Zhang L, Qin Y, Zou G, Li Z, et al. (2013) Long-term strain improvements accumulate mutations in regulatory elements responsible for hyper-production of cellulolytic enzymes. *Sci Rep* 3: 1569.
178. Perkins DD, Turner BC (1988) *Neurospora* from natural populations: toward the population biology of a haploid eukaryote. *Exp Mycol* 12: 91-131.
179. Tretter ED, Johnson EM, Wang Y, Kandel P, White MM (2013) Examining new phylogenetic markers to uncover the evolutionary history of early-diverging fungi: comparing MCM7, TSR1 and rRNA genes for single- and multi-gene analyses of the Kickxellomycotina. *Persoonia* 30: 106-125.
180. Ogawa M, Kobayashi T, Koyama Y (2013) ManR, a transcriptional regulator of the beta-mannan utilization system, controls the cellulose utilization system in *Aspergillus oryzae*. *Biosci Biotechnol Biochem* 77: 426-429.
181. Updegraff DM (1969) Semimicro determination of cellulose in biological materials. *Anal Biochem* 32: 420-424.
182. Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen JC, et al. (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci U S A* 108: 15079-15084.
183. Beeson WT, Phillips CM, Cate JHD, Marletta MA (2012) Oxidative Cleavage of Cellulose by Fungal Copper-Dependent Polysaccharide Monooxygenases. *Journal of the American Chemical Society* 134: 890-892.
184. Vu VV, Beeson WT, Span EA, Farquhar ER, Marletta MA (2014) A family of starch-active polysaccharide monooxygenases. *Proc Natl Acad Sci U S A* 111: 13822-13827.
185. Agger JW, Isaksen T, Varnai A, Vidal-Melgosa S, Willats WG, et al. (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.
186. Li X, Beeson WT, Phillips CM, Marletta MA, Cate JH (2012) Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure* 20: 1051-1061.
187. Konno N, Igarashi K, Habu N, Samejima M, Isogai A (2009) Cloning of the *Trichoderma reesei* cDNA encoding a glucuronan lyase belonging to a novel polysaccharide lyase family. *Appl Environ Microbiol* 75: 101-107.
188. Bauer S, Vasu P, Persson S, Mort AJ, Somerville CR (2006) Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. *Proc Natl Acad Sci U S A* 103: 11417-11422.
189. 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.
190. McCluskey K (2003) The fungal genetics stock center from molds to molecules. *Advances in Applied Microbiology* 52: 245-262.

191. Szewczyk E, Nayak T, Oakley CE, Edgerton H, Xiong Y, et al. (2006) Fusion PCR and gene targeting in *Aspergillus nidulans*. *Nat Protoc* 1: 3111-3120.
192. Punt PJ, Dingemanse MA, Kuyvenhoven A, Soede RD, Pouwels PH, et al. (1990) Functional elements in the promoter region of the *Aspergillus nidulans* *gpdA* gene encoding glyceraldehyde-3-phosphate dehydrogenase. *Gene* 93: 101-109.
193. Waring RB, May GS, Morris NR (1989) Characterization of an inducible expression system in *Aspergillus nidulans* using *alcA* and tubulin-coding genes. *Gene* 79: 119-130.
194. Nayak T, Szewczyk E, Oakley CE, Osmani A, Ukil L, et al. (2006) A versatile and efficient gene-targeting system for *Aspergillus nidulans*. *Genetics* 172: 1557-1566.
195. Szewczyk E, Nayak T, Oakley CE, Edgerton H, Xiong Y, et al. (2006) Fusion PCR and gene targeting in *Aspergillus nidulans*. *Nature Protocols* 1: 3111-3120.
196. Vishniac W, Santer M (1957) The thiobacilli. *Bacteriol Rev* 21: 195-213.
197. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, et al. (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438: 1105-1115.
198. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9: 286-298.
199. Price MN, Dehal PS, Arkin AP (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. *Plos One* 5: e9490.
200. Battaglia E, Hansen SF, Leendertse A, Madrid S, Mulder H, et al. (2011) Regulation of pentose utilisation by *AraR*, but not *XlnR*, differs in *Aspergillus nidulans* and *Aspergillus niger*. *Appl Microbiol Biotechnol* 91: 387-397.
201. Benz JP, Chau BH, Zheng D, Bauer S, Glass NL, et al. (2014) A comparative systems analysis of polysaccharide-elicited responses in *Neurospora crassa* reveals carbon source-specific cellular adaptations. *Mol Microbiol* 91: 275-299.
202. Gritz L, Davies J (1983) Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* 25: 179-188.
203. Austin B, Hall RM, Tyler BM (1990) Optimized vectors and selection for transformation of *Neurospora crassa* and *Aspergillus nidulans* to bleomycin and phleomycin resistance. *Gene* 93: 157-162.
204. Malling HV, De Serres FJ (1970) Genetic effects of N-methyl-N'-nitro-N-nitrosoguanidine in *Neurospora crassa*. *Mol Gen Genet* 106: 195-207.
205. Palma-Guerrero J, Leeder AC, Welch J, Glass NL (2014) Identification and characterization of LFD1, a novel protein involved in membrane merger during cell fusion in *Neurospora crassa*. *Mol Microbiol* 92: 164-182.
206. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43: 491-498.
207. Sandoval NR, Kim JY, Glebes TY, Reeder PJ, Aucoin HR, et al. (2012) Strategy for directing combinatorial genome engineering in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 109: 10540-10545.

208. Orth JD, Palsson B (2012) Gap-filling analysis of the iJO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions. *BMC systems biology* 6: 30.
209. Liu C-x, Wu F-l, Jiang H-w, He X, Guo S-j, et al. (2014) Global identification of CobB interactors by an *Escherichia coli* proteome microarray. *Acta Biochimica et Biophysica Sinica*: gmu038.
210. Gerik KJ, Lam W, Lodge JK (2014) Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution leading to virulence attenuation.
211. Ghosh S, Rao KH, Bhavesh NS, Das G, Dwivedi VP, et al. (2014) N-Acetylglucosamine (GlcNAc)-Inducible Gene GIG2 Is a Novel Component of GlcNAc Metabolism in *Candida albicans*. *Eukaryotic Cell* 13: 66-76.
212. Guillemette T, van Peij N, Goosen T, Lanthaler K, Robson GD, et al. (2007) Genomic analysis of the secretion stress response in the enzyme-producing cell factory *Aspergillus niger*. *Bmc Genomics* 8: 158.
213. Pakula TM, Laxell M, Huuskonen A, Uusitalo J, Saloheimo M, et al. (2003) The effects of drugs inhibiting protein secretion in the filamentous fungus *Trichoderma reesei*. Evidence for down-regulation of genes that encode secreted proteins in the stressed cells. *J Biol Chem* 278: 45011-45020.
214. Delmas S, Pullan ST, Gaddipati S, Kokolski M, Malla S, et al. (2012) Uncovering the genome-wide transcriptional responses of the filamentous fungus *Aspergillus niger* to lignocellulose using RNA sequencing. *PLoS Genet* 8: e1002875.
215. Ries L, Pullan ST, Delmas S, Malla S, Blythe MJ, et al. (2013) Genome-wide transcriptional response of *Trichoderma reesei* to lignocellulose using RNA sequencing and comparison with *Aspergillus niger*. *Bmc Genomics* 14: 541.
216. Berka RM, Grigoriev IV, Otillar R, Salamov A, Grimwood J, et al. (2011) Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nat Biotechnol* 29: 922-927.
217. Vu VV, Beeson WT, Phillips CM, Cate JH, Marletta MA (2014) Determinants of regioselective hydroxylation in the fungal polysaccharide monooxygenases. *Journal of the American Chemical Society* 136: 562-565.

Appendix A: The *N. crassa* Avicel Regulon

NCU	Annotation (Broad Inst)	XLR-1 Bound*	CLR-2 Bound*	Bound* CLR-1	Bound* CLR-1	WT Avicel	WT No Carbon	Δclr-1 Avicel	Δclr-2 Avicel	Pccg-1:clr-2 No Carbon	ΔNCU05846 No Carbon
NCU04689	hypothetical protein	-	-	-	-	1	0	1	1	0	0
NCU04169	hypothetical protein	-	-	-	-	487	99	1101	1319	53	63
NCU04170	hypothetical protein	-	-	-	-	35	12	78	86	12	16
NCU06983	hypothetical protein	-	-	-	-	72	26	143	134	17	29
NCU07737	salicylate hydroxylase	-	-	-	-	96	34	269	193	22	57
NCU09904	glucan 1,3-beta-glucosidase	-	-	-	-	10	4	20	19	2	4
NCU04556	hypothetical protein	-	-	-	-	1	0	1	1	0	2
NCU04557	hypothetical protein	-	-	-	-	3	1	3	2	1	3
NCU01866	lactonohydrolase	-	-	-	-	31	9	27	30	5	3
NCU10106	mitochondrial triosephosphate isomerase	Y	-	-	-	54	13	40	38	14	11
NCU04400	hypothetical protein	Y	-	-	-	233	14	169	153	62	5
NCU10107	ribose 5-phosphate isomerase	Y	-	-	-	591	40	355	351	118	11
NCU00729	hypothetical protein	-	-	-	-	239	74	373	352	90	131
NCU11353	D-xylulose kinase	Y	-	-	-	181	30	226	194	47	52
NCU07510	hypothetical protein	Y	-	-	-	27	0	22	25	1	1
NCU08384	xylose reductase	Y	-	Y	-	2293	11	1585	1464	18	46
NCU09705	GAL10	Y	-	-	-	521	39	344	315	79	71
NCU06143	hypothetical protein	Y	-	-	-	587	20	497	433	62	36
NCU10110	3-hydroxyisobutyrate dehydrogenase	Y	-	-	-	116	9	82	76	14	8
NCU08189	glycosylhydrolase family 10-2	Y	-	-	-	5675	50	7061	6223	510	249
NCU00891	xylitol dehydrogenase	Y	-	-	-	546	30	348	314	89	79
NCU01900	xylosidase/arabinoxidase	Y	Y	-	-	1340	26	694	726	152	45
NCU04401	fructose-bisphosphate aldolase	Y	-	-	-	411	11	187	190	36	11
NCU08383	hypothetical protein	-	-	-	-	2	0	2	2	0	0
NCU00292	cholinesterase	Y	-	-	-	419	20	583	506	35	19
NCU09652	beta-xylosidase	Y	-	-	-	364	50	417	361	84	56
NCU00709	glycosylhydrolase family 3-8	Y	-	-	-	32	2	44	34	3	3
NCU02343	alpha-L-arabinofuranosidase 2	Y	Y	-	-	1116	116	1786	1642	140	189
NCU05350	hypothetical protein	Y	-	-	-	48	8	38	39	6	6
NCU10697	hypothetical protein	-	-	Y	-	21	3	21	22	3	4
NCU09924	BNR/Asp-box repeat protein	-	-	-	-	18	3	8	7	3	4
NCU09923	glycosylhydrolase family 3-7	-	-	-	-	32	2	15	12	3	5
NCU03903	lipase/esterase	-	-	-	-	46	16	37	31	15	16
NCU08752	acetylcholinesterase	Y	-	-	-	324	23	60	99	19	22
NCU06364	GDH lipase/acylhydrolase	-	-	-	-	128	23	64	64	26	18
NCU09926	hypothetical protein	-	-	-	-	6	1	3	5	1	2
NCU06138	quinase permease	Y	-	-	-	560	22	548	576	430	64
NCU04475	lipase B	-	-	-	-	470	22	300	156	148	46
NCU06384	MFS sugar transporter	Y	-	-	-	10	3	10	11	4	1
NCU05591	ABC transporter CDR4	-	-	-	-	50	14	83	89	24	3
NCU07546	multidrug resistance protein MDR	-	-	-	-	33	12	42	48	14	6
NCU01136	hypothetical protein	-	-	-	-	14	6	15	18	6	11
NCU05068	hypothetical protein	-	-	Y	-	347	9	464	433	4	18
NCU01430	hypothetical protein	Y	-	-	-	69	19	47	45	19	29
NCU07055	monooxygenase	-	-	-	-	22	5	15	13	3	12
NCU09415	hypothetical protein	-	-	-	-	20	3	10	14	2	4
NCU06490	hypothetical protein	Y	Y	-	-	183	44	183	153	21	20
NCU06991	hypothetical protein	-	-	-	-	12	3	11	12	1	1
NCU05134	hypothetical protein	-	-	Y	-	222	68	130	160	4	14
NCU08038	CAS1	-	-	-	-	19	11	14	14	6	7
NCU05580	membrane transporter	-	-	-	-	2	0	1	1	0	0
NCU01320	glutathione S-transferase-3	-	-	-	-	204	33	123	97	19	11

