Early-branching gut fungi possess a large, comprehensive array of biomass-degrading enzymes

by Solomon, K.V., Haitjema, C.H., Henske, J.K.,Gilmore, S.P., Borges-Rivera, D., Lipzen, A., Theodorou, M.K., Grigoriev, I., Regev, A., Thompson, D.A. and O'Malley, M.A.

DOI: 10.1126/science.aad1431

Primitive gut fungi have extraordinary degradation capabilities

Authors: Kevin V. Solomon[1], Charles H. Haitjema[1], John K. Henske[1], Sean P. Gilmore[1], Diego Borges-Rivera[2], Anna Lipzen[4], Michael K. Theodorou[3], Igor Grigoriev[4], Aviv Regev[2], Dawn A. Thompson[2], Michelle A. O'Malley[1]*

Affiliations:

[1] Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, CA 93106
[2] Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02143
[3] Animal Production, Welfare and Veterinary Sciences, Harper Adams University, Newport, Shropshire, TF10 8NB, United Kingdom
[4] US DOE Joint Genome Institute, 2800 Mitchell Dr, Walnut Creek, CA 94598
*Correspondence to: momalley@engineering.ucsb.edu

One Sentence Summary: Early branching anaerobic fungi produce the highest number of lignocellulose degrading enzymes recorded in nature, which are unbiased in substrate preference.

Abstract: The fungal kingdom is the source of almost all industrial enzymes in use for lignocellulose bioprocessing. Its more primitive members, however, remain completely unexploited due to culture recalcitrance and poor characterization. We developed a systems-level approach that integrates RNA-Seq, proteomics, phenotype and biochemical studies, allowing for the first comprehensive insight into the lignocellulose degradation abilities in the earliest diverging free-living fungi. Anaerobic gut fungi isolated from herbivores produce the largest known array of biomass-degrading enzymes identified in nature. These enzymes synergistically degrade crude, unpretreated plant biomass, and are competitive with optimized commercial preparations from *Aspergillus* and *Trichoderma*. Compared to these model platforms, gut fungal enzymes are unbiased in substrate preference due to a wealth of xylan-degrading enzymes. Our work reveals that these enzymes are universally catabolite repressed, and we establish that a rich landscape of noncoding regulatory RNAs fine-tunes the hydrolytic response. This study elucidates the dynamic nature of lignocellulose degradation in primitive gut fungi, and illuminates many promising sequence divergent enzyme candidates for lignocellulosic bioprocessing.

Main Text: Lignocellulosic biomass from agricultural and forestry wastes, energy crops, invasive plant species, and pectin-rich food scraps are an abundant, renewable source of fermentable sugars to produce biofuels and sustainable chemicals (*1*, *2*). Industrial routes to make these value-added compounds rely on a suite of enzymes sourced from fungi, nature's recyclers, to convert biomass into the sugars needed for microbial fermentation. However, lignin and other biopolymers must be removed from crude biomass with costly pretreatment processes (*3*) to permit enzymatic degradation and sugar release (*4*). The need for multiple enzyme production processes increases this cost further, as genetically modified fungal platforms such as *Trichoderma reesei* and *Aspergillus nidulans* over produce only limited subsets of enzymes that are unable to independently digest even pretreated substrates completely to sugars (Fig. 1, Table S1) (*5*–*7*). A promising path to economical chemical production is a versatile, unbiased platform capable of producing all the enzymes needed to efficiently hydrolyze diverse lignocellulose feedstocks into fermentable sugars without pretreatment.

Attractive new enzyme platforms that degrade recalcitrant feedstocks reside within microbial communities that routinely process lignocellulose, such as those found in the digestive tract of large herbivores (*8*). Central to these communities are the most primitive free-living fungi that persist to this day, Neocallimastigomycota or anaerobic gut fungi, which are the primary colonizers of biomass in the herbivore gut (*9*, *10*). Ironically, the anaerobic fungi evolved at a time when the Earth's atmosphere lacked oxygen, prior to the emergence of plants; thus, they developed machinery to scavenge the biopolymer-rich cell walls of their primitive neighbors (*11*). As the Earth's atmosphere changed, the anaerobic fungi capitalized on their degradation abilities to thrive in herbivores, where their animal hosts supply an equally diverse diet of lignin-, xylan-, cellulose-, and pectin-rich biomass (Fig. 1, Table S1). As a result, the anaerobic fungi contain a rich repertoire of novel biomass degrading enzymes far exceeding those of other more evolved fungi and bacteria (*12*). However, unlike their aerobic relatives, Neocallimastigomycota remain relatively unspecialized in their choice of biomass substrate with an equal distribution of enzymes. Therefore, the anaerobic fungi are versatile biomass degrading platforms, and even rich untapped sources for new lignocellulolytic enzymes (Fig. 1, 2) (*13*).

