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

Kevin V. Solomon[1,*], **Charles H. Haitjema**[1], **John K. Henske**[1], **Sean P. Gilmore**[1], **Diego Borges-Rivera**[2], **Anna Lipzen**[3], **Heather M. Brewer**[4,5], **Samuel O. Purvine**[4,5], **Aaron T. Wright**[4], **Michael K. Theodorou**[6], **Igor V. Grigoriev**[3], **Aviv Regev**[2], **Dawn A. Thompson**[2,†], and **Michelle A. O'Malley**[1,‡]

[1]Department of Chemical Engineering, University of California, Santa Barbara (UCSB), Santa Barbara, CA 93106, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA 02143, USA

[3]U.S. Department of Energy (DOE) Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

[4]Earth and Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA

[5]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352, USA

[6]Animal Production, Welfare and Veterinary Sciences, Harper Adams University, Newport, Shropshire TF10 8NB, UK

Abstract

The fungal kingdom is the source of almost all industrial enzymes in use for lignocellulose bioprocessing. We developed a systems-level approach that integrates transcriptomic sequencing, proteomics, phenotype, and biochemical studies of relatively unexplored basal fungi. Anaerobic gut fungi isolated from herbivores produce a large array of biomass-degrading enzymes that synergistically degrade crude, untreated 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. These enzymes are universally catabolite-repressed and are further regulated by a rich

landscape of noncoding regulatory RNAs. Additionally, we identified several promising sequence-divergent enzyme candidates for lignocellulosic bioprocessing.

Lignocellulosic biomass from plant matter is an abundant, renewable starting material for biofuel and industrial chemical production (1, 2). Industrial-scale processes require fungal enzymes to convert biomass into fermentable sugars. However, to permit enzymatic degradation and sugar release (3), lignin must be removed from crude biomass via costly pretreatment processes (1). The need for multiple enzyme production processes increases this cost further, as genetically modified fungal platforms such as *Trichoderma reesei* and *Aspergillus nidulans* overproduce limited subsets of enzymes that are unable to independently digest even pretreated substrates completely to sugars (table S1) (4). Economical chemical production will require a versatile, unbiased platform to produce all of the enzymes needed to hydrolyze diverse lignocellulose feedstocks into fermentable sugars without pretreatment.

Microbes found in the digestive tract of large herbivores are attractive enzyme platforms for lignocellulose processing (5). Among these are Neocallimastigomycota (anaerobic gut fungi), the primary colonizers of biomass in ruminants and the earliest-branching nonparasitic fungi still living (6). Although Neocallimastigomycota account for ~8% of the gut microflora, they degrade up to 50% of the untreated biomass through invasive growth and enzyme secretion (7-9). Neocallimastigomycota contain a diverse repertoire of biomass-degrading enzymes (table S1) that degrade a range of feedstocks with equal efficiency (Fig. 1), making them rich untapped sources for previously unidentified lignocellulolytic enzymes. However, their strict anaerobic lifestyle, complex nutritional requirements, and culture recalcitrance have severely hindered early attempts at isolation, exploitation, and molecular characterization (10).

We isolated three previously uncharacterized cultures from the feces of different herbivorous mammals with varied diets. We used microscopy and ITS1 (internal transcribed spacer 1) sequencing (11) to verify that the isolates were distinct species, representing separate genera of Neocallimastigomycota (*Anaeromyces robustus*, *Neocallimastix californiae*, and *Piromyces finnis*). Each grew on C_3 and/or C_4 grasses at rates comparable to its growth on soluble substrates (Fig. 1A). *Anaeromyces* had a clear preference for glucose and grew more slowly on switch grass (~20% of the glucose growth rate). In contrast, the monocentric fungi, *Piromyces* and *Neocallimastix*, displayed limited substrate preference, with growth rates varying no more than 20% from the mean growth rate across all substrates. Similarly, these fungi had slight growth advantages on crude lignocellulose, growing up to 20% faster on reed canary grass (*Phalaris arundinacea*), an invasive species and bioenergy crop (12), when compared with glucose.