NCU04720	nitrate nonutilizer-6	Y	Y	-	211	50	138	114	34	23
NCU01003	hypothetical protein	-	-	-	0	0	0	0	0	0
NCU00761	triacylglycerol lipase	-	-	-	2	2	1	1	2	1
NCU04467	hypothetical protein	-	-	-	49	17	51	46	136	12
NCU07705	cellulose degradation regulator-1	Y	-	-	108	32	2	74	137	48
NCU08371	hypothetical protein	-	-	-	78	35	66	55	142	57
NCU00710	acetyl xylan esterase	Y	-	-	140	2	5	4	189	6
NCU06373	hypothetical protein	-	-	-	25	6	10	10	85	7
NCU08867	hypothetical protein	-	Y	-	14	3	5	5	8	7
NCU04870	acetyl xylan esterase	Y	Y	-	1230	1	59	37	1525	422
NCU05159	acetylxytan esterase	Y	-	-	2764	16	147	92	1218	521
NCU07224	monooxygenase	-	-	-	16	0	4	2	14	3
NCU07225	glycosylhydrolase 11-2	Y	Y	Y	15308	30	1647	826	15714	1443
NCU03322	GDSL family lipase	Y	-	-	256	16	164	127	255	74
NCU07222	hypothetical protein	-	-	-	19	2	4	3	29	2
NCU10039	hypothetical protein	Y	-	-	104	15	41	40	63	20
NCU08687	galactokinase	-	-	-	202	31	47	49	81	72
NCU09485	chaperone dnaK	-	-	-	114	37	46	47	65	62
NCU00169	translocation complex component	-	-	-	164	58	67	65	98	76
NCU00813	disulfide isomerase	-	-	-	238	92	91	92	183	141
NCU06333	translocation protein SEC62	-	-	-	155	44	42	38	124	58
NCU09223	protein disulfide-isomerase	-	-	-	1274	303	393	361	979	381
NCU03982	glucose regulated protein 78	-	-	-	1763	310	426	417	1163	406
NCU04854	glycosylhydrolase family 7-2	-	Y	-	182	5	5	6	55	8
NCU09098	tetracycline transporter	-	-	-	35	9	9	9	25	11
NCU11278	hypothetical protein	-	-	-	62	10	10	9	51	16
NCU08897	protein transporter SEC61 subunit alpha	-	-	-	431	122	146	149	214	130
NCU03083	hypothetical protein	-	-	-	99	35	45	43	60	40
NCU09265	calreticulin	Y	-	-	640	157	195	198	255	204
NCU10721	solute carrier family 35 member B1 protein	-	-	-	64	17	21	19	24	22
NCU06704	hypothetical protein	-	-	Y	160	44	37	44	53	30
NCU08379	hypothetical protein	-	-	-	481	153	141	151	196	104
NCU04127	hypothetical protein	-	-	-	494	193	166	172	290	110
NCU00871	hypothetical protein	-	-	-	3	1	1	0	9	2
NCU01148	methyltransferase	-	-	-	14	6	6	7	24	5
NCU07143	6-phosphogluconolactonase	-	Y	-	1733	23	68	56	2039	206
NCU10045	pectinesterase	-	-	-	254	19	53	41	487	53
NCU09976	rhamnogalacturonan acetyltransferase	-	-	-	12	1	1	1	36	3
NCU03328	glycosylhydrolase family 61-6	-	Y	-	919	30	27	30	4463	73
NCU04494	acetyl xylan esterase	-	-	-	62	12	11	11	190	18
NCU05751	cellulose-binding protein	-	-	-	161	2	1	2	515	6
NCU02855	glycosylhydrolase family 11-1	Y	Y	-	325	3	4	3	3054	25
NCU09664	acetylxytan esterase	-	-	-	143	1	1	1	428	5
NCU09582	chitin deacetylase	Y	Y	-	661	4	9	9	2461	31
NCU08176	pectate lyase A	-	-	-	110	1	3	2	464	15
NCU05598	ascus development-1	-	-	-	51	5	6	6	296	21
NCU04997	glycosylhydrolase family 10-3	-	-	-	19	1	2	1	322	5
NCU03329	hypothetical protein	-	Y	-	119	37	42	38	341	52
NCU05121	glycosylhydrolase family 45-1	-	Y	-	89	8	6	6	2104	20
NCU09774	cellulase	-	-	-	2	0	0	0	248	0
NCU01077	hypothetical protein	-	-	-	6	2	2	2	85	4
NCU09775	alpha-N-arabinofuranosidase	-	Y	-	13	1	2	2	89	7
NCU05104	glycosylhydrolase family 7-4	-	-	-	1	0	0	0	4	1
NCU07897	hypothetical protein	-	Y	-	11	3	3	4	14	7
NCU09523	hypothetical protein	-	Y	-	26	7	6	5	31	14
NCU00764	hypothetical protein	-	-	-	2	0	0	0	1	0
NCU09491	feruloyl esterase B	-	-	-	51	3	3	3	256	314
NCU06607	hypothetical protein	-	-	-	93	3	4	5	329	276
NCU08785	fungal cellulose binding domain-containing protein	-	-	-	929	5	3	0	12430	1681
NCU01049	hypothetical protein	-	Y	Y	7	0	0	0	24	1
NCU01076	hypothetical protein	-	Y	-	157	1	1	1	2626	34
NCU07339	hypothetical protein	-	Y	Y	33	2	2	2	31	16

NCU00870	hypothetical protein	-	Y	-	965	19	43	24	1072	279
NCU08398	aldose 1-epimerase	-	Y	-	682	11	18	4	3107	327
NCU03181	acetylxyylan esterase	-	Y	-	300	7	8	5	561	123
NCU05924	glycosylhydrolase family 10-1	-	Y	-	3081	1	4	1	11856	956
NCU07326	hypothetical protein	-	Y	-	741	7	6	7	2789	304
NCU00449	hypothetical protein	-	-	-	178	2	2	2	499	55
NCU07898	endoglucanase IV	-	Y	-	1340	4	5	6	2885	387
NCU00206	cellobiose dehydrogenase-1	-	Y	-	1491	3	3	3	2177	181
NCU08760	glycosylhydrolase family 61-5	-	Y	-	4091	5	6	4	9788	690
NCU02240	glycosylhydrolase family 61-1	-	Y	-	2501	10	9	9	5241	376
NCU00763	hypothetical protein	-	-	-	16	0	0	0	11	2
NCU00762	glycosylhydrolase family 5-1	-	Y	-	4625	2	10	5	2411	180
NCU05955	Cel74a	-	Y	Y	681	15	15	17	446	104
NCU05056	hypothetical protein	-	-	-	4	0	0	0	3	2
NCU05864	hypothetical protein	-	Y	Y	1365	10	21	12	434	503
NCU09524	hypothetical protein	-	Y	-	252	3	3	2	128	91
NCU05057	glycosylhydrolase family 7-1	-	Y	Y	1614	4	6	7	677	350
NCU00836	glycosylhydrolase family 61-7	-	Y	-	1879	2	3	2	1042	399
NCU09680	glycosylhydrolase family 6-2	-	Y	Y	13444	21	24	21	8942	7797
NCU07340	cellobiohydrolase-1	-	Y	Y	15537	20	27	25	12144	4987
NCU01050	glycosylhydrolase family 61-4	-	Y	Y	4946	0	1	0	1323	1321
NCU08042	cellulose degradation regulator-2	-	-	Y	172	2	2	0	278	34
NCU07760	glycosylhydrolase family 61-2	-	Y	-	34	3	5	5	39	18
NCU04952	glycosyl hydrolase family 3-4	-	-	-	91	13	6	15	200	146
NCU02059	endothiapepsin	-	Y	-	135	19	26	32	183	143
NCU04871	hypothetical protein	-	Y	-	2	0	1	0	3	3
NCU03433	hypothetical protein	-	-	-	13	2	2	3	6	4
NCU03725	vegetative incompatibility blocked-1	Y	-	Y	167	41	45	52	63	72
NCU08624	hypothetical protein	-	-	-	32	4	4	4	5	12
NCU02344	fungal cellulose binding domain-containing protein	-	Y	-	51	2	2	2	6	14
NCU00800	hypothetical protein	-	-	-	1	0	0	0	0	0
NCU00801	celloextrin transport-1	-	Y	-	1900	49	73	203	1040	1500
NCU02915	hypothetical protein	-	Y	Y	121	17	23	25	66	63
NCU02009	FreB	-	Y	Y	488	65	92	99	270	235
NCU09416	cellulose-binding GDSL lipase/acylhydrolase	-	Y	-	291	0	1	1	22	20
NCU09764	hypothetical protein	-	Y	-	534	2	2	3	16	62
NCU02916	glycosylhydrolase family 61-3	-	Y	Y	1295	6	7	9	130	821
NCU07190	glycosylhydrolase family 6-3	-	Y	Y	4341	11	13	33	535	1007
NCU05846	hypothetical protein	-	Y	Y	395	22	31	66	128	0
NCU00130	glycosylhydrolase family 1-1	Y	-	Y	5606	31	107	263	600	2473
NCU09445	Cip2	-	-	-	4	0	0	0	2	1
NCU08412	glycosylhydrolase family 2-1	-	Y	Y	1233	26	116	1	575	586
NCU00890	glycosylhydrolase family 2-1	-	Y	-	48	2	5	3	16	16
NCU07336	hypothetical protein	-	-	-	12	3	4	4	9	6
NCU04998	hypothetical protein	-	-	-	3	1	2	4	3	2
NCU05501	hypothetical protein	-	-	-	55	12	20	19	19	59
NCU08750	isoamyl alcohol oxidase	-	-	Y	204	51	61	115	65	214
NCU04618	hypothetical protein	-	-	-	20	6	8	14	7	18
NCU02600	DUF1479 domain-containing protein	-	-	-	4	1	1	1	1	13
NCU08746	starch binding domain-containing protein	-	-	-	56	10	9	26	8	216
NCU04948	hypothetical protein	-	-	-	2	0	1	1	0	20
NCU05855	O-methyltransferase	-	Y	Y	26	6	9	9	10	41
NCU07979	hypothetical protein	-	-	Y	14	5	6	8	7	12
NCU06971	xylan degradation regulator-1	-	Y	Y	159	57	60	79	67	94
NCU09689	hypothetical protein	-	-	Y	622	16	17	172	15	405
NCU08755	glycosylhydrolase 3-3	-	-	Y	1126	62	67	330	39	423
NCU07487	glycosylhydrolase family 3-6	-	Y	Y	129	1	1	2	1	96
NCU05574	hypothetical protein	-	-	Y	74	15	9	28	16	67
NCU04522	hypothetical protein	-	-	-	185	3	2	10	10	296
NCU05848	cytochrome P450 monooxygenase	-	-	-	3	0	0	0	0	9
NCU05863	hypothetical protein	-	-	Y	75	5	2	8	4	39