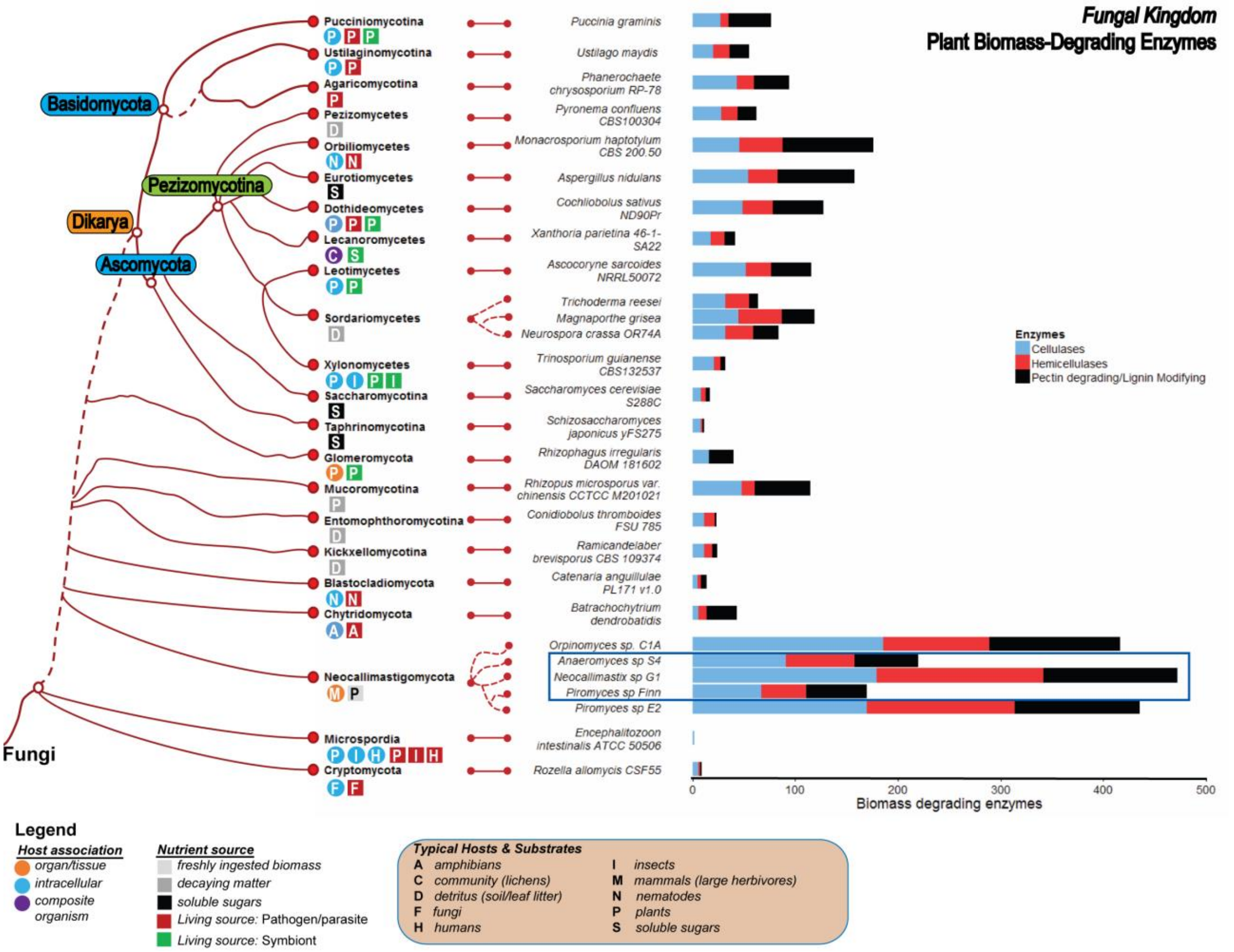

Fig. 1| Biomass degrading machinery in the fungal kingdom. Biomass degrading genes (Table S1) within the genomes of representative fungal species. Boxed species were isolated and their transcriptomes sequenced in this paper (Database S1-S3). Gene numbers for these isolates are estimated from the transcriptome. Fungal Tree of Life adapted from that at Mycocosm (*14*). Common host associations and substrate preferences are indicated below each fungal division.

As unbiased biomass degraders, anaerobic fungi perform an integral role in the decomposition of plant material within the guts of large herbivores. Despite their small numbers (< 8% of the gut microbial community), they rapidly colonize all plant fibers within the gut (*15*) and are capable of degrading 50% of the untreated biomass (*12*). Gut fungi achieve these extraordinary capabilities through a complex lifecycle resembling that of the pathogenic chytrids. Like the chytrids, gut fungi reproduce asexually with motile zoospores that colonize new substrates. When fresh plant biomass is encountered, the zoospores germinate and degrade the substrate through combined invasive growth and secretion of powerful enzymes. Many of these enzymes, including a majority of the hemicellulases, have arisen from horizontal gene transfer with their bacterial counterparts in the herbivore gut (*12*). Due to the intense competition of these microbial communities, horizontal gene transfer, and varied host diet, gut fungi have expanded into six well-established genera (*13*) each expressing a wealth of diverse degrading enzymes (Fig. 1, Table S1) allowing them to effectively degrade crude plant biomass regardless of source. Their strict anaerobic lifestyle coupled with complex nutritional requirements and culture recalcitrance, however, have severely hindered early attempts at isolation, exploitation, and molecular characterization (*13*).

We bridge this knowledge gap by integrating environmental isolation & selection, transcriptome profiling, proteomics, and enzymatic characterization to reveal the hydrolytic capacity of these remarkable microbes across genera for the first time. Included in this analysis is a rich landscape of novel biomass degrading enzymes, long non-coding antisense RNA, and new mechanisms to support their metabolic reprogramming. For this study, we isolated novel specimens that represent half of the six known genera from herbivore fecal samples (*Anaeromyces*, *Neocallimastix*, and *Piromyces*). Hydrolytic capability of each isolate was established before we assembled their transcriptomes *de novo* with next generation sequencing, later verified by proteomics – this offers the first sequence-based insight into the biomass-degrading complexes that anaerobic fungi produce. The global expression profiles of the universal biomass degrader, *Piromyces* sp. *Finn*, was then studied in great detail with catabolite profiling to identify new biomass-degrading genes and shed insight into the conserved mechanisms that regulate them (*16*). More importantly, this regulatory information identifies powerful new cellulase candidates that co-regulate with well-characterized glycosyl hydrolases, which are otherwise invisible to conventional sequence based discovery approaches. Given the primitive positioning of the anaerobic fungi, we also reconstruct the evolutionary inheritance of their capabilities, including conserved and clade specific expansions of function, and identify early ancestors of conserved fungal genes. Here, we demonstrate this method to characterize the unique array of biomass degrading enzymes in the universal degrader, *Piromyces* sp. *Finn*, and explore its powerful hydrolytic response against diverse unpretreated lignocellulosic substrates in extraordinary depth.

Due to the fastidious nature of gut fungi, only a handful of live isolated cultures currently exist. Nonetheless, gut fungi persist in a variety of hosts from which we cultivated our own specimens. We isolated three unique specimens from the fecal samples of herbivorous mammals with very different diets found on opposite sides of the United States. These isolated strains were identified with microscopy and ITS1 sequencing (*17*) as unique gut fungal strains that represent 3 separate genera of Neocallimastigomycota: *Anaeromyces*, *Neocallimastix*, and *Piromyces*. These isolates grew readily on C3/C4 grasses with growth comparable to that on soluble substrates (Fig 2A). *Anaeromyces* displayed some bias in substrate utilization and a clear preference for glucose. In contrast, the monocentric fungi, *Piromyces* and *Neocallimastix*, displayed half the bias in substrate preference with growth rates varying no more that 20% from the mean growth rate across all substrates. Similarly, these fungi had a slight growth advantage on crude lignocellulose, growing up to 20% faster on reed canary grass (*Phalaris arundinacea*), an invasive species and model bioenergy crop (*18*), when compared to glucose.

To evaluate the specific cellulolytic properties of these isolates, we collected and rapidly purified the biomass degrading enzymes from the supernatant of gut fungal cultures by exploiting the ability of many cellulases to bind to cellulose. These purified extracts, which represent a subset of the fungal biomass-degrading enzyme repertoire, were then tested against a number of cellulosic substrates and analogs (Fig S1). Gut fungal secretions were active against all tested substrates demonstrating clear cellulase (Fig S1A-C), β-glucosidase (Fig S1D), and hemicellulase activities (Fig S1E) that were comparable to those from heavily optimized and engineered preparations of *Trichoderma* and *Aspergillus*. Gut fungi, and *Piromyces* in particular, displayed a remarkable ability to access the sugars found within hemicellulose, displaying as much as 300% more activity when compared to commercial enzyme formulations from *Trichoderma* and *Aspergillus* (Fig 2B). Despite this extraordinary hemicellulose activity, gut fungi perform equally well on cellulosic substrates such as carboxymethyl cellulose and display relatively little substrate bias (Fig 2C) in agreement with predictions from the genomic survey (Fig 1). This even distribution of diverse biomass degrading enzymes, and their inherent synergy, broadens the range of substrates that can be degraded effectively (Fig 2) and make gut fungi better suited than their less primitive cousins to effectively degrade both cellulosic and hemicellulosic materials found within crude plant biomass. More importantly, it is this synergy, and not enzyme number, that is responsible for the superior biomass degradation abilities of *Piromyces* (Fig 1, 2). This remarkable ability to degrade diverse substrates without preference and relatively low gene numbers make *Piromyces* a particularly attractive universal degrader and model system for further study.