We collected and purified the biomass-degrading enzymes from fungal supernatants by exploiting the ability of many cellulases to bind to cellulose. We then tested the purified extracts for hydrolytic activity against several cellulosic substrates and analogs (fig. S1). Gut fungal secretions were active against all tested substrates, demonstrating cellulase, β-glucosidase, and hemicellulase activities comparable to those from engineered preparations of *Trichoderma* and *Aspergillus*. Neocallimastigomycota, and *Piromyces* in particular,

displayed as much as a 300% increase in xylan-degradation activity when compared with commercial *Aspergillus* enzyme formulations (Fig. 1B). Gut fungi degrade cellulose at similar rates, demonstrating little preference for cellulose or hemicellulose (Fig. 1C), in agreement with their enzymatic distribution from genomic sequencing (table S1). This comprehensive array of biomass-degrading enzymes, and their inherent synergy, broadens the range of substrates that can be degraded effectively, making gut fungi better suited than later-diverging species to degrade diverse polymers found within crude plant biomass. More importantly, it is this synergy, and not enzyme diversity, that is responsible for the superior biomass-degradation abilities of *Piromyces*, making it an intriguing model system for further study.

Transcripts encoding biomass-degrading enzymes comprise ~2% of the gut fungal transcriptomes (data S1 to S3) containing diverse functions classified into distinct lignocellulolytic glycosyl hydrolase (GH) and other carbohydrate-active enzyme (CAZyme) domains (13) (Fig. 2A). The majority of these transcripts also encode non-catalytic dockerin domains thought to mediate self-assembly of an extracellular catalytic complex or cellulosome (Fig. 2, B and C) for synergistic degradation of lignocellulose (14). The hydrolytic capabilities of gut fungi on crude biomass are well explained by the functional expansions of many CAZyme families (table S1 and fig. S3). Neocallimastigomycota are rich in hemicellulases (notably GH10) and polysaccharide deacetylases, which allow gut fungi to remove hemicellulose and access the energy-rich cellulose core of plant biomass (15) in the absence of pretreatment. This process is greatly aided by pectin removal (16) with a number of polysaccharide lyases, carbohydrate esterases, and GH88s, allowing the anaerobic fungi to readily degrade an array of lignin-rich C_3 and C_4 bioenergy crops without pretreatment (Fig. 1A).

Functional annotations of the transcriptome were validated within *Piromyces*, *Anaeromyces*, and *Neocallimastix* via a proteomic survey (data S5 to S7). Proteins secreted from *Piromyces* in the presence of reed canary grass were isolated by cellulose precipitation (Fig. 2D and fig. S4) and individually mapped by mass spectrometry (MS) (17) to more than 50 cellulolytic transcripts including 25 GH families enriched in or specific to the anaerobic fungal lineage (e.g., GH9, GH45, GH48, GH10, and GH11). Also present was the full complement of endoglucanases, exoglucanases, and β-glucosidases needed to fully depolymerize cellulose (GH5, GH6, GH9, GH45, and GH48) and hemicellulases (GH10 and GH11) (table S2), with many transcripts containing dockerin domains for extracellular fungal cellulosome formation.

A pervasive feature of gut fungal transcriptomes is long noncoding antisense transcripts (asRNA) (data S1 to S3). At least 11% of the *Piromyces* transcriptome is noncoding and complementary to putative targets involved in a range of catalytic and developmental pathways, including biomass degradation (Fig. 2A and fig. S2). asRNA is functionally enriched (hypergeometric test) in a number of gene ontology (GO) processes, such as the cellulose catabolic process ($P = 0.02$), ribosome biogenesis ($P = 10^{-11}$), RNA-dependent DNA replication ($P = 6 \times 10^{-6}$), and amino acid transmembrane transport ($P = 0.003$) (data S4). These results infer a role for asRNA regulation in fungal cellulose catabolism and

suggest that noncoding asRNAs may be as critical for function in early-branching Neocallimastigomycota as they are in higher fungal lineages (18-20).