NCU00326	calcium homeostasis protein Regucalcin	-	Y	Y	175	59	35	74	76	154
NCU04623	glycosylhydrolase family 35-2	-	-	-	19	8	7	13	13	35
NCU01059	glycosylhydrolase family 47-3	-	-	Y	33	7	5	13	8	29
NCU01058	hypothetical protein	-	-	-	12	2	2	6	4	12
NCU08756	hypothetical protein	-	-	-	17	2	2	7	1	14
NCU08043	hypothetical protein	-	Y	Y	8	2	1	2	0	2
NCU09176	hypothetical protein	-	-	-	53	9	16	20	1	16
NCU02882	hypothetical protein	-	Y	Y	26	7	10	13	2	16
NCU10521	glutathione S-transferase-4	-	-	-	63	12	23	21	8	15
NCU08117	hypothetical protein	-	-	-	24	7	8	9	5	8
NCU08113	hypothetical protein	-	-	-	27	1	1	5	1	2
NCU11342	MFS hexose transporter	-	-	Y	17	4	5	7	3	9
NCU09172	hypothetical protein	-	-	-	1	0	0	0	0	0
NCU08790	hypothetical protein	-	-	-	65	2	5	7	2	7
NCU09295	hypothetical protein	Y	-	-	32	13	19	21	6	21
NCU08116	hypothetical protein	-	-	-	2	0	1	2	1	1
NCU05853	MFS sugar transporter	-	-	Y	4114	322	1446	3237	758	2092
NCU08114	cellodextrin transport-2	Y	Y	Y	8798	93	1834	3996	503	3510
NCU07270	hypothetical protein	-	-	-	10	4	5	5	3	5
NCU02485	hypothetical protein	-	Y	Y	31	3	8	10	3	10
NCU08115	mutS ortholog 3	Y	Y	Y	56	7	13	24	6	21
NCU09425	Cellobionic acid phosphorylase	-	-	Y	433	128	196	258	151	202
NCU10344	cytochrome P450 98A3	-	-	-	1	0	1	1	0	2
NCU11118	hypothetical protein	-	-	Y	54	21	13	19	18	18
NCU07668	MFS multidrug transporter	-	-	-	22	6	9	10	9	7
NCU01944	hypothetical protein	Y	Y	Y	58	8	5	7	3	11
NCU05842	hypothetical protein	-	-	-	4	1	0	0	0	1
NCU03208	hypothetical protein	-	-	-	9	1	5	4	1	2
NCU04928	hypothetical protein	-	-	-	3	1	2	2	1	2
NCU10383	Removed									
NCU11068	Removed									
NCU11198	Removed									
NCU11542	Removed									
NCU11722	Removed									
NCU11828	Removed									
NCU11932	Removed									
NCU11963	Removed									
NCU12103	Removed									

*ChIPseq Data from J Craig et al., Manuscript in preparation

Appendix B: RT-PCR primers

Locus	Gene	Direction	Primer ID	Sequence
NCU04173	<i>act</i>	Forward	OSC185	TGATCTTACCGACTACCT
NCU04173	<i>act</i>	Reverse	OSC186	CAGAGCTTCTCCTTGATG
NCU07705	<i>clr-1</i>	Forward	OSC189	ATGACGCCGAACCGAGTG
NCU07705	<i>clr-1</i>	Reverse	OSC190	CAACAACACCAGAATGCGG
NCU08042	<i>clr-2</i>	Forward	OSC183	TCCCGGCCATCAGACAGA
NCU08042	<i>clr-2</i>	Reverse	OSC184	ATCGGCACGGAAGGTTGTT
NCU08114	<i>cdt-2</i>	Forward	OSC175	CCACCGGTTACGATGGCATGTT
NCU08114	<i>cdt-2</i>	Reverse	OSC176	AGGGGGACAAAGGGGATGGAGC
NCU07340	<i>cbh-1</i>	Forward	OSC187	ATCTGGGAAGCGAACAAAG
NCU07340	<i>cbh-1</i>	Reverse	OSC188	TAGCGGTCGTCGGAATAG
AN6542	<i>actA</i>	Forward	YX52	TTCTTGGTCTGGAAAGCGG
AN6542	<i>actA</i>	Reverse	YX53	CTGCATACGGTCGGAGATAC
AN5808	<i>clrA</i>	Forward	OSC600	TGCAGACTGTCCTCCGACTA
AN5808	<i>clrA</i>	Reverse	OSC601	CAACATTCCCTCACCCACGA
AN3369	<i>clrB</i>	Forward	OSC590	TCTCAGACGACGGGTATGGT
AN3369	<i>clrB</i>	Reverse	OSC591	TAATCGCGGGACTGAATCGG
AN7320	<i>cdh</i>	Forward	YX76	CAGCCATGCTTCGCCCAGTCT
AN7320	<i>cdh</i>	Reverse	YX77	TGAGCTGCAGGACAGGTATCCG
AN5176	<i>cbhA</i>	Forward	OSC576	CATCAACGGATTGGGCAACG
AN5176	<i>cbhA</i>	Reverse	OSC577	ACATGGTCTGCTCGACACTG
AN0494	<i>cbhB</i>	Forward	YX96	CGGCACGGGATACTGCGACT
AN0494	<i>cbhB</i>	Reverse	YX97	GCAGCAGGAGCCGTGATTGC
AN5282	<i>cbhC</i>	Forward	OSC588	TTCCATCCCTCACTGGGACA
AN5282	<i>cbhC</i>	Reverse	OSC589	TGCGGCAACATCTAGCCAT
AN1273	<i>cbhD</i>	Forward	YX58	AGGGATACCAGCTCTACGCCAAC
AN1273	<i>cbhD</i>	Reverse	YX59	CCGTCGTATCCAGCCAATGGAAC

Appendix C: Plasmid Sequences

pSC04

LOCUS pSC04 10387 bp ds-DNA circular 25-NOV-2014
 DEFINITION Expression vector pMF272, complete sequence.
 COMMENT ApEinfo:methylated:1
 FEATURES
 source Location/Qualifiers
 join(1..4981,8186..10387)
 /organism="Expression vector pMF272"
 /mol_type="other DNA"
 /db_xref="taxon:274625"
 /label=source:Expression vector pMF272
 CDS 468..1328
 /gene="bla"
 /function="ampicillin resistance"
 /codon_start=1
 /product="beta-lactamase"
 /protein_id="AAT34981.1"
 /db_xref="GI:47524498"
 /translation="MSIQHFRVALIPFFAAFLPVFAHPETLVKVKDAEDQLGARVGY
 IELDLNSGKILESFRPEERFPMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
 YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTIGGPKELTAFLHNMGDHVTSL
 DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
 LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
 EIGASLIKHW"
 /label=beta-lactamase
 misc_feature complement(2374..4006)
 /note="3' flank region of Neurospora crassa his-3 gene"
 /label=3' flank region of Neurospora crassa his-3 gene
 promoter 4059..4980
 /note="Neurospora crassa ccg-1 (grg-1) promoter"
 /label=Neurospora crassa ccg-1 (grg-1) promoter
 misc_feature complement(8267..10368)
 /note="5' truncated histidine biosynthesis trifunctional
 protein; includes phosphoribosyl-AMP cyclohydrolase,
 phosphoribosyl-ATP cyclohydrolase and histidinol
 dehydrogenase"
 /label=5' truncated histidine biosynthesis trifunctional
 misc_feature 7538..8067
 /label=NCU08042 3-prime UTR
 rep_origin 1423..2105
 /label=ColE1 origin
 CDS 666..1325
 /label=AmpR
 misc_feature 4990..7537
 /label=NCU08042 ORF

ORIGIN

```

1 tatagtgtca cctaaatcgt atgtgtatga tacataaggt tatgtattaa ttgtagccgc
61 gttctaacga caatatgtcc atatggtgca ctctcagtag aatctgctct gatgccgcat
121 agttaagcca gccccgacac ccgccaacac ccgctgacgc gccctgacgg gcttgtctgc
181 tccccggcatc cgcttacaga caagctgtga ccgtctccgg gagctgcatg tgtcagagggt
241 ttccaccgtc atcaccgaaa cgcgcgagac gaaagggcct cgtgatacgc ctatttttat
301 aggttaatgt catgataata atggttttctt agacgtcagg tggcactttt cggggaaatg
361 tgcgcggaac ccctatttgt ttatttttctt aaatacattc aaatatgtat ccgctcatga
421 gacaataacc ctgataaatg cttcaataat attgaaaaag gaagagtatg agtattcaac
481 atttccgtgt cgcccttatt cccttttttg cggcattttg ccttctctgt ttgtctacc
541 cagaaacgct ggtgaaagta aaagatgctg aagatcagtt ggggtgcacga gtgggttaca
601 tcgaactgga tctcaacagc ggtaagatcc ttgagagttt tcgccccgaa gaacgttttc
661 caatgatgag cactttttaa gttctgtctat gtggcgcggt attatcccgt attgacgccg
721 ggcaagagca actcgtgcgc cgcatacact attctcagaa tgacttggtt gactactcac
781 cagtcacaga aaagcatctt acggatggca tgacagtaag agaattatgc agtgctgcca
841 taaccatgag tgataacact gcggccaact tacttctgac aacgatcgga ggaccgaagg
901 agctaaccgc ttttttgcac aacatggggg atcatgtaac tcgccttgat cgttgggaac
961 cggagctgaa tgaagccata ccaaagcagc agcgtgacac cagcatgcct gtagcaatgg
1021 caacaacggt gcgcaacta ttaactggcg aactacttac tctagcttcc cggaacaat
1081 taatagactg gatggaggcg gataaagttg caggaccact tctgcgctcg gcccttccgg
```