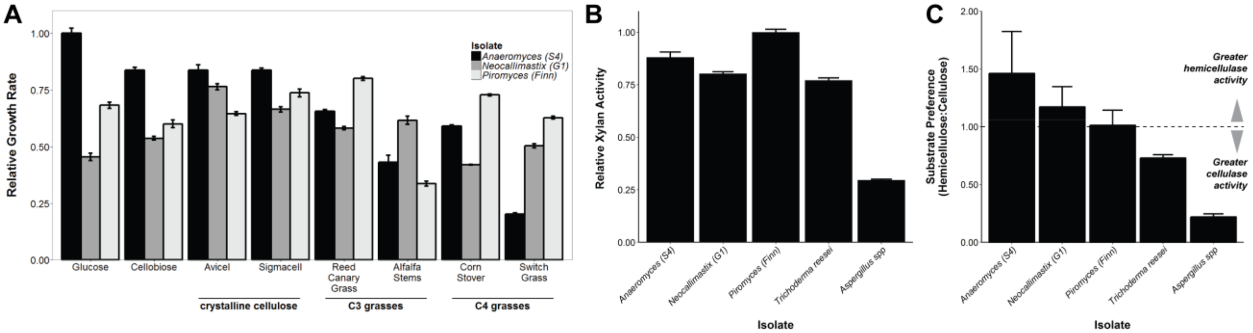

Fig. 2| Functional validation of anaerobic gut fungal biomass degrading capability. (A) Relative growth of gut fungal isolates on a diversity of crystalline cellulose and crude representative C3/C4 bioenergy crops (see Table S3 for specific growth rates). (B) Relative xylan activity of cellulose precipitated gut fungal secretions and commercial *Trichoderma* (Celluclast™) and *Aspergillus* (Viscozyme™) (C) Relative hemicellulose:cellulose activity (xylan vs. carboxymethylcellulose [CMC]) activity of cellulose precipitated gut fungal secretions and commercial preparations. Data represent mean ± SEM of at least 3 samples.

To better understand the remarkable biomass degrading properties of gut fungi, we deep sequenced their transcriptomes, establishing a catalog of genes (Database S1-S3). We collected RNA samples from the fungi grown on a number of soluble and cellulosic substrates to generate a strand specific cDNA library (*19*). These libraries were sequenced and assembled *de novo* (*20*) into transcriptomes each containing roughly 20,000 transcripts transcribed from at least 14,000 genes. The transcriptome of the model *Piromyces* contained more than 27,000 transcripts transcribed from at least 18,000 genes (Database S1). The high quality of this *de novo* assembly was verified by the amplification and Sanger sequencing of selected transcripts, in full or part, which displayed an average identity of 99% to the assembled sequence (Methods). Roughly a third or 8,833 of these transcripts could be annotated either by BLAST or protein domain identification (Database S1) (*21*).

At least 11% of the *Piromyces* transcriptome (2,979 transcripts) are consistent with long noncoding antisense transcripts (asRNA), as established by the orientation of their annotations (Database S1), with strong complementarity to putative target sequences (Fig S2A) within the transcriptome. Putative targets for these asRNA are involved in a number of catalytic and developmental pathways, including biomass degradation, suggesting a broad regulatory role (Fig 3A, Fig S2B). This interpretation is supported by the functional enrichment of antisense in a number of biological process GO terms such as *cellulose catabolic process* ($p_{val} = 0.02$), *ribosome biogenesis* ($p_{val} = 10^{-11}$), *RNA-dependent DNA replication* ($p_{val} = 6 \times 10^{-6}$), and *amino acid transmembrane transport* ($p_{val} = 0.003$) (Database S4). There is a growing consensus that asRNA fulfill a number of regulatory functions (*22*, *23*) and have critical roles in higher fungi (*23*) such as in meiosis in *Saccharomyces cerevisiae* (*24*) and the circadian clock in *Neurospora crassa* (*25*). While analogous roles for asRNA in gut fungi were not examined, our results suggest that these regulatory non-coding transcripts form a pervasive feature of gut fungal genomes and arose early in the evolution of fungal lineages.

Transcripts encoding biomass degrading enzymes comprise ~2% of the gut fungal transcriptomes and contain diverse catalytic functions broadly classified into distinct lignocellulolytic glycosyl hydrolase (GH) families and other carbohydrate active enzyme (CAZyme) domains as recorded in the CAZy database (http://www.cazy.org) (*26*) (Fig. 3A). More than half of these transcripts also encode non-catalytic dockerin domains that are thought to mediate self-assembly of an extracellular catalytic complex or fungal cellulosome (Fig. 3B, C) for synergistic degradation of lignocellulose (*27*). The unique hydrolytic capabilities of gut fungi on native unpretreated biomass are well explained by the functional expansions of many CAZyme families (Table S1, Fig S3). Neocallimastigomycota are rich in hemicellulases (most notably GH10) and polysaccharide deacetylases (Table S1, Fig. 1A), which allow these fungi to effectively remove hemicellulose and access the energy-rich cellulose core of plant biomass in the absence of pretreatment (*28*). This process is greatly aided by pectin removal (*29*) with a number of polysaccharide lyases, carbohydrate esterases and GH88s. This diversity of CAZyme activities confers extraordinary hemicellulase activity to gut fungal extracts, increasing xylan-specific activity relative to commercial preparations of *Trichoderma* and *Aspergillus* by up to 337% (Fig 2B). More importantly, however, it allows these anaerobic fungi to readily degrade an array of lignin-rich C3/C4 bioenergy crops without pretreatment (Fig. 2A).

Functional annotations of the transcriptome were validated within *Piromyces, Anaeromyces,* and *Neocallimastix* (Databases S5-S6) isolates via a proteomic survey of fungal secretions, allowing us to directly link sequence data to protein expression and activity. Proteins secreted from *Piromyces* sp. *Finn* in the presence of reed canary grass were isolated by cellulose precipitation (Fig. 3D, Fig. S4) and mapped using mass spectrometry (*30*) to over 50 cellulolytic transcripts including 25 GH families enriched in or specific to the anaerobic fungal lineage (GH9, GH45, GH48, GH10, GH11). Also present were the full complement of endoglucanases, exoglucanases and β-glucosidases needed to fully depolymerize cellulose (GH5, GH6, GH9, GH45, GH48) and hemicellulases (GH10, GH11) (Fig. 3D, Fig. S4, Table S2), with many transcripts containing dockerin domains for extracellular complex formation (e.g. fungal cellulosomes).

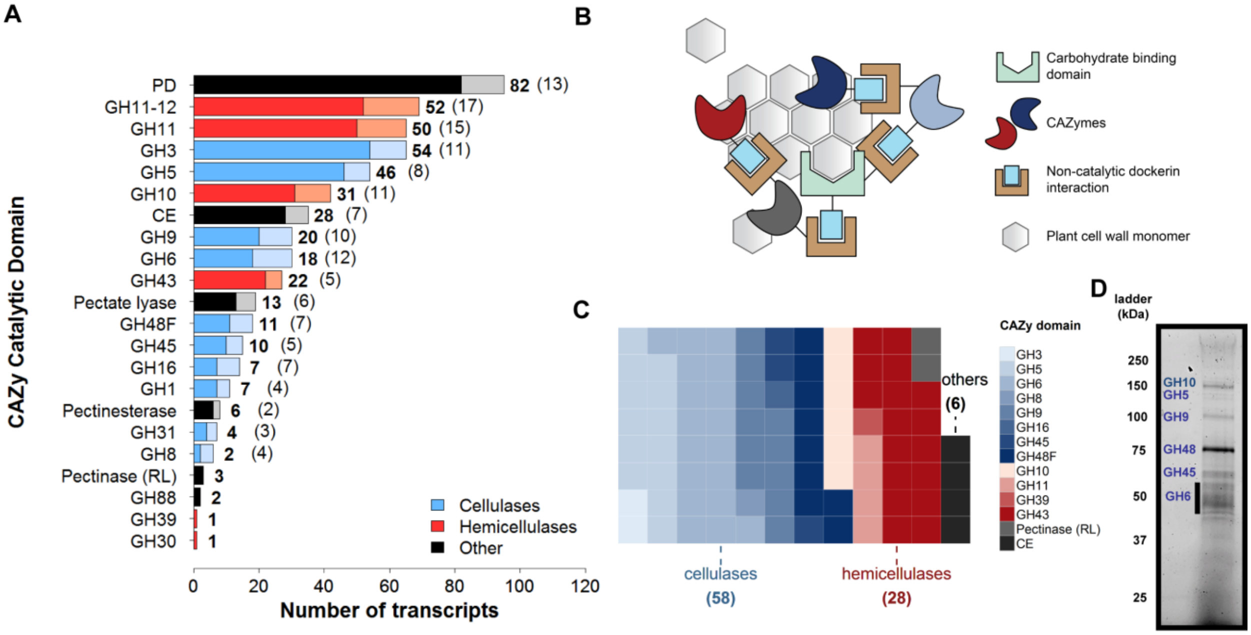

Fig. 3| Biomass degrading machinery in anaerobic gut fungi. (A) Distribution of cellulolytic carbohydrate-active enzyme (CAZyme) transcripts and their regulatory antisense expressed by *Piromyces* sp. *Finn*. Transcripts encoding an enzyme are indicated in bold while antisense transcripts that target them are plotted in a lighter shade and indicated in parentheses. These transcripts are classified into cellulases (blue) that process the cellulose of lignocellulose, hemicellulases (red) that hydrolyze hemicellulose, and other (black), which form the accessory enzymes needed to separate these components from other cell wall constituents such as lignin and pectin. (B) A proposed model for an extracellular catalytic complex for cellulose degradation (*13*). (C) CAZyme composition of the putative extracellular complex. Each square represents a unique enzyme that encodes a CAZyme fused to at least one dockerin domain. PD = polysaccharide deacetylase (acetylxylan esterase), CE = carbohydrate esterase (excluding pectinesterases), RL = Rhamnogalacturonate lyase. (D) Identity of predominant secreted gut fungal CAZYmes in the cellulose-precipitated fraction. In a similar gel (Fig. S4), bands were individually excised and mapped to catalytic functions identified within the transcriptome by tandem MS.

Microbes are parsimonious organisms that typically repress alternate catabolic pathways in favor of glucose when it becomes available. Based on this principle, we hypothesized that cultures grown on lignocellulose down regulate expensive biomass-degrading enzymes in response to glucose addition. Thus, this catabolite repression can be exploited to answer 2 central questions: 1) How are the activities of biomass degrading enzymes coordinated?; and 2) Are divergent proteins present that co-regulate, whose function we may assign through 'guilt-by-association' (*31*, *32*)? We grew *Piromyces* cultures on reed canary grass and then perturbed the system with a small pulse of glucose to induce catabolite repression, collecting RNA samples until the glucose was fully consumed (Fig. 4A). 374 transcripts showed more than a 2-fold change in expression (p ≤ 0.01) with a third of these transcripts containing cellulolytic domains (Fig. 4B). Among these regulated cellulolytic transcripts were all the MS-validated proteins expressed under growth on reed canary grass (Table S2), with the exception of GH45 and XylA. The transcripts associated with biomass degradation were almost exclusively repressed in response to glucose, as expected, and reflected activity trends from cellulose isolated secretions. Expression levels of these transcripts returned to their initial baselines once the glucose was fully consumed (Fig. 4C, Fig. S5). The regulatory patterns of these transcripts also revealed coordinated expression signatures of biomass degradation through cluster analysis (*32*).