1141	ctggctggtt	tattgctgat	aaatctggag	ccggtgagcg	tgggtctcgc	ggtatcattg
1201	cagcactggg	gccagatggg	aagccctccc	gtatcgtagt	tatctacacg	acggggagtc
1261	aggcaactat	ggatgaacga	aatagacaga	tcgctgagat	aggtgcctca	ctgattaagc
1321	attggtaact	gtcagaccaa	gtttactcat	atatacttta	gattgattta	aaacttcatt
1381	tttaatttaa	aaggatctag	gtgaagatcc	tttttgataa	tctcatgacc	aaatccctt
1441	aacgtgagtt	ttcgttccac	tgagcgctag	accccgtaga	aaagatcaaa	ggtcttctt
1501	gagatccttt	ttttctgcgc	gtaatctgct	gcttgcaaac	aaaaaaacca	ccgtaccag
1561	cgggtggttg	tttgccggat	caagagctac	caactctttt	tccgaaggta	actggcttca
1621	gcagagcgca	gataccaaat	actgtccttc	tagtgtagcc	gtagtttagc	caccacttca
1681	agaactctgt	agcaccgcct	acatacctcg	ctctgctaat	cctgttacca	gtggctgctg
1741	ccagtggcga	taagtcgtgt	cttaccgggt	tggactcaag	acgatagtta	ccggataagg
1801	cgcagcggtc	gggctgaacg	gggggttcgt	gcacacagcc	cagcttgag	cgaacgacct
1861	acaccgaact	gagataccta	cagcgtgagc	attgagaaa	cgccacgctt	cccgaaggga
1921	gaaaggcgga	caggtatccg	gtaagcggca	gggtcggaac	aggagagcgc	acgagggagc
1981	ttccaggggg	aaacgcctgg	tatctttata	gtcctgtcgg	gtttcgccac	ctctgacttg
2041	agcgtcgatt	tttgtgatgc	tcgtcagggg	ggcggagcct	atggaaaaac	gccagcaacg
2101	cggccttttt	acggttcctg	gccttttgct	ggccttttgc	tcacatgttc	tttctcgctg
2161	tatcccctga	ttctgtggat	aaccgtatta	ccgcctttga	gtgagctgat	accgctcgcc
2221	gcagccgaac	gaccgagcgc	agcaggtcag	tgagcgagga	agcgggaagag	cgcccaatag
2281	gcaaacccgc	tctcccgcgc	cgttggccga	ttcattaatg	caggttaacc	tggttatctg
2341	aaattaatac	gactcactat	agggagaccg	gcagatccaa	tgcggatgga	ttcgctcact
2401	cggtgcccgc	gcggtgcgaa	gttctgccc	tggattgggg	gattgcacag	acggtggagt
2461	gcttgctaca	gctaggtacc	cagttgcaca	ccaaacgggtg	cccctgtgcc	cctgctctca
2521	ggtgtgtcct	gtgagcaact	gcgcccgcgc	gcggcaccgg	gaagctgcgc	gaacctcaat
2581	ggatgattgg	cgccagatcc	ccggcatcac	ctcagcacat	gggcagcccg	agaaggaatc
2641	gagggtagaa	gatattccag	ggtttcgct	agcacaactt	ccaagcgctt	ctaccgctac
2701	tccacacacc	cacgctgaat	ggcacaacca	ctgtaggcct	agcgggtcca	gctccccgta
2761	accacgcagc	cccagcgacc	gaggtcatca	tgacactatg	acgacaatct	ggcgggagac
2821	ttgggaaaaa	cgatgacggg	cacctacctt	gggaacaata	gagcaagtcc	agtcaacgcc
2881	ctaggggaag	gggaaaaaaa	aaaagagccc	attctcacaa	ggagcaaacac	tgtgctgcta
2941	tcatagtaca	ctacaagcaa	agtcgggatg	ttgctagaat	ccgtcggcta	cgtcggtaga
3001	tacctcccat	aggcagtaag	tctaggtatc	caatttatag	ttccgcgtat	ttcttcgtcg
3061	aacggggcac	agcaaaacct	aacagccgca	gtgcctcgca	accaacacca	agcagtaatc
3121	tcatccatcc	atacagatcc	tcccacaata	tctttacatg	cgcacttgcc	ctgtgttagc
3181	agtttaggtt	ccactaccaa	ggctgacacg	agcagcactg	acagcaaaact	gctgcttggt
3241	gccaagaaat	tggcagtttg	ctgagcacct	acatatcagt	acacatcaac	cctggtttaa
3301	ttgttagcct	ggtactgtta	accgtcatga	acgaggacgt	ggaggaagaa	ccacgctaac
3361	aattgaacca	gggttgatgt	atacctctct	taccatattt	gcttgcttcc	acgtccctat
3421	cctcaagcga	ttttgttctt	gactttttgt	atttcaacgc	aaaggccatt	caggtttctc
3481	atgttcaacc	ctttggatgg	cgatgcgctt	cgtcccagga	cctcgtgacc	tttgaaagac
3541	gagaaatata	tccaaaagca	acttctgctt	atggacaagc	aactgcatgg	cggcccacaa
3601	ctcatatctc	gcgtaatcat	ccaaaacgcta	ggcccagagta	ctctactgtt	ctcgagacac
3661	ctaccataat	atgtacccaa	tcagacatga	caacagtcaa	ccaaacatac	cctgatccgg
3721	ccaaaccctg	acatccctag	atgatggtta	ggcacacaaa	tggaaaacgg	gactcgttgg
3781	ttagggtggga	acgcttgttat	tcgagccagc	caaccgactc	tcctcgtccc	tctcagatcc
3841	cgttattgcc	gtttgaccca	gccgcccagc	cgccaaacag	ctgagggagc	caatggtcta
3901	atacaccgtc	caacggcccc	agtgtatgtt	caaacttgat	tgacagcgaa	cgaaacccct
3961	gaaacccggt	atagatatgt	atagatttga	cggggtagct	tggcccttaa	tacgactcac
4021	tataggcgga	attggagctc	caccgcgggtg	gcggccgcta	gaaggagcag	tccatctcgc
4081	tgaatcacga	gagaatcagc	tactttgaat	cgatggatgc	agctaccaga	agtcactcag
4141	ttcgttcaaa	gccacatcac	tgggcacttc	cattgggaca	ggcattgatc	ggacgagacc
4201	gacttctggc	cgctttcaac	agccacatta	tatccatgtc	acggctacgc	gtcggccttc
4261	ggtaacagaa	aagcacacag	acagcgattg	tgacatggat	tcgggcaaa	cattggtggg
4321	cgcaccaggt	catcctgagt	gtgcagtgcc	tgctattcag	atttcatcta	actcggggag
4381	aggggttcaa	aggggcgtga	cgtcacagac	aacgggtgaa	ggacgaagat	tgccctcact
4441	ctttgctagc	aatttgccctg	caaagaaggc	gcacatgaca	agcaaacaa	tgggaaacca
4501	ctattgaata	cccacaatgc	aaagctcgga	aggtacgtct	tgattgcagt	gtgtcgagtg
4561	tcaaaaaaga	agcaagtgtt	catgcaagcc	aaaattggca	cctcctccac	ttctccgagt
4621	gccccacccg	aacctccagg	cgagatggcc	ggaacatacc	atccgcgttg	ggattatgac
4681	gtatctcctt	cttcttcaca	tgattccatc	ccgttggttg	ttgtttgcga	gctgtgacgg
4741	gagatcgtag	atgccacttc	gggccaggca	ggcagtgcag	gcagccagga	acacaagctt
4801	ccaaacttgg	catctcgatt	gccgattaag	ggaaccaa	gcctatat	gactgtcctc
4861	ccacctcccc	aataaccattc	ttttcttctt	ccatcatcag	ccaacaaagc	aatcacatct
4921	gacttacttc	aaatcaacac	aacactcaaa	ccactttcac	aaccctccac	atcaaccaaa
4981	tCTAGAACCA	TGGCACCATC	AATGTCGATA	CCTACACAAG	GCGGCATGTT	TCACACATGT
5041	AAGTTTGATG	ACCTTTTCCA	ACCAAGGGGT	CCATTACTAA	TGAGCAAAATC	CAGTCCAAGG
5101	AGTCACACCA	CGAAAGCCGG	CTGTTGACTC	TCGAGATAGT	GTTAAGTCCA	ATGGTAGTGG
5161	GGCAGCCCAAG	AGGATAACTA	CTCCTCATGC	TTGTGCTGAG	TGTAAGAGAC	GAAAGATGTA
5221	AGTCAATGCG	TGTCGATCAG	CCTCACCATC	AATCTAACTT	GAGTGCCCTA	GTCGGTGTGA
5281	TGGTCAACAA	CCATGTGGCC	AATGTCTCTC	AAGCCGAGCA	CCGAAACGTT	GTTTCTATGA
5341	CAAGCACCGC	CAAAGGGTGA	TCCCATCGCG	CAAGACACTC	GAAGCCCTCT	CCAGTCCCT
5401	CGAAGAATGC	CGATCGATTT	TGAAGCGACT	GTATCCTCAC	CAAGAAGTCC	ACGCCTTGCT
5461	TCCCTTGTC	AGGCAAGAAC	TCCTCAACCT	CCTGGATCGA	CCCGTCCCGG	ACAGCACAA

5521 AGCGAACGGA CTACCGTCAC CTCCTCTAAA CACAACACCG ATGGCCGAGT CAGAGACACC
5581 AACCAAGTCT GAAAACATTT TGGAAACAGAT TCCGACTCGG GATACAGAAT GGGACGAGGA
5641 ACCCGGGGAG AGGGACCATA TCCCTGTAGA GGCGGACGAT ATCAACGCGC TGTCACTCTC
5701 GGTGGACAGA CAAGCTTCTT ATCTTGGAGC CTCATCCATC AAGGCTGCCC TTATGGTCAT
5761 GCTCAAGGTT CAACCAGGCT TGAGGAGCTC CTTGGCCGCC CCGCTCAACA GTATCGAGAT
5821 TTCTCATAAC TTCCCAGCCA TCAGACAGAA ATCATCGAGT CAGAAAGATC CCCAGCGGAT
5881 CCCATGGTCA TGGAAAGGAC AGACACTTAT CGATGCCTAC TTCAAGCGAG TCCACGTCTT
5941 CATCCCAATG CTCGACGAAA CAACCTTCCG TGCCGATTAC CTTGAAGGTC AACGATTTGA
6001 CGCACCCTGG TTAGCCTTGT TGAACATGGT ATTCCGCATG GGTAGCATTG TGCAATGAA
6061 GTCGGACGAT TACAATCACG TCAACTATTA CAACCGCGCC ATGGAGCATC TTCCCATGGA
6121 TTCTTTTGGC AGCAGCCATA TCGAGACTGT CCAAGCATTG GCACTCATCG GTGGTTACTA
6181 TCTTCACTAT ATTAACCGAC CCAACATGGC CAACGCTGTT TTGGGTGCCG CCATCCGCAT
6241 GGCCAGCGCT CTCGGTCTCC ACCGTGAGAG TCTTGCCCAA GGAGGAAGTG ATATGGTCGC
6301 TGCCGAGACC AGGAGACGCA CATGGTGGGC TCTTTCTGCT CTTGACACGT GGGCCACGAC
6361 AACCATGGGA CGTCCCTCTT TTGGCCGTTG GGGACCCGCC ATCAACATTC CCGCCGCCGA
6421 GTTCGGTGTG AATGCCGCTC GGGACTCGTC TCAACACGCG GGTATTCTCC CCATGATTGA
6481 AAACGTCAAG TTCTGCAAAA TCGCCACTCA GATCCAAGAC ATGCTCGCCA TTTCTCCTTT
6541 GCTGCGTACC GAGGACCGCT GCAATCTGGA CGGACAGCTC GTTGTCTGGT ATGAAAAGCT
6601 ACCTTGGCTC CTGCGTACGA CCGACCCATG CGCTGAGCCG CTCTACATGG CCCGCTGCAT
6661 CATGAAGTGG CGGTATCAGA ACCTCCGTAT GCTTCTTCAT CGCCCAGTCC TCCTCTCCAT
6721 GGCTCATCA GGTCTTAACC CACATACTCA GGCTGTGAC GCGGATCTCG CCGCCATTGA
7781 GACGTGCCGA GAGCTTGTCT GGGCAACCAT TGAGGACATT GGTCTGAAT GGACACGTAA
6841 CCAGATGTCC GGCTGGAACG CAGTGTGGTT TCTCTACCAG GCTGCCATGG TTCTCTTGT
6901 CTCGCTCTTT TGGCAATGGG GCAACCTCG TGTCCTTGAG TGGCTCAAGC AGATTGAGGC
6961 GGTATTGGAG CTCCTTGAGG CCATGGAAGA GTGGTCACTC GCTGCTCGCA GGAGCCGGGA
7021 GGTGGTTCTT CGCATGTACG AGGCCAGCAG AGTCATCCAG GCACAGGGTG CCGCAGCGCA
7081 CCAACTACAG CAGAGGGGAT CGCAGAGCCC TCATTCTCTC AGCAGCACTA CTGCGTCTTT
7141 GGCAACATG CACATCAACA ACGACTTGCT CATGGGGTCG TCACCGGGTT CCGCGGAGCT
7201 CTACATGACT CCTATCGACC TGGAGCCCGT CGAAGGCTTG GGCATGACGG GTGTGCTGGA
7261 CCACGGCGGG ATGTGGGATC TCGACGGCAT GCTTTGGGGT CCCAGTCCAA GCCCGTCGGT
7321 TCACCATGGT CAACACTCGA CTCACCCGGA TGATGCGGTG CCCACGGTGG CGGAGTACGC
7381 TGACGCGTTC CCGACGGCTG ATGTAGTAGA TGGGTTCCTG CAACATCACT CGGCTATGGA
7441 GTTTGGGAAC ATGATGGGTC ACCACCACCA CCACCACCAT GCAGGTGCAC ATGCTGGCCA
7501 GGTCTAGTTT GGAGGTATGG ATGGGTACAC CTACTGAAGG CCCCTCGCAG TGCTCCAATT
7561 TTCTAGTTTA CTACTCATTT CTCTCTTTCT TTCTTTCTTT CTTCCTCTTT TCACACAAAA
7621 AATAGTTGGG CTGGAAGCCC GGGGTATATG ACAGTTGGCG CGACGCAGGC TTGAAGGACC
7681 GAAATACCA CACATACGAT GGCTGGATGC GGTGGTGCG GATTGGCTT ACGAATGGTT
7741 GCTTTTTGCG GGGTACATCA TATACGGAGA AGGATAAGGG TCTCATGGTA AAAAGGAAAA
7801 GGAGCTTTTG TTTTTTGGCG GTTGATTTCT CAGGGAATGT TGGTTTGATT TTTGGCTGAT
7861 TTTGGATGGT TACAACGGCG CCTTCGCGGG TACACTCTCC CTCTTGTTAG GTGTAGTGTC
7921 CCACTCCCT TCCATTCAAC AGCCACCGGC CTTACCTAGG CCGTTGTCAT GTGGGCCTAG
7981 TTCAGGGCAC CACCATGTTC ATTGTATTAT TATTATAGCC CTTTTTACAT TTCATTTCCA
8041 CTTTAGCTTC TTAACACAGT CATCAACACT TCTTGTGCCT ACACTGCGGC ATCTCTTTTC
8101 GCCATCGCTT CGAGTACACT AGTTCTGATG TCACATGCGG TTATGTTTTT ATCCCGTTGT
8161 AACACTTTGT GCCTCTTGGT CCATAgatcc ccacatcaac ccactacatc cttttgtctc
8221 tgtttttttt tttttttttt gctacttcga aaagaatttc cgtttagtgt ctcttgctca
8281 tbtgctcaag acgggatgctg accgcccctc tbtgagcctc gagctcctca acctagcca
8341 actgcatgac agcctggccg acgttttttg gaccctcggc agtcaagtgt gagctggtaa
8401 tbtgcttgac gaacgagcgc aaattgacgc cagagtactg cttgccaaag ccatagtag
8461 ctatgtgtct tgacgttagc agaagaatct gtgaagcggg gctatttatg atatgtaact
8521 tacgcagcga gtgggttaaca ccagcagagt aatcgccaac ggactcagga gtccaggcgc
8581 caatgaagac actaccggcg ttcatgacaa gatcgacage cttctcggcc tccttgatct
8641 ggaggatcaa gtgctcagga gcgtacttgt tgctgagctc catggcctcc tcgacgtctc
8701 tgacctgcac ggtgatcgag tgggcatgg agccacggac aatctggacg cgaggaaagt
8761 ccgtagcctg acgggtaacc tcgtcctcaa tagcctgaag atgctcctcg tcgaggtcaa
8821 tagcgatcag gatgacctga ctgtcaacgc cgtgctcagc ctgggacagg agatccgagg
8881 caacgaacgc ggggttgcg tccttgctcag cgatgaccag cacctcggac gggccagcgg
8941 gcatgtcaat accaacggca gcgttggtgt cgttgctgac gaacatcttg gcagcagtga
9001 cgaactggtt accggggccg agaattctgt cgaccttggt gatgctctcg gtgccgtagg
9061 ccattggcagc tacggcctgg gcaccgcccg caagcacgat ggactcggcc ccaaccttgt
9121 gagcgacgtg gacaatctcg ggagtgatgg ttccgtcggc gcggggagga gaggcgaaca
9181 caatcttgtt gcagccggcg accatggcgg gaacaccag cataagggca gtgctgggga
9241 gaacggcggt accgcccggg atgtagcagc cgacggcctc gatgggacga gagaaacggc
9301 tgcagacaac accgggcatg gtctcgacct ggaggggcct ctccctcctc tggcgcgctg
9361 ggaacttgcg gatgttctcg aaggacacgt cgatggcagc aatggtctcc tcaggagact
9421 gcataagctc cttggggaag ggcgccttca ggacggggct agtaagagag gtacgctctc
9481 cgaacttggt agtgtagcag agaacagcct tgtcgccgtt cttgggacg tcctcgatga
9541 tggggacaat gatctttag atggcatcg acgacttttg cgcaggacgc ttgagagcag
9601 catcgagctc ctcggttag agcttgagg cgtcgaaacg tctcatggtg atcttctcag
9661 ggggtggtctc ctgggcccgg tccttggtga caggggccga agtggcagcg agagcggacg
9721 ccgagggctt gatgcccctc ttctcagccc acttaccctt agcatctcca gtccctgcgt
9781 tgaccttcca gctcttggcg tcaaggctcc tttcgatatc ggcaagagta acgccggcag
9841 caacggccct ggtaagagca aagtagaaga gatcggcagc ctcaaaggcg atttctgagg

```

9901 gggctctgagc ggtgcagagc tcctcagcct cctccatgat cttggcccgg actagcttct
9961 catcggagaa gagacgggca gtgtaggagc cctcgggggc agactgtttc ctcgaaatca
10021 aagtctgctc gagcttgga aggcctttga gctgaccaa gcagccggac tgatcgaggt
10081 ggcagaaacg acccttctgc ttcacgacaa acttgagagc atcggttatcg cagtcaagcg
10141 agatgcggac gagctcctga gtgtctccgg aagtagcacc cttgtaccag agaccgcgct
10201 tccggctctg atagacacca gtctgtgtcc tgagggcctc gttcacactc tcggcactgc
10261 tgtagaccag acccagagca gtgtcgtgct catcaacgac aacggtgggg agaagaccat
10321 cgggacggtc agacttcag acgctcgaga ggatggtgga gatggcaagc ttcagctgct
10381 cgagttc
//

```

pSC65

```

LOCUS          pSC65          10231 bp ds-DNA   circular   25-NOV-2014
COMMENT        ApEinfo:methylated:1
FEATURES             Location/Qualifiers
     source          join(1..4986,7527..10231)
                     /organism="Expression vector pMF272"
                     /mol_type="other DNA"
                     /db_xref="taxon:274625"
                     /label=source:Expression vector pMF272
     CDS              468..1328
                     /gene="bla"
                     /function="ampicillin resistance"
                     /codon_start=1
                     /product="beta-lactamase"
                     /protein_id="AAT34981.1"
                     /db_xref="GI:47524498"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTIGGPKELTAFLHNMGDHVTSL
                     DRWEPELNEAIPNDERDTTMPVAMATLRKLLTGELLTLASRQQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
                     /label=beta-lactamase
     misc_feature     complement(2374..4006)
                     /note="3' flank region of Neurospora crassa his-3 gene"
                     /label=3' flank region of Neurospora crassa his-3 gene
     promoter         4059..4980
                     /note="Neurospora crassa ccg-1 (grg-1) promoter"
                     /label=Neurospora crassa ccg-1 (grg-1) promoter
     exon             4991..5392
                     /label=NCU07705 Exon
     misc_feature     complement(8111..10212)
                     /note="5' truncated histidine biosynthesis trifunctional
                     protein; includes phosphoribosyl-AMP cyclohydrolase,
                     phosphoribosyl-ATP cyclohydrolase and histidinol
                     dehydrogenase"
                     /label=5' truncated histidine biosynthesis trifunctional
     exon             5464..6932
                     /label=NCU07705 Exon(1)
     exon             7016..7274
                     /label=NCU07705 Exon(2)
     terminator       7375..7526
                     /label=3-prime region from NCU00762
     misc_feature     7275..7295
                     /label=linker
     misc_binding     7296..7346
                     /label=V5 Tag
     misc_feature     7347..7364
                     /label=His Tag

```

```

ORIGIN
      1 tatagtgtca cctaaatcgt atgtgtatga tacataaggt tatgtattaa ttgtagccgc
     61 gttctaacga caatatgtcc atatggtgca ctctcagtag aatctgctct gatgccgcac
    121 agttaagcca gccccgacac ccgccaacac ccgctgacgc gccctgacgg gcttgtctgc
    181 tcccgcatc cgcttacaga caagctgtga ccgtctccgg gagctgcatg tgtagaggt
    241 tttcaccgtc atcaccgaaa cgcgcgagac gaaagggcct cgtgatacgc ctatttttat
    301 aggttaatgt catgataata atggtttctt agacgtcagg tggcactttt cggggaaatg
    361 tgcgcggaac ccctatttgt ttatttttct aaatacattc aaatatgtat ccgctcatga
    421 gacaataacc ctgataaatg cttcaataat attgaaaaag gaagagtatg agtattcaac
    481 atttccgtgt cgcccttatt cccttttttg cggcattttg ccttcctgtt ttgtctcacc
    541 cagaaacgct ggtgaaaagta aaagatgctg aagatcagtt ggggtgcacga gtgggttaca
    601 tcgaactgga tctcaacagc ggtaagatcc ttgagagttt tcgccccgaa gaacggtttt

```

661 caatgatgag cactttttaa gttctgctat gtggcgcggt attatcccgt attgacgccc
721 ggcaagagca actcggtcgc cgcatacact attctcagaa tgacttggtt gactactcac
781 cagtcacaga aaagcatctt acggatggca tgacagtaag agaattatgc agtgctgcca
841 taaccatgag tgataacact cgggccaact tacttctgac aacgatcgga ggaccgaagg
901 agctaaccgc ttttttgac aacatggggg atcatgtaac tcgccttgat cgttgggaac
961 cggagctgaa tgaagccata ccaaacgacg agcgtgacac cacgatgcct gtagcaatgg
1021 caacaacggt gcgcaaaacta ttaactggcg aactacttac tctagcttcc cggaacaat
1081 taatagactg gatggaggcg gataaagttg caggaccact tctgcgctcg gcccttccgg
1141 ctggctggtt tattgctgat aaatctggag ccggtgagcg tgggtctcgc ggtatcattg
1201 cagcactggg gccagatggt aagccctccc gtatcgtagt tatctacacg acggggagtc
1261 aggcaactat ggatgaacga aatagacaga tcgctgagat aggtgcctca ctgattaagc
1321 attggtaact gtcagaccaa gtttactcat atatacttta gattgattta aaacttcatt
1381 ttttaatttaa aaggatctag gtgaagatcc tttttgataa tctcatgacc aaaatccctt
1441 aacgtgagtt ttcgttccac tgagcgtcag accccgtaga aaagatcaaa ggatcttctt
1501 cagtgctctt ttttctgcgc gtaatctgct gcttgcaaac aaaaaaacca ccgctacag
1561 cgggtggtttg tttgccggat caagagctac caactctttt tccgaaggta actggcttca
1621 gcagagcgca gataccaaat actgtccttc tagtgtagcc gtagtttagc caccacttca
1681 agaactctgt agcaccgcct acatacctcg ctctgctaact cctgttacca gtggtgctc
1741 ccagtgccga taagtctgtt ctaccgggtg tggactcaag acgatagtta ccggataagg
1801 ccagcggttc gggctgaacg ggggggttcgt gcacacagcc cagcttgag cgaacgacct
1861 acaccgaact gagataccta cagcgtgagc attgagaaag cgccacgctt ccgaaggga
1921 gaaagcgcca caggatctcg gtaagcgcca ggtcggaaac aggagagcgc acgaggagc
1981 ttccaggggg aaacgcctgg tatctttata gtctgtcgg gtttcgccac ctctgacttg
2041 agcgtcgatt tttgtgatgc tcgtcagggg ggccggagcct atggaaaaac gccagcaacg
2101 cggccttttt acggttcctg gccttttctg ggctttttgc tcacatgttc tttctgcgt
2161 taatccctga tctgtggat aaccgtatta ccgcctttga gtgactgat accgctgcc
2221 gcagccgaac gaccgagcgc agcagatcag tgagcgagga agcgaagag cgcaccaatc
2281 gcaaacgcc tctccccgcg cgttgccga ttcattaatg caggttaacc tggcttatcg
2341 aaattaatac gactcactat agggagaccg gcagatccaa tgcggatgga ttcgctcact
2401 cggtgccgcg gcggtgcgaa gttcctgccc tggattgggg gattgcacag acggtggagt
2461 gcttgctaca gctaggtacc cagttgcaca ccaaacggtg cccctgtgcc cctgcgtcta
2521 ggtgtgtcct gtgagcaact gcgcccgcgc gcggcaccgc gaagctgcgg gaacctcaat
2581 ggatgattgg cgccagatcc ccggcatcac ctcagcacat gggcagccc agaaggaatc
2641 gagggtagaa gatattccag ggtttccgct agcacaactt ccaagcgct ctaccgtac
2701 tccacacacc cagctgaat ggcacaacca ctgtaggcct agcgggtcca gctccccgta
2761 accacgcagc cccagcgacc gaggtcatca tgacactatg acgacaatct ggcgggagac
2821 ttgggaaaaa cgtatgacgg ggaacacata ggaacaataa gagcaagtcc agtcaacgoc
2881 ctagggaagg gggaaaaaaa aaaagagccc attctcaciaa ggagcaaac tgtgctgcta
2941 tcatagtaca ctacaagcaa agtcgggatg ttgctagaat ccgtcggcta cgtcggtaga
3001 tacctcccat aggcagtaag tctaggtatc caatttatag ttcgcgctat tctctcgtc
3061 aacggggcac agcaaacgct acacgcgca gtgcctcgca accaacacca agcagtaatc
3121 tcatccatcc atacagatcc tcccacaata tctttacatg cgcacttgcg ctgtgttagc
3181 agtttaggtt ccactaccaa ggctgacacg agcagcactg acagcaaaact gctgcttggg
3241 gccaagaaat tggcagtttg ctgagcacct acatatcagt acacatcaac cctgggttaa
3301 ttgttagcct ggtactgtta accgtcatga acgaggaagt ggaggaagaa ccacgctaac
3361 aattgaacca ggggtgatgt atacctctct taccatattt gcttgcttcc acgtccctat
3421 cctcaagcga ttttgttctt gactttttgt atttcaacgc aaaggccatt cagggttctc
3481 atgttcaacc ctttggatgg cgtgcccgtt cgtcccagga cctcgtgacc tttgaaagac
3541 gagaaatata tcccaagca acttctgctt atggacaagc aactgcatgg cggcccacaa
3601 ctcatatctc gcgtaatcat ccaaacgcta ggcccagta ctctactgtt ctcgagacac
3661 ctaccataat atgtacccaa tcagacatga caacagtcaa ccaaacatac cctgatccgg
3721 tcaaacctg acatccctag atgatggtta ggcaacaaa tggaaaacgg gactcgttgg
3781 ttagggtgga acgcttgtat tgcagccagc caaccgactc tcctcgtccc tctcagtc
3841 cgttattgcc gtttgacca gccgcccagc cgccaaacag ctgaggagag caatggtcta
3901 atacaccgtc caacggcccc agtgatgttt caaacttgat tgacagcgaa cgaaccctc
3961 gaaacccggt atagatatgt atagatttga cggggtagct tggcccttaa tacgactcac
4021 tataggcgca attggagctc caccgcggtg gcggccgcta gaaggagcag tccatctcgc
4081 tgaatcacga gagaatcagc tactttgaat cgatggatgc agctaccaga agtcaactcag
4141 ttcgttcaaa gccacatcac tgggcacttc cattgggaca ggcattgatc ggacgagacc
4201 gacttctggc cgctttcaac agccacatta tatccatgtc acggctacgc gtcggccttc
4261 ggtaacagaa aagcacacag acagcgattg tgacatggat tcgggcaaag cattgggtgg
4321 cgaccagggt catcctgagt gtgcagtgcc tgctattcag atttcatcta actgcccagg
4381 aggggttcaa aggggctgta cgtcacagac aacgggtgaa ggacgaagat tgcctcactt
4441 ctattgctagc aatttgcctg caaagaaggc gcacatgaca agcaaaacac tgggaaacca
4501 ctattgaata cccacaatgc aaagctcgga aggtacgtct tgattgcagt gtgtcgagt
4561 tcaaaaaaga agcaagtgtt catgcaagcc aaaattggca cctcctccac ttctccgagt
4621 gccccaccgg aacctccagg cgagatggcc ggaacatacc atccgcgttg ggattatgac
4681 gttatctcct cttcttcaca tgattccatc cgttggttgc ttgtttgcga gctgtgacgg
4741 gagatcgtag atgccacttc gggccaggca ggcagtgcag gcagccagga acacaagctt
4801 ccaacttggg catctcgatt gccgattaag ggaaccaaact gcctatataa gactgtcctc
4861 ccacctcccc aataccattc ttttcttctt ccatcatcag ccaacaaagc aatcacactc
4921 tcaactcttc aaatcaacac aacactcaaa ccactttcac aacccctcac atcaaccaaa
4981 tctagaCGAC ATGTCGCCGT CTTCCAACAG CGACTCTCCC GCCTCGGTAG TGGCAGCAT