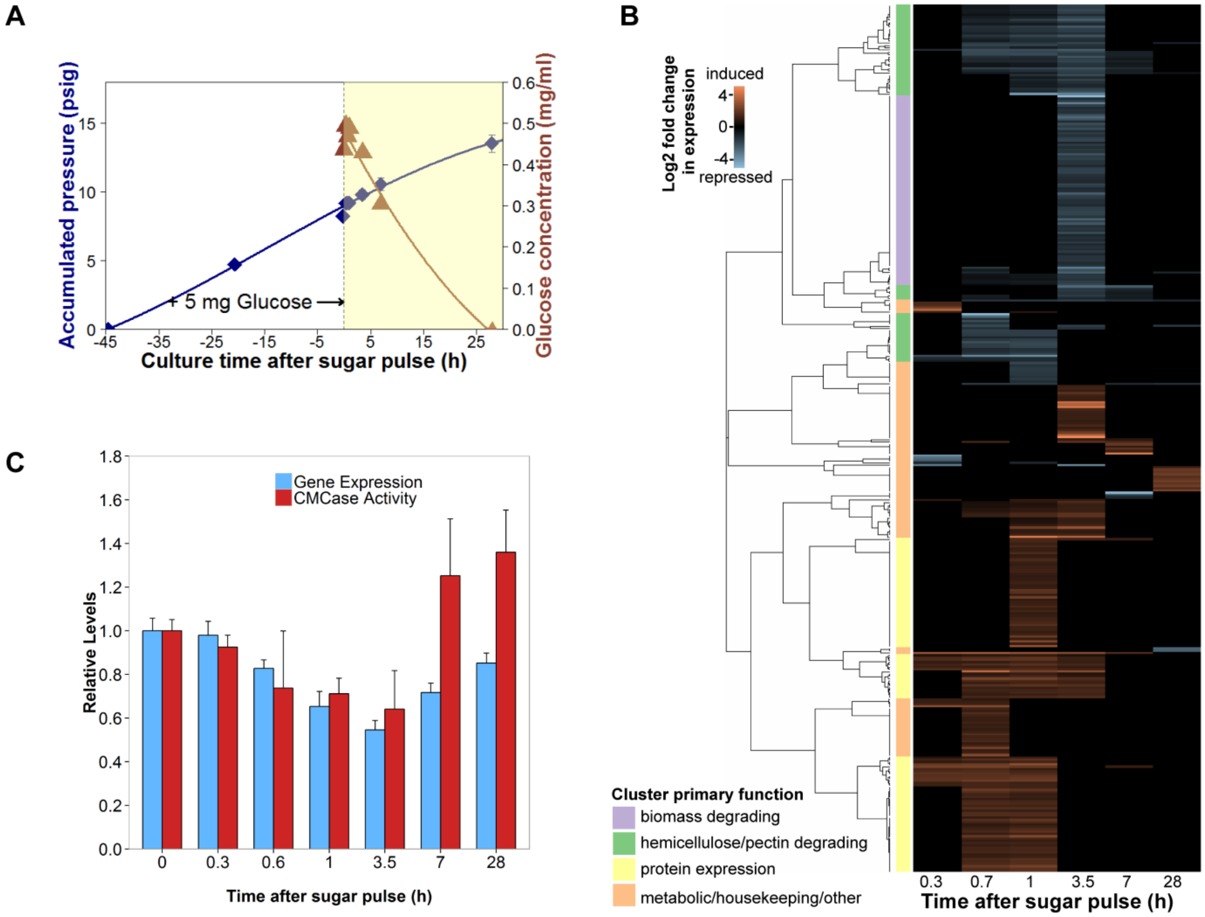

Fig. 4| Global dynamic response to glucose pulse (A) Growth (pressure) and glucose concentration of the sugar perturbation experiments. Cultures were pulsed with 5 mg glucose. mRNA and secretome samples were regularly collected and analyzed after glucose addition (yellow region) until complete consumption of the glucose. (B) Cluster analysis of genes strongly regulated by glucose. Transcript abundance data were compared to uninduced samples at t=0 to calculate the $\log_2$ fold change in expression (*33*). These results were filtered for statistical significance (p≤0.01) and only transcripts with significant regulation (≥2 fold change) are displayed. Clusters are manually annotated based on the most common protein domains/BLAST hits. (C) Relative expression levels (FPKM) of biomass degrading enzymes (Table S1) and their corresponding activity (cellulosome fraction) on carboxy methylcellulose (CMC) (*34*). Data represent the mean ± SEM of ≥2 replicate samples.

Hierarchical cluster analysis revealed 21 distinct clusters or ‘regulons’ of glucose-responsive transcripts containing genes of similar or related function coordinately regulated to achieve a specific goal (Fig. 4C). Biomass degrading enzyme regulons were further specialized into primarily hemicellulose and pectin degrading, or regulons with a broad array of biomass degrading enzymes. Due to the functional enrichment of these clusters, divergent transcripts of unknown function co-regulated with other biomass degrading transcripts may be novel biomass degrading enzymes for biotechnology. Here, we identified 17 such candidates from *Piromyces* (Table S4) that are likely to have unique roles in lignocellulose hydrolysis and are currently being screened.

Biomass degrading enzymes were almost exclusively down regulated in response to glucose at one of two timescales: 40 minutes or 3.5 hours (Fig. 4B). Pectinases, hemicellulases and related accessory enzymes formed distinct regulons, which were rapidly repressed within 40 minutes (Fig. 4B, Database S7). In contrast, cellulases and the remaining biomass degrading machinery responded much later at 3.5 h. This regulatory pattern of more responsive hemicellulases is conserved in a number of contexts in higher fungi (*35*–*38*) and is believed to arise due to the selection pressure of the structure of lignocellulose itself. Hemicellulose and pectin serve to strengthen plant cell walls by surrounding the desired cellulose. Thus, cellulases are needed only after the hemicellulases and pectinases have removed this outer coating. Coordinated expression in this manner will give rise to regulatory pathways for hemicellulases and pectinases that are more responsive than those of cellulases for a common regulatory input, explaining the behavior observed.

Upregulated clusters, in contrast, were consistent with those used for logarithmic growth on glucose and, likely, mediated the cellular response to this sugar pulse. Chief among them were protein expression clusters containing chaperone proteins, rRNA processing proteins, elongation factors and key enzymes in amino acid and nucleotide biosynthesis. Due to the dynamic nature of the glucose pulse, different protein expression clusters, with distinct expression profiles, were upregulated over the course of the experiment (Fig. 4B). One set of clusters was upregulated almost immediately upon glucose addition to deactivate cellulase expression while another set of clusters was induced upon glucose depletion to reactivate cellulase expression. The remaining clusters were less functionally distinct, including a broad array of metabolic, protein expression and housekeeping genes involved in processes such as cell wall synthesis, central metabolism and intracellular transport.

Future platform engineering efforts will rely on the identification and control of regulatory proteins that are responsible for substrate recognition and transcriptional remodeling within gut fungi. Thus, we sought to identify those responsible for the glucose catabolite repression observed. As no receptors or other obvious sensing/signaling proteins were transcriptionally regulated by glucose addition in *Piromyces*, we broadened our search to include unregulated sensors such as the orthologous transcription factors responsible for the conserved hemicellulase/pectinase response in both primitive gut and higher fungi (Table S5). Among these were Cre1/CreABC, the master regulators of fungal carbon assimilation that suppress cellulolytic enzymes in response to glucose, and Xlr-1/XlnR that induces hemicellulase expression upon xylose recognition (*39*). In *Piromyces*, *Anaeromyces*, and *Neocallimastix* isolates , we found complete orthologs of *creB* and *creC*, and strong homologs of *creA* (*40*) suggesting an early evolutionary origin to the CreABC regulatory network. These transcripts, however, share less than 50% sequence similarity with genes from later branching phyla, due in part to the significant A-T bias in gut fungal genomes (*13*) and the corresponding changes in DNA operator sites and transcription factor binding motifs (*41*). Similarly, transcription factor homologs of Xyr-1/XlnR were identified with protein domains characteristic of XlnR.

Unlike *xlnR* in *Aspergillus*, the identified Xyr-1/XlnR homolog in *Piromyes* was not transcriptionally repressed by CreABC activation on the timescales examined. Nonetheless, it is not uncommon over evolutionary timescales for regulation to be handled by transcription factors with different regulatory mechanisms, while still preserving their logical output (*42*). The putative XlnR transcripts were also homologous to other conserved cellulolytic activators across all our *Neocallimastigomycota* isolates suggesting a common evolutionary ancestor to *Ascomycota* cellulolytic transcription factors ACE1-2, ClbR, Clr1-2, and Xyr-1/XlnR (Table S5). Given the high degree of sequence homology, their putative role in the regulation of fungal biomass degradation, and potential for engineering applications, these highly conserved factors should be investigated further to identify specific operator sites and their mechanism of action.

To better understand the regulatory role of key biomass degrading enzymes, we interrogated the system to determine how they were expressed as a function of substrate. *Piromyces* cultures were grown on either glucose, cellobiose, microcrystalline cellulose (Avicel®), filter paper or reed canary grass and transcriptomes were analyzed for differential expression relative to that on glucose. These studies showed significant remodeling of the transcriptome as a function of substrate (2,596 transcripts or ~10% of all transcripts) reflecting changes in both the metabolism and morphology of our gut fungal cultures (Fig. S6). Among these were 194, or half, of the differentially regulated transcripts from the glucose perturbation experiment described above. Overall, a 2-fold change in the expression of biomass degrading enzymes occurred during the switch from glucose to more complex reed canary grass. This trend was mirrored in the activity of cellulose-precipitated secretions (Fig. 5A). Discernible changes in the composition of the biomass degradation machinery also accompanied these variations in expression levels (Fig S7).