5041 GACCATGGCC GCGCAGGCTG CTCCTAACGT TGTAAGGCAA TATCTCTGCG CGAATCGAGA
5101 CTTCACAATC GAAATTCCTT CGCCCCAAGT TGCACCCACT ACCAATGGGG TCAACACCGT
5161 CAAGTCACCC TCCACTCTCA AGTCGTCCCG TGCTCCCAAC TTCAGCCGCG AAGGTATTCT
5221 TGGATCGGCC CAGAAAGCCC GAAATCTCTC GCAGTCGTCT GAAAACCGGC CTGACCATGT
5281 CAATGTCTCT AATGGCTTTC ACAAGGCTCC TAGTGAGGAG AGCACCAATC CGCTGAAAAG
5341 AAGAAATACG GATGCTGGCG TCGACTACCC TAGAAGAAGG GCAACCATTG CGGTATGTTT
5401 GGGCTCTGGC GCAGTGGAAG AATCTGGTGT GGTGAATCTA CACGCTGACA TTTCGTTGAA
5461 CAGTGCAGAG TATGCCGCTC GAGGAAGTCG CGGTGTGACG GCACAAAACC AAAATGCAAG
5521 CTGTGCACAG AGTTGGGTGC CGAGTGCATA TATCGCGAGC CTGGTATCAA GCTGGATGCC
5581 GCGCACAAGC TCATCTCGA ACGCTCAAC AGGATAGAAA GTCTGCTGCA ATTAACATG
5641 GGCTTAACCG GTCATACCCA TGGCCTAGGA TTATCCATA ACTCCCCGGG GATGAGCAAC
5701 GGGACAGGGT TGAGTGGTGA CAACCTGATG ATGCAAAATC CCGACGCCGC CGGCTTTGTC
5761 TCGGTGATCC CCAGCGGAGG CTTGGGCACA TGGTCCGCTA CCGCTACCAA CAGCTTGCTA
5821 ACAATGCCCA AGGTCCACAC CAATGCCGCC CTTTATTTGT TGCAATGGCC CATGATTCCG
5881 GATCTCGTCT CACGACCTTA CGACCCACAA ATTCTCTTCC AACTCGAAT GGCAGGGAG
5941 CCACTCCACT CGCTCACAAA GACCCCTTGT GTGGACCTCT CCAACACTTC CGCTTACATC
6001 GAAGCCTATT TCGATCGTGT CAACGTTTGG TACGCTGTG TCAACCTTAA TACATGGAGG
6061 AGCCACTATC GTATCGCCTT CTCGAACGGT TTCAGGGAAG GACCAGAAAG TTGTATTGTG
6121 CTCTCTGCTT TGGCTCTCGG TCAGGCAAGT TTACGGGGGA GCATATCCAG GATCGTGCCA
6181 CATGAGGATC CCCCAGGCTT CCAATATTTC ACGGCCGCAT GGGCTTGTG ACCGGGCATG
6241 ATGACCTCCA ACAGCGTCTT CGCCGCTCAA TGCCACCTTT TAGCTGCCGC TTACCTATTC
6301 TACCTCGTCC GACCCCTGGA AGCATGGAAC CTTCTCTGCA CAACAAGTAC CAAGTTGCGA
6361 CTCTCTCTTA TGACGCCGAA CCGAGTGCCT GTCGACCAA GGGAACTGAG TGAAAGGATC
6421 TACTGGAATG CTCTGCTGTT CGAGAGCGAC CTCTCGCCG AGCTGGACCT TCCGCTTCT
6481 GGTGTTGTTG CATTGAAGA AAATGTTGGT CTGCCATGTG GCTTCGAGGG GGATGAGCAA
6541 TTGCGCTTGT GTCGCGATGA TTTGTGGTAT TTCTCGCCG AAATAGCACT TCGGCGACTA
6601 CTCAACCGCG TCAGCCAGCT CATTACTTCC AAGGACTCCT TGGCTCTAC AACGAGTCTC
6661 GAACCCATCG TGACCGAGTT GGATTTCCTA CTAACGCAGT GGTACGAGAG TCTACCTCTA
6721 CCGCTGCAAT TTCCGTTTAC TCGGACAATG CTCCAGATC CTGTCCAGAC GGTGTTGCGG
6781 TTGCGCTTCT TTGCTTGCCG GACCATTTAT TACCGGCCAT ACATTCTCGC GGTCTCGAC
6841 AATGAACAGG CAATACTGGA CCCCTCAGTG CGTGACGCGT GCACAAAGTG TCTGGAGGCT
6901 TCATATCAGC AATTGGAGCA TATCAGTGCA CAGTAAGTCT TCGCCCGTTA ATATTCGGGT
6961 TTTGATTGGC TGCTTTTTC TTTTATTATC GTCTCAATCT AACAAAAGGC CGCAGTCACG
7021 CCGGACATCG GCCGTACCTT TGGCAAGGGG CACTCTCGAT TGTATCGCAG AACTTCTGTA
7081 TTTATGGGTG CACTATGTCA CCGTCACTAA ACAGTATTCT GGTGAGCTTG GTCTCACACC
7141 GCGACGTTAT CGACAGATC ATCAACGATG TCGTATGGA GATCGAGCGT TATTCGGTAC
7201 TCTCACCAG TCTGAGCTC TCGGCGGAGA TTATCAAGGA AGCGGAGGTT AGCGGACGAA
7261 CGTTCCTGAG TGGTGGCGGA GCGGAGGCG GAGGCTTAAT TAACGGTAAG CCTATCCCTA
7321 ACCTCTCTCT CGGTCTCGAT TCTACGCACC ACCACCACCA CCActaggaa ttcgTAAGGG
7381 AAAGTGACAG AGTTGCTGTA GGGAAAGCGA TCAATCGCCC AACGAGAAAC AGCCTGTTTT
7441 TCTCGTTGAAA AGCATTGATC CCCAAATCGT TTAACTTTTT CTTGTACATA CATACTACCT
7501 TTGAACATAT TAAATCGCAT TGATTGgggc ccggtaccga gcttttgttc cctttagtga
7561 gggttaatta gggcggtggc ggctaattgg gtctgaatgc taaaagacac catttccac
7621 actccctctg atttttttta tggttttttt ttatcttcat ttttcgccgt cttttattcg
7681 gttgttttgg aggttcacat gcaactgca agactgcaag ttttttgaga tggcgcttc
7741 atgcaatgca gcatcactgc agggacattg tgacatccac gtttcgatcg tctgtgcttc
7801 ccccggtcat gtacgcatga tctgaaaaaa aaaaaaaaaa aaaagggttg tgtaataatg
7861 aaaataaaca cccagggagc agcacctcat ctgctggaat agggagaaaa actcaagggc
7921 cgttttgggtc tccacactta ccaacagatt catgtactga tcttcagcct tcacagaca
7981 gtacaatgtc caaaactcgt tatcattact ctatgtgccc aaaatgacct agatatacat
8041 caaccacta catccttttg ttcttgtttt tttttttttt ttttgctact tcgaaaagaa
8101 tttccgttta gttgctcttg ctcatgtgct caagacggat gctgaccgcc cttctgtgag
8161 cctcgagctc ctcaacctta gccaactgca tgacagcctg gccgacgttt ttgagacct
8221 cggcagtcga gttggagctg gtaatgtgct tgacgaacga ggcgaaattg acgccagagt
8281 actgcttgcc aaagccatag gtagctatgt gtcttgacgt tagcagaaga atctgtgaag
8341 cggggtctatt tatgatattg aacttacgca gcgagtgggt aacaccagca gagnaattgc
8401 caacggactc aggagtcag gcgccaatga agacactacc ggcgttcatg acaagatcga
8461 cagccttctc ggcctccttg atctggagga tcaagtgtc aggagcgtac ttgttgctga
8521 gctccatgac ctcctcgagc gtcttgacct gcacggtgat cgagtggcg atggagccac
8581 gacgaactcg gacgcgagga agctccgtag cctgacgggt aacctcgtcc tcaatagcct
8641 gaagatgctc ctcgtcgagg tcaatagcga tcaggatgac ctgactgtca acgctgtgct
8701 cagcctggga caggagatcc gaggaacga acgcggggtt ggcgtccttg tcagcgatga
8761 ccagcacctc ggacgggcca gcgggcatgt caataccaac gccagcgttg gtgtcgttgc
8821 tgacgaacat cttggcagca gtgacgaact ggttaccggy ggcgagaatc ttgtcgacct
8881 tggatgagct ctcggtgccc taggccatgg cagctacggc ctgggacccg ccggcaagca
8941 cagtgagctc gggcccaacc ttgtgagcga cgtggacaat ctcgggagtg atggttcctg
9001 cggcgcgggg aggagagggc aacacaatct tgttgagcgc ggcgacctg gcgggaacac
9061 gcagcataag ggcagtgtg gggagaacgg cgttaccgcc ggggatgtag cagccgacgg
9121 cctcgatggg acgagagaaa cggctgcaga caacaccggg catggtctcg acctggaggg
9181 gcttctctc cttctggggc gcgtggaact tgcggtatgt ctcgaaggac acgtcgatgg
9241 cagcaatggt ctctcaggg agctgcataa gctccttggg gaagggcgcc tcagggacgg
9301 ggcctagtaag agaggtagcc ttctcgaact tgtgagtgtg cgacagaaca gcctgtctgc
9361 cgttcttgcg gacgtcctcg atgatggga caatgatctt gtagatggca tcggacgact

```

9421 ttgcgccagg acgcttgaga gcagcatcga gctcctcggt agagaccttg gaggcgtcga
9481 aacgtctcat ggtgatcttc tcaggggttg tctcctgggc ggcctccttg gtgacagggg
9541 ccgaagtggc agcgagagcg gacgccgcag gcttgatgcc ctccttctca gccacttac
9601 ccttagcatc tccagtcctg cgcttgacct tccagctctt ggcgtcaagg ctcctttcga
9661 tatcggaag agtaacgccg gcagcaacgg ccctggtaag agcaaagtag aagagatcgg
9721 cagcctcaaa ggcgatttcc tgggggggtct gagcgggtgca gagctcctca gcctcctcca
9781 tgatcttggc ccggactagc ttctcatcgg agaagagacg ggcagtgtag gagccctcgg
9841 gggcagactg tttcctcgaa atcaaagtct gctcgagctt ggggaaggcct ttgagctgac
9901 caaagcagcc ggactgatcg aggtggcaga aacgacctt ctgcttcacg acaaacttga
9961 gagcatcggt atcgagtcga agcgagatgc ggacgagctc ctgagtgtct ccggaagtag
10021 cacccttgta ccagagaccg cgcttcgggc tctgatagac accagtctgt gtccagaggg
10081 cctcggtcac actctcggca ctgctgtaga ccagaccag agcagtgtcg tgctcatcaa
10141 cgacaacggt ggggagaaga ccatcgggac ggtcagactt ccagacgctc gagaggatgg
10201 tggagatggc aagcttcagc tgctcgagtt c