Gene set enrichment analysis (GSEA) (*43*) was used to analyze the composition of the biomass degrading machinery as a function of substrate. As expected, the number and functional diversity of CAZyme domains increased as a function of substrate complexity (Fig. 5B). Moreover, insoluble filter paper, Avicel and reed canary grass induced the expression of dockerin tagged transcripts, presumably for synergistic degradation through cellulosome formation. Non-hemicellulosic substrates (cellobiose, filter paper and Avicel) induced expression of a number of seemingly unnecessary hemicellulases such as GH10 suggesting a common regulatory network for many cellulases and hemicellulases. Nonetheless, there still exist independent regulatory networks to induce the additional enzymes needed to degrade crude reed canary grass (Fig. 5B). Our analyses also revealed shifts between enzyme types for similar reactions as a function of substrate, suggesting a highly tailored catabolic response. Cellobiose is a common soluble product of cellulose hydrolysis, which requires β-glucosidases (GH5, GH9) to cleave it into glucose. *Piromyces* sp. *Finn*, however, finely tuned its machinery preferring GH5s for this reaction when grown on cellobiose, and GH9s for reed canary grass, Avicel and filter paper. This flexibility of enzyme choice for a given reaction suggests hidden synergies between all expressed enzymes, and has potential implications for industrial enzyme formulations.

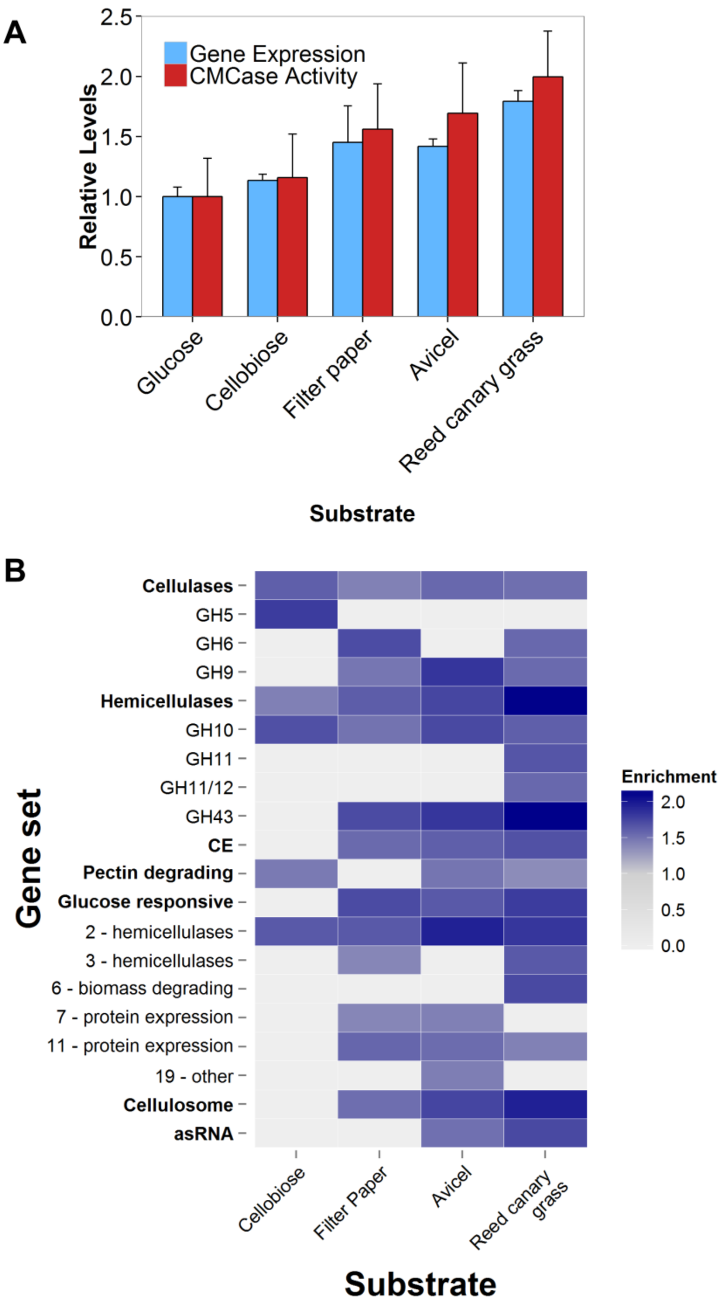

Fig. 5| Substrate specific hydrolytic response (A) Relative expression levels (FPKM) of biomass degrading enzymes (Table S1) and their corresponding activity (cellulosome fraction) on carboxy methylcellulose (CMC) (*34*). (B) Normalized enrichment scores of positively enriched specified gene sets relative to growth on glucose. Gene sets that contain genes that are expressed more highly in a given substrate are indicated (FDR ≤ 10%). Enrichment scores are directly proportional to their expression level. Gene sets indicated in bold are analyzed in aggregate and in subsets (unbolded sets below). asRNA = antisense RNA that target CAZy domains (Fig 3A), Cellulosome = dockerin tagged transcripts. Figures represent the mean ± SEM of ≥ 2 replicates.

Gene sets of the clusters identified in the glucose perturbation experiment (putative regulons) were among those tested for functional enrichment on the array of substrates using GSEA (Fig. 5B). Previously identified protein expression clusters (Fig. 4B), which include proteins such as chaperonins and rRNA processing proteins, were enriched on insoluble substrates (Fig. 5B), confirming their role in mediating expression of lignocellulolytic enzymes. Another regulon, 2- hemicellulases encoding diverse hemicellulases and a handful of cellulases, was central to all growth phenotypes other than glucose. The prevalence of these enzymes, even in the face of non-polymeric carbohydrates, suggests that they play an integral role in the sensing and consumption of insoluble substrates (*39*): in the absence of glucose these enzymes are expressed at a basal level to partially solubilize available cellulosic materials which can then be recognized and trigger a more specific catabolic response. Consistent with this hypothesis is the 6-fold upregulation (p_{val} ~0.02) of the conserved transcription factor XlnR on reed canary grass and Avicel to better recognize these solubilized sugars and induce the gut fungus' extraordinary xylan degrading capabilities. This response is further regulated by asRNA targeting CAZyme domains as evidenced by their functional enrichment on Avicel (p_{val} = 0.003, FDR = 0.03) and reed canary grass cultures (p_{val} ~ 0, FDR = 0.003) (Fig. 5B). An independent analysis using a hypergeometric statistical test confirms that antisense transcripts targeting CAZyme domains (antisense transcripts of *cellulose catabolic process* GO annotation) are functionally enriched under these conditions ($p_{val} \approx 0.01$) (Database S8). The identities of the expressed asRNA, however, are substrate-specific to fine tune the catabolic response through a number of mechanisms (*44*) and conserve cellular resources (Table S6). For example, Avicel induces expression of an antisense transcript that targets, and presumably downregulates, a highly expressed pectate lyase domain, a catalytic function that is superfluous for Avicel hydrolysis. Similarly, reed canary grass induces expression of a GH10 antisense transcript to fine-tune the expression level of the hemicellulase in a substrate-specific manner.