```

//

pSC107

```

LOCUS          pSC107          10345 bp ds-DNA   circular   25-NOV-2014
COMMENT       ApEinfo:methylated:1
FEATURES             Location/Qualifiers
     source          join(1..4050,9773..10345)
                     /organism="Expression vector pMF272"
                     /mol_type="other DNA"
                     /db_xref="taxon:274625"
                     /label=source:Expression vector pMF272
     misc_feature    7596..8035
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
                     width 5 offset 0
                     /label=TrpC Promoter
     misc_feature    join(8036..8370,8372..8409)
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
                     width 5 offset 0
                     /label=Phleomycin
     CDS             468..1328
                     /gene="bla"
                     /function="ampicillin resistance"
                     /codon_start=1
                     /product="beta-lactamase"
                     /protein_id="AAT34981.1"
                     /db_xref="GI:47524498"
                     /translation="MSIQHFRVALIPFFAAFLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTITGGPKELTAPLHNMGDHSVTRL
                     DRWEPELNEAIPNDERDTMPVAMATTLRKLLTGELTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
                     /label=beta-lactamase
     misc_feature    join(5357..6032,6034..6382)
                     /label=HPH
     misc_feature    join(8454..8461,8463..8746)
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
                     width 5 offset 0
                     /label=Sc CYC1 Terminator
     misc_feature    join(7057..7225,7228..7232,7234..8033)
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1}} {} 0}
                     width 5 offset 0
                     /label=A.nidulans gpdA promoter
     promoter        4059..5356
                     /label=NCU01050 Promoter
     3'UTR           6383..7056
                     /label=NCU01050 3-prime region
     misc_feature    8747..9772
                     /label=NCU01050 downstream 1kb
ORIGIN
1  tatagtgtca cctaaatcgt atgtgtatga tacataaggt tatgtattaa ttgtagccgc
61  gttctaacga caatatgtcc atatggtgca ctctcagtag aatctgctct gatgccgcac
121 agttaagcca gccccgacac ccgccaacac ccgtgacgc gccctgacgg gcttgctcgc
181 tcccgcatc cgcttacaga caagctgtga ccgtctccgg gagctgcatg tgcagaggt
241 tttcaccgtc atcaccgaaa cgcgcgagac gaaagggcct cgtgatacgc ctatttttat
301 aggttaatgt catgataata atggtttctt agacgtcagc tggcactttt cggggaaatg

```

361 tgcgcggaac ccctatttgt ttatttttct aaatacattc aaatatgtat ccgctcatga
421 gacaataaac ctgataaatg cttcaataat attgaaaaag gaagagtatg agtattcaac
481 atttccgtgt cccccttatt cccctttttt cggcattttg ccttctctgt tttgctcacc
541 cagaaacgct ggtgaaagta aaagatgctg aagatcagtt ggggtgcacga gtgggttaca
601 tcgaactgga tctcaacagc ggtaagatcc ttgagagttt tcgccccgaa gaacgttttc
661 caatgatgag cactttttaa gttctgctat gtggcgcggt attatcccgt attgacgccg
721 ggcaagagca actcgtgcgc cgcatacact attctcagaa tgacttggtt gagtactcac
781 cagtcacaga aaagcatctt acggatggca tgacagtaag agaattatgc agtgctgcca
841 taaccatgag tgataacact gcggccaact tacttctgac aacgatcgga ggaccgaagg
901 agctaaccgc ttttttgcac aacatggggg atcatgtaac tcgccttgat cgttgggaaac
961 cggagctgaa tgaagccata ccaaacgacg agcgtgacac cacgatgcct gtagcaatgg
1021 caacaacgtt gcgcaaaacta ttaactggcg aactacttac tctagcttcc cggcaacaat
1081 taatagactg gatggaggcg gataaagttg caggaccact tctgcgctcg gcccttccgg
1141 ctggctgggt tattgctgat aaatctggag ccggtgagcg tgggtctcgc ggtatcattg
1201 cagcactggg gccagatggt aagccctccc gtatcgtagt tatctacacg acggggagtc
1261 aggcaactat ggatgaacga aatagacaga tcgctgagat aggtgcctca ctgattaagc
1321 attggtaact gtcagaccaa gtttactcat atatacttta gattgattta aaacttcatt
1381 ttttaatttaa aaggatctag gtgaagatcc tttttgataa tctcatgacc aaaatcccct
1441 aacgtgagtt ttcgttccac tgagcgtcag accccgtaga aaagatcaaa ggatcttctt
1501 gagatccttt ttttctgcgc gtaatctgct gcttgcaaac aaaaaaacca ccgctaccag
1561 cggtggttgg tttgcccgat caagagctac caactctttt tccgaaggta actggcttca
1621 gcagagcgca gataccaaat actgtccttc tagtgtagcc gtagttaggc caccacttca
1681 agaactctgt agcaccgcct acatacctcg ctctgctaatt cctgttacca gtggctgctg
1741 ccagtgcgca taagtctgtt cttaccgggt tggactcaag acgatagtta ccggataagg
1801 cgcagcggtc ggggtgaacg ggggttctgt gcacacagcc cagcttgagg cgaacgacct
1861 acaccgaact gagataccta cagcgtgagc attgagaaag cgccacgctt cccgaaggga
1921 gaaaggcgga caggtatccg gtaagcgcca gggctcggaac aggagagcgc acgagggagc
1981 ttcacagggg aaacgcctgg tatctttata gtccctgctcg gtttcgccac ctctgacttg
2041 agcgtcgatt tttgtgatgc tcgtcagggg ggcggagcct atggaaaaac gccagcaacg
2101 cggccttttt acggttctct gccttttgct ggccttttgc tcacatgttc tttctgcgt
2161 tatcccctga tctgttggtt aaccgtatta ccgcctttga gtgagctgat accgctcgcc
2221 gcagccgaac gaccgagcgc agcagctcag tgagcgagga agcggaagag cgcccaatc
2281 gcaaaccgcg tctcccgcgc cgttggccga ttcattaatg caggttaacc tggcttatcg
2341 aaattaatac gactcactat agggagaccg gcagatccaa tgcggatgga ttcgctcact
2401 cggtgccgcg gcggtgcgaa gttcctgccg tggattgggg gattgcacag acggtggagt
2461 gcttgctaca gctaggtacc cagttgcaca ccaaaccggt cccctgtgcc cctgcgtcta
2521 ggtgtgtcct gtgagcaact gcgcccgcgc gcggcaccgg gaagctgcgg gaacctcaat
2581 ggatgattgg ccgccagatcc ccggcatcac ctcagcacat gggcagcccg agaaggaaac
2641 gagggtagaa gatattccag ggtttccgct agcacaactt ccaagcgctt ctaccgctac
2701 tccacacacc cacgctgaat ggcacaacca ctgtaggcct agcgggtccca gctccccgta
2761 accacgcgac cccagcgacc gaggtcatca tgacactatg acgacaatct ggcgggagac
2821 ttgggaaaaa cgatgacggg cacctacctt gggaacaata gagcaagtcc agtcaacgcc
2881 ctagggaagg gggaaaaaaa aaaagagccc attctcacaa ggagcaacac tgtgctgcta
2941 tcatagtaca ctacaagcaa agtcgggatg ttgctagaat ccgtcggcta cgtcggtaga
3001 tacttcccat aggcagtaag tctaggtatc caatttatag ttccgcgtat ttcttcgtcg
3061 aacggggcac agcaaaacct aacagccgca gtgcctcgca accaacacca agcagtaatc
3121 tcatccatcc atacagatcc tcccacaata tctttacatg cgcacttgcg ctgtgttagc
3181 agtttaggtt ccactacca ggtgacacg agcagcactg acagcaaaact gctgcttggg
3241 gccaaagaaat tggcagtttg ctgagcacct acatatcagt acacatcaac cctggtttaa
3301 ttgttagcct ggtactgtta accgtcatga acgaggacgt ggaggaagaa ccacgctaac
3361 aattgaacca ggggtgatgt atacctctct taccatattt gcttgcttcc acgtccctat
3421 cctcaagcga ttttgttctt gactttttgt atttcaacgc aaaggccatt caggtttctc
3481 atgttcaacc ctttgatgg cgatgcggtt cgtcccagga cctcgtgacc tttgaaagac
3541 gagaaatata tcccaaagca acttctgctt atggacaagc aactgcatgg cggcccacaa
3601 ctcatatctc gcgtaatcat ccaaacgcta ggcccagta ctctactgtt ctcgagacac
3661 ctaccataat atgtacccaa tcagacatga caacagtc aaacacatac cctgatccgg
3721 ccaaaccctg acatccctag atgatggtta ggcacacaaa tggaaaacgg gactcgttgg
3781 ttagggtggg acgcttgatg tgcagccagc caaccgactc tcctcgtccc tctcagatcc
3841 cgttattgccc gtttgaccca gccgcccagc cgccaaacag ctgagggagc caatggtcta
3901 atacaccgct caacgcgccc agtgatggtt caaacttgat tgacagcgaa cgaaccctct
3961 gaaacccggt atagatatgt atagatttga cggggttagct tggcccttaa tacgactcac
4021 tataggcgca attggagctc caccgcgggt gcgcccgcTC ACTGCACATG GTGATCGTGG
4081 TGGTCATGAT GATACATCAC AAACGGATGA TTTGCCTACT GAGCCGTGGC TTCGGGATTG
4141 AATTCCTGCT GATGTTTGCA TTAATTCGGC AGGACTTGGG GAAACTGTAC CTGAAATGTC
4201 GCACAGTGGC TTGGTGCAGG CAGCTGTTGC TCGGAAATC TCCTGCTGGA CTGCCACAAC
4261 ATGTGTGCTT GCTCTGCAGG GCATGTAGGT GTATTGAGTT GGTAGGCAGG CAAGGTGCAA
4321 ACTTTATCAA GCTCTGCAGG GTTCTGTTAGT GTTATCTTCA ATAACCGAT AACACTTAGC
4381 CAGCCCAAGTA TCCTCCTTTC TCCTCTGAGC TCCCTGTTGT GGGCCTGGGG AAGTGCGAAT
4441 CGAAATCACC TTCCGAGGGA GCATCTAGCG ATGAAAGGAA ATTCCTAACC CTATCAATT
4501 GAAAAGGACT GCTCTCGTCA GTTCTGGGCC ATTTGAGTTC TACCAACACT CGTGTAAGGA
4561 ATGAGGGATG TTATGCGACC TTCTTCACAC AATAATTGGA AGTGTCACTT TGCTCCGCTT
4621 TCATCGCCCA GTTAAAGTGA TGGATGAGAA TAGCTACCTA TTGGATCAAA ATATCACAGC
4681 ATAATGCCAC GCCCAGGACA TTGTTCTCGC GTGCCCTTCC ATCGAAAGAA TCACAACGCC

4741 TCCATAAAGC CTGGTGTCTT AGTGTGACAT TCCAATTGGG GGTAGGGAC CGGAGTCCCG
4801 GAAGAGACTC CTTAGGTACC GTGTCTATCA CAAGATTTGT TGTTCCTAAG CCACAGCCAA
4861 GATGTCTGGA TCAAGATATA ACGATGCATT TCAACAGTGA AGGAGGTTCA AAACACGGCG
4921 CCAGAACCTT GGCTCAACCG CTCTCGGGGA TACCGGTAAG GTTGCCGAGA CATATCCAAA
4981 CTCGGGCCGG ACTGTCATCC TCCGGAGGGG GGTTCCAAAGA ACGCGGGGAG TATGATTCTG
5041 ACTGTCTTAC ACCAACCTGG AAAAGTCGCC CAGACGGCAA GATCTTATGG GATCCGTCOA
5101 TATTTGGATCT ATTTGTGTTG CGAACGTCTC AACGTTTCTT GACAGCCGTG TTAATGACTT
5161 CAACAAAGCC ATAAACCGCT GATGTTGCTT TTATCAAAAAG TACACCAAGA AATCCGGGGC
5221 GAAAATTGAC AAGACCACAG GTGAAGAGCT CGAGTATAAG TACGCCACAG ATGACCTTAA
5281 CCGCCTCATG TTCTGGTGAT CCACAACGAG CAAGTTTGAG CAAGTTCCAA TCTCATTTCT
5341 CCCTCTTCCT ATCACCATGA AAAAGCCTGA ACTCACCGCG ACGTCTGTGC AGAAGTTTCT
5401 GATCGAAAAG TTCGACAGCG TCTCCGACCT GATGCAGCTC TCGGAGGGCG AAGAATCTCG
5461 TGCTTTTCAGC TTTCGATGTAG GAGGGCGTGG ATATGTCTCT CGGGTAAATA GCTGCGCCGA
5521 TGGTTTCTAC AAAGATCGTT ATGTTTATCG GCACTTTGCA TCGGCCGCGC TCCCGATTCC
5581 GGAAGTGCTT GACATTGGGG AATTACGCGA GAGCCTGACC TATTGCATCT CCCGCCGTGC
5641 ACAGGGTGTC ACGTTGCAAG ACCTGCCTGA AACCGAACGT CCCGCTGTTC TGACGCCGGT
5701 CGCGGAGGCC ATGGATGCGA TCGCTGCGGC CGATCTTAGC CAGACGAGCG GGTTCGGCCC
5761 ATTTCGGACCG CAAGGAATCG GTCAATACAC TACATGGCGT GATTTCATAT GCGCGATTGC
5821 TGATCCCCAT GTGTATCACT GGCAAACTGT GATGGACGAC ACCGTCAGTG CGTCCGTCGC
5881 GCAGGCTCTC GATGAGCTGA TGCTTTGGGC CGAGGACTGC CCCGAAGTCC GGCACCTCGT
5941 GCACGCGGAT TTCGGCTCCA ACAATGTCCT GACGGACAAT GGCCGCATAA CAGCGGTCTAT
6001 TACTGTGAGC GAGGCGATGT TCGGGGATTTC CCgATACGAG GTCGCCAACA TCTTCTTCTG
6061 GAGGCCGTGG TTGGCTTGTA TGGAGCAGCA GACGCGCTAC TTCGAGCGGA GGCATCCGGA
6121 GCTTGACAGA TCGCCGCGGC TCCGGGCGTA TATGCTCCGC ATTTGGTCTG ACCAACTCTA
6181 TCAGAGCTTG GTTGACGGCA ATTTTCGATGA TGCAGCTTGG GCGCAGGGTC GATGCGACGC
6241 AATCTGTCCGA TCCGGAGCCG GGACTGTCCG GCGTACACAA ATCGCCCGCA GAAGCGCGGC
6301 CGTCTGGACC GATGGCTGTG TAGAAGTACT CGCCGATAGT GGAACCCGAC GCGCCAGCAC
6361 TCGTCCGAGG GCAAAAGGAAT AGTCTAGAAT GGGTTAACGG TATTGGGTAT GGTGGGAACA
6421 TGGTAGTACA TCGTTGAAAC ATGTGAGAGT CAGTACACTT GGTTTAAAGA CGAACCTCC
6481 TTTTACTCCG CGTACTACCC CTATTCTTGT ACATAGTCTC ATCAACCCAA CTTCGCACAT
6541 GCATACCTAC TTGACGTATA CATGATTAAC GCATTTCCCT TTCATCTCA ATCTAAATA
6601 AGCTATCTCT ACCGCGCTAC TCCTCCGCCT TTCTGAAATT CGGTGCCTCC CTCCAATACT
6661 CCTCATCTCT CAACTGATAC TTCTCGTCC AGTCCTGTAT CCTTCAAAG CTCTACACT
6721 GATGTTCTAC ATCCCACAAC AACGTAGTCC CATAAGCCCT CTCCTTCTTA TCCATCCGCA
6781 AATACTCCTC CGTGATACCT TCCTCTTTTG TATCCTCATC AACATACGTA TAAGTCTTGT
6841 TCCACCAATA CGTCATTTGT ACATCCCAT TACACATTAT CGCCTGTCTC AGAAAATCGA
6901 GACAGTGCGC CCGGTGTCG TGATATTCCG CGTCCGAGGT GTTGGGGAAG TAATGCCCTC
6961 GCCAAAAGGA GCGCCGGATG AGGTGAGGCG AGTGAAGCTG GTGGAACATG GAGACGCCGT
7021 AGATGTAGGA ACCCGGCGCG TAGTAGATCT CGGAGGctag ctgattctgg agtgaccag
7081 agggctcatga cttgagccta aaatccgccc cctccaccat ttgtagaaaa atgtgacgaa
7141 ctgctgagct ctgtacagtg accggtgact ctttctggca tgcggagaga cggacggacg
7201 cagagagaag ggtctgagtaa taagcgccac tgcgcccagac agctctggcg gctctgaggt
7261 gcagtggatg attattaatc cgggaccggc cgccctcccg ccccgaaagt gaaaggctgg
7321 tgtgcccctc gttgaccaag aatctattgc atcatcggag aatatggagc ttcacgaaat
7381 caccggcaag aagcgaagga gaatgtgaag ccagggtgtg atagccgtcg gcgaaatagc
7441 atgcccattaa cctaggtaca gaagtccaat tgcttccgat ctggtaaaaag attcacgaga
7501 tagtaccttc tccgaagtag gtagagcgag taccggcgcg gtaagctccc taattggccc
7561 atccggcatc tgtagggcgt ccaaatatcg tgcctctcct gctttgcccg gtgtatgaaa
7621 ccgggaaaggc cgtcaggag ctggccagcg ggcagagacc ggaacacaag ctggcagctg
7681 acccatccgg tgctctgcac tcgacctgct gaggtccctc agtccctggt aggcagcttt
7741 gccccgtctg tccgcccggg gtgtcggcg gggtgacaag gtcgttgctg cagtccaaca
7801 tttgttgcca tattttctg ctctcccccac cagctgctct tttcttttct ctttctttt
7861 ccatcttcag tatattcatc ttcccatcca agaaccctta tttcccctaa gtaagtactt
7921 tgctacatcc atactccatc cttcccatcc cttattcctt tgaaccttcc agttcgagct
7981 ttcccacttc atcgcagctt gactaacagc taccgccgtt gagcagacat caccatGGCC
8041 AAGTTGACCA GTGCCGTTCC GGTGCTCACC GCGCGCAGC TCGCCGAGC GGTGAGTTT
8101 TGGACCGACC GGCTCGGGTT CTCCCGGGAC TTCGTGGAGG ACGACTTCGC CGGTGTGGTC
8161 CGGACGACG TGACCTGTG CATCAGCGCG GTCCAGGACC AGGTGGTGCC GGACAACACC
8221 CTGGCCTGGG TGTGGGTGCG CGGCCGTGGC GAGCTGTACG CCGAGTGGTC GGAGGTCGTG
8281 TCCACGAAC TCCGGGACGC CTCGGGCGCG GCCATGACCG AGATCGGCCA GCAGCCGTGG
8341 GGGCGGGAGT TCGCCCTGCG CGACCCGGCC tGCAACTGCG TGCACCTTCGT GGCCGAGGAG
8401 CAGGACTGAC cgacgcccag caacaccgccc ggtccgagcg ggcccagcgg gtccgagggc
8461 tcggagatcc gtcccccttt tcccttctg atcatcatgta attagttatg tcacgcttac
8521 attcagcccc tccccccaca tccgctctaa ccgaaaagga aggaagttaga caacctgaag
8581 tctaggtccc tatttatttt tttatagtta tgtagtatt aagaacgtta tttatatttc
8641 aaatttttct tttttttctg tacagacgcg tgtacgcatg taacattata ctgaaaacct
8701 tgcttgagaa gggttttggga cgctcgaagg ctttaatttg caagctAACT GCCACCACCA
8761 CGTTTAACAT CCCACGAAC CAAGAAATGCG ATCTATCATT CTGCTCTCCC TCTCACGGC
8821 CTTATCGCC GCCTCTCCCA TCCGCTCTCC TCGAAGCCCC AATACTGCAC TCGTACCGAT
8881 CAACAAGCGC TTCGAAGCCA GTGTCCCTCG GATTTGGGCG CCTAAGAAGG CGGACTTCGA
8941 GCAATTTGAA GATGAGGGCG ACGAGACCTC CACCGAGACC AATGCCGAG CCAATGCCGA
9001 CGCCAATGCC GTAATCAGCA AGCGAAACAA TGGTCTTGG CACTGGCAGT CCTGGAACGA
9061 GGTTCTGAG AAGACTGAGG ATGGTGCTGC TAATGCTGCT GAGGGTGTGA AGGGTGGGCA

9121	GGAGTGAGCT	GCTACTTGGT	ACCTAGCTGA	GCCTTGCCGG	TGAGTAACTT	GGTCGATATT
9181	GATTCCTGCG	CTAGGAGTAT	TGACCGTGTG	CTAATGCTTG	CGTCAGTTAC	CATCAAGTGA
9241	AGTGGCTTTG	TTTGGGTAC	ATGGTTGACG	TCCTGGGATA	CTGTGTGAGA	CTAGGATGCA
9301	TGATGAATTG	AAGCGGTCGG	TACCTTCGTT	TACTTTGGCG	GCTGGAAGCT	AGGTGGAATA
9361	GTCTGAAACA	TATGGAAGAA	AGGGTCATGG	CCATCTTTAC	TTACATGGGA	GGTCCTGGCT
9421	GCTTGCCATT	ATTCAACATA	AGACACACCG	GCCCTTCCCC	TTCTTTCTTT	CTTCCTCTTT
9481	TTTTTCTCTT	GATCAATTCA	GACTCGATAT	AAGCCATTTC	CCTCTTGTA	TTTGCTTCCT
9541	CTTACCAGTC	AGTGATCCCA	TGTGAGGTGC	TGAGTGAGAC	AGGTAGTGAT	ATTGGAATCC
9601	TTACATACCT	GCAAAACTCC	ACCTACAATT	TCCCATCCCT	CCATCGCTCT	CGGCTCTTGC
9661	AACACTCGCG	CTTGCTACC	TCCACCACGC	CCAGGTCCTC	TTACTCCAAG	TCCTCCTTAA
9721	ACTTGGTGAC	GCCGAAGATG	GTGAATGATG	CGGCATTTGC	GTGATGTGGA	GAgatatacgg
9781	caagagtaac	gccggcagca	acggccctgg	taagagcaaa	gtagaagaga	tcggcagcct
9841	caaaggcgat	ttcctggggg	gtctgagcgg	tgcagagctc	ctcagcctcc	tccatgatct
9901	tggccccgac	tagcttctca	tcggagaaga	gacgggcagt	gtaggagccc	tcgggggcag
9961	actgtttcct	cgaaatcaaa	gtctgctcga	gcttgggaag	gcctttgagc	tgaccaaagc
10021	agccggactg	atcgaggtgg	cagaaacgac	ccttctgctt	cacgacaaac	ttgagagcat
10081	cgttatcgca	gtcaagcgag	atgcggaacga	gtccttgagt	gtctccgga	gtagcaccct
10141	tgtaccagag	accgcgcttc	cggctctgat	agacaccagt	ctgtgtcctg	agggcctcgt
10201	tcacactctc	ggcactgctg	tagaccagac	ccagagcagt	gtcgtgctca	tcaacgacaa
10261	cggtggggag	aagaccatcg	ggacggtcag	acttccagac	gctcgagagg	atggtggaga
10321	tggcaagctt	cagctgctcg	agttc			

//