The rich enzymatic repertoire of anaerobic fungi and their versatile substrate degradation capabilities make Neocallimastigomycota particularly attractive targets for the discovery of new biomass degrading enzymes with interesting properties (*12*, *45*). In the absence of standard molecular and genetic tools, we integrated the latest advances in –OMICS technologies with traditional phenotypic and biochemical characterization to obtain the most comprehensive picture of lignocellulose hydrolysis to date in these primitive, unexploited microbes. From new isolates of *Piromyces*, *Anaeromyces*, and *Neocallimastix*, we were able to identify and validate hundreds of novel biomass degrading genes with performance comparable to those from highly engineered and optimized strains of *Trichoderma* and *Aspergillus*. Our catabolic profiling studies in *Piromyces* also revealed the subtle programming of these enzymes that enables these unexploited microbes to degrade diverse substrates with equal efficiency. More importantly, we identified several highly conserved transcription factors that control the expression of key enzymes and establish that putative non-coding antisense RNA tune the cellulolytic response for the first time. Collectively, our data paints the first in-depth picture of transcriptomic remodeling in gut fungi and provides a roadmap for future platform and enzyme engineering efforts.

This study also demonstrates the power of -OMICs based approaches and phenotypic studies to reveal the versatility of these difficult-to-isolate, non-model organisms from nature, and to capture the dynamics of their gene regulatory networks. The characteristic expression signatures captured in these studies may also be used to formulate hypotheses regarding unknown transcripts and to identify novel divergent enzymes for wide use in biotechnology. Leveraging these tools, we obtained a holistic view of the highly tunable biomass degradation machinery in gut fungi, informing industrial hydrolytic strategies, and identified novel candidate enzymes with no homologues in nature. These approaches are readily generalizable to other applications, organisms, and even consortia when genetic tools and reference genomic information are lacking, informing a number of studies aimed at gene discovery and network reconstruction.

References and Notes

1. K. Sanderson, Nature. 444, 673–676 (2006).
2. D. R. Dodds, R. A. Gross, Science. 318, 1250–1251 (2007).
3. K. Sanderson, Nature. 474, S12–S14 (2011).
4. A. Berlin et al., J. Biotechnol. 125, 198–209 (2006).
5. D. Martinez et al., Nat. Biotechnol. 26, 553–560 (2008).
6. J. E. Galagan et al., Nature. 438, 1105–1115 (2005).
7. M. Schülein, in Methods in Enzymology, S. T. K. Willis A. Wood, Ed. (Academic Press, 1988), vol. Volume 160 of Biomass Part A: Cellulose and Hemicellulose, pp. 234–242.
8. M. Hess et al., Science. 331, 463–467 (2011).
9. M. J. Nicholson, M. K. Theodorou, J. L. Brookman, Microbiology. 151, 121–133 (2005).
10. T. Y. James et al., Nature. 443, 818–822 (2006).
11. Y. Chang et al., Genome Biol. Evol. 7, 1590–1601 (2015).
12. N. H. Youssef et al., Appl. Environ. Microbiol. 79, 4620–4634 (2013).
13. C. H. Haitjema, K. V. Solomon, J. K. Henske, M. K. Theodorou, M. A. O'Malley, Biotechnol. Bioeng. 111, 1471–1482 (2014).
14. I. V. Grigoriev et al., Nucleic Acids Res. 42, D699–704 (2014).
15. M. K. Theodorou et al., Proc. Nutr. Soc. 55, 913–926 (1996).
16. K. V. Solomon, C. H. Haitjema, D. A. Thompson, M. A. O'Malley, Curr. Opin. Biotechnol. 28, 103–110 (2014).
17. D. S. Tuckwell, M. J. Nicholson, C. S. McSweeney, M. K. Theodorou, J. L. Brookman, Microbiology. 151, 1557–1567 (2005).
18. S. Lavergne, J. Molofsky, Crit. Rev. Plant Sci. 23, 415–429 (2004).
19. D. Parkhomchuk et al., Nucleic Acids Res. 37, e123–e123 (2009).
20. M. G. Grabherr et al., Nat. Biotechnol. 29, 644–652 (2011).
21. S. Götz et al., Nucleic Acids Res. 36, 3420–3435 (2008).
22. M. A. Faghihi, C. Wahlestedt, Nat. Rev. Mol. Cell Biol. 10, 637–643 (2009).
23. M. E. Donaldson, B. J. Saville, Mol. Microbiol. 85, 405–417 (2012).
24. M. Yassour et al., Genome Biol. 11, R87 (2010).
25. C. Kramer, J. J. Loros, J. C. Dunlap, S. K. Crosthwaite, Nature. 421, 948–952 (2003).
26. V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, Nucleic Acids Res. 42, D490–D495 (2014).
27. S. Raghothama et al., Nat. Struct. Mol. Biol. 8, 775–778 (2001).
28. M. E. Himmel et al., Science. 315, 804–807 (2007).
29. V. Lionetti et al., Proc. Natl. Acad. Sci. 107, 616–621 (2010).
30. E. J. Finehout, K. H. Lee, Electrophoresis. 24, 3508–3516 (2003).
31. J. M. Stuart, E. Segal, D. Koller, S. K. Kim, Science. 302, 249–255 (2003).

32. M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, Proc. Natl. Acad. Sci. 95, 14863–14868 (1998).
33. S. Anders, W. Huber, Genome Biol. 11, 1–12 (2010).
34. B. C. King, M. K. Donnelly, G. C. Bergstrom, L. P. Walker, D. M. Gibson, Biotechnol. Bioeng. 102, 1033–1044 (2009).
35. S. T. Coradetti, Y. Xiong, N. L. Glass, MicrobiologyOpen. 2, 595–609 (2013).
36. U. Bakir, S. Yavascaoglu, F. Guvenc, A. Ersayin, Enzyme Microb. Technol. 29, 328–334 (2001).
37. C. E. Todero Ritter et al., Enzyme Res. 2013, e240219 (2013).
38. M. Hrmová, P. Biely, M. Vršanská, Arch. Microbiol. 144, 307–311 (1986).
39. N. L. Glass, M. Schmoll, J. H. D. Cate, S. Coradetti, Annu. Rev. Microbiol. 67, 477–498 (2013).
40. F. Chen, A. J. Mackey, C. J. Stoeckert, D. S. Roos, Nucleic Acids Res. 34, D363–D368 (2006).
41. T. Portnoy et al., BMC Genomics. 12, 269 (2011).
42. A. E. Tsong, B. B. Tuch, H. Li, A. D. Johnson, Nature. 443, 415–420 (2006).
43. A. Subramanian et al., Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550 (2005).
44. V. Pelechano, L. M. Steinmetz, Nat. Rev. Genet. 14, 880–893 (2013).
45. T.-Y. Wang et al., Biotechnol. Biofuels. 4, 24 (2011).
46. R. Edgar, M. Domrachev, A. E. Lash, Nucleic Acids Res. 30, 207–210 (2002).
47. M. K. Theodorou, J. Brookman, A. P. J. Trinci, in Methods in Gut Microbial Ecology for Ruminants, H. P. S. Makkar, C. S. McSweeney, Eds. (Springer Netherlands, Dordrecht, 2005), pp. 55–66.
48. J. Martin et al., BMC Genomics. 11, 663 (2010).
49. M. W. Duncan, R. Aebersold, R. M. Caprioli, Nat. Biotechnol. 28, 659–664 (2010).
50. M. K. Theodorou, B. A. Williams, M. S. Dhanoa, A. B. McAllan, J. France, Anim. Feed Sci. Technol. 48, 185–197 (1994).
51. T. M. Wood, in Methods in Enzymology, S. T. K. Willis A. Wood, Ed. (Academic Press, 1988; http://www.sciencedirect.com/science/article/pii/0076687988601030), vol. 160 of Biomass Part A: Cellulose and Hemicellulose, pp. 19–25.
52. H. McWilliam et al., Nucleic Acids Res. 41, W597–W600 (2013).
53. I. Letunic, P. Bork, Bioinformatics. 23, 127–128 (2007).

Acknowledgements: We thank Kelvin Lee and Leila Choe (University of Delaware) for proteomics support; Paul J. Weimer (US Dairy Forage Research Center) for lignocellulosic substrates; the Broad Institute core facilities for sequencing and computational assistance in assembling the transcriptome. Sequence and cluster descriptions are included in Databases S1 – S3 and S8 in the supplementary materials. Raw sequence data and transcriptomic profiles reported in this study have been deposited in the NCBI BioProject database (www.ncbi.nlm.nih.gov/bioproject) under BioProject Accession No. PRJNA 291757 (http://www.ncbi.nlm.nih.gov/bioproject/291757). The expression data discussed in this publication have been deposited in the NCBI's Gene Expression Omnibus (*46*) and are accessible through GEO Series accession number GSE64834 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64834). This work was supported by the Office of Science (BER), U.S. Department of Energy (DE-SC0010352), the U.S. Department of Agriculture (Award 2011-67017-20459), and the Institute for Collaborative Biotechnologies through grant W911NF-09-0001. A portion of this research was performed under the JGI-EMSL Collaborative Science Initiative and used resources at the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory, which are DOE Office of Science User Facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under Contract Nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL).

Author Contributions:

K.V.S., C.H.H, D.A.T., and M.A.O. planned the experiments. M.A.O, J.K.H, C.H. H, M.K.T, and K.V.S isolated pure cultures and provided presumptive identification of gut fungi. K.V.S., C.H.H., J.K.H, and M.A.O. performed growth and transcriptomic experiments, C.H.H. performed proteomic analyses, S.P.G. performed enzyme characterization, K.V.S., D.B.R., J.K.H., A.R., I.G. and S.P.G. facilitated bioinformatics analyses of the datasets. K.V.S., D.A.T., and M.A.O. wrote the manuscript.

Figure Legends

Fig. 1| Biomass degrading machinery in the fungal kingdom. Biomass degrading genes (Table S1) within the genomes of representative members in Mycocosm (*14*). Highlighted species were isolated and their transcriptome sequenced in this paper (Database S1-S3). Gene numbers for these isolates are estimated from the transcriptome. Fungal Tree of Life adapted from that at Mycocosm (*14*).

Fig. 2| Functional validation of anaerobic gut fungal biomass degrading capability. (A) Relative growth of gut fungal isolates on a diversity of crystalline cellulose and crude representative C3/C4 bioenergy crops (see Table S3 for specific growth rates). (B) Relative xylan activity of cellulose precipitated gut fungal secretions and commercial *Trichoderma* (Celluclast™) and *Aspergillus* (Viscozyme™) (C) Relative hemicellulose:cellulose activity (xylan vs. carboxymethyl cellulose [CMC]) activity of cellulose precipitated gut fungal secretions and commercial preparations. Data represent mean ± SEM of at least 3 samples.

Fig. 3| Biomass degrading machinery in anaerobic gut fungi. (A) Distribution of cellulolytic carbohydrate-active enzyme (CAZyme) transcripts expressed by *Piromyces* sp*. finn* on either glucose or reed canary grass. Transcripts that encode an enzyme are indicated in bold while antisense transcripts that target them are plotted in a lighter shade and indicated in parentheses. These transcripts are classified into cellulases (blue) that process the cellulose of lignocellulose, hemicellulases (red) that hydrolyze hemicellulose, and other (black) which form the accessory enzymes needed to separate these components from other cell wall constituents such as lignin and pectin. (B) A proposed model for an extracellular catalytic complex for cellulose degradation (*13*). (C) CAZyme composition of the putative extracellular complex. Each square represents a unique gene family that encodes a CAZyme fused to at least one dockerin domain. PD = polysaccharide deacetylase (acetylxylan esterase), CE = carbohydrate esterase (excluding pectinesterases), RL = Rhamnogalacturonate lyase. (D) Identity of predominant secreted gut fungal CAZYmes in the cellulose-precipitated fraction. In a similar gel (Fig. S3), bands were individually excised and mapped to catalytic functions identified within the transcriptome by tandem MS.

Fig. 4| Global dynamic response to glucose pulse (A) Growth (pressure) and glucose concentration of the sugar perturbation experiments. Cultures were pulsed with 5 mg glucose. mRNA and secretome samples were regularly collected and analyzed after glucose addition (yellow region) until complete consumption of the glucose. (B) Cluster analysis of genes strongly regulated by glucose. Transcript abundance data were compared to uninduced samples at t=0 to calculate the log$_2$ fold change in expression (*33*). These results were filtered for statistical significance ($p \leq 0.01$) and only transcripts with significant regulation (≥2 fold change) are displayed. Clusters are manually annotated based on the most common protein domains/BLAST hits. (C) Relative expression levels (FPKM) of biomass degrading enzymes (Table S1) and their corresponding activity (cellulosome fraction) on carboxy methylcellulose (CMC) (*34*). Data represent the mean ± SEM of ≥2 replicate samples.

Fig. 5| Substrate specific hydrolytic response (A) Relative expression levels (FPKM) of biomass degrading enzymes (Table S1) and their corresponding activity (cellulosome fraction) on carboxy methylcellulose (CMC) (*34*). (B) Normalized enrichment scores of positively enriched specified gene sets relative to growth on glucose. Gene sets that contain genes that are expressed more highly in a given substrate are indicated (FDR ≤ 10%). Enrichment scores are directly proportional to their expression level. Gene sets indicated in bold are analyzed in aggregate and in subsets (unbolded sets below). asRNA = antisense RNA that target CAZy domains (Fig 3A), Cellulosome = dockerin tagged transcripts. Figures represent the mean ± SEM of ≥ 2 replicates.

Supplementary Materials:

Materials and Methods

Figures S1-S7

Tables S1-S6

Databases S1-S8