To assess how the activities of biomass-degrading enzymes are coordinated, we grew *Piromyces* cultures on lignocellulose and perturbed the system with a small pulse of glucose to induce catabolite repression, after which we collected RNA samples until the glucose was consumed (Fig. 3A). Three hundred seventy-four transcripts showed more than a twofold change in expression (P 0.01), with a third of these transcripts containing cellulolytic domains (Fig. 3B). Among these regulated cellulolytic transcripts were all of the MS-validated proteins expressed under growth on reed canary grass (table S2), with the exception of GH45 and XylA. Transcripts associated with biomass degradation were almost exclusively repressed in response to glucose, as expected, and reflected activity trends from cellulose-isolated secretions (21). Expression levels of these transcripts returned to initial baselines once glucose was fully consumed (Fig. 3C and fig. S5). Cluster analysis revealed coordinated expression signatures of biomass degradation in the regulatory patterns of these transcripts (22).

Hierarchical cluster analysis revealed that glucose-regulated genes performing a common function grouped into 21 distinct clusters or regulons (Fig. 3B). Owing to the functional enrichment of these regulons, divergent transcripts of unknown function that co-regulate with biomass-degrading transcripts may be previously unidentified biomass-degrading enzymes for biotechnology. We identified 17 such candidates from *Piromyces* (table S4). Biomass-degrading regulons were either hemicellulose- and pectin-degrading and rapidly repressed within 40 min or contained a broad array of biomass-degrading enzymes that responded more slowly at 3.5 hours (Fig. 3B and data S8). The faster regulatory response of hemicellulases is conserved in higher fungi (23, 24) and thought to be an adaptation to lignocellulose structure. Hemicellulose and pectin surround cellulose; thus, cellulases act only after the hemicellulases and pectinases remove this outer coating. Given a common regulatory input, coordinating this expression leads to quicker regulation of hemicellulases and pectinases than of cellulases, in agreement with observation. Candidate mediators of this response include conserved orthologs of the fungal master carbon regulator (CreABC); xylose-sensitive transcription factors (Xlr-1 and XlnR); and other conserved cellulolytic activators such as ACE1-2, ClbR, and Clr1-2 (table S5). In contrast, up-regulated clusters contained an array of metabolic and housekeeping genes consistent with logarithmic growth, along with protein expression genes (such as those encoding chaperonins and ribosomal RNA processing proteins) that probably mediated the cellular response to the sugar pulse (data S8).

To better understand the regulation of key biomass-degrading enzymes, we analyzed expression as a function of substrate. *Piromyces* showed substantial remodeling of the transcriptome as the carbon source was varied (~10% of all transcripts), reflecting changes in the biomass-degrading machinery and internal processes of gut fungal cultures (fig. S6 and data S9). Among these transcripts were 194 of the differentially regulated transcripts from the glucose perturbation experiment described above. Overall, a twofold change in the expression of biomass-degrading enzymes occurred during the switch from glucose to complex reed canary grass. This trend was mirrored in the activity of cellulose-precipitated

secretions (Fig. 4A). Discernible changes in the composition of the biomass-degradation machinery also accompanied variations in expression level (fig. S7).

Gene set enrichment analysis (25) of the transcriptomes confirmed that the number and functional diversity of CAZyme domains increased as a function of substrate complexity (Fig. 4B), with insoluble substrates [filter paper, Avicel (Sigma-Aldrich), and reed canary grass] inducing fungal cellulosomes for enhanced degradation. Nonhemicellulosic substrates (cellobiose, filter paper, and Avicel) up-regulated unneeded hemicellulases such as GH10, suggesting a common regulatory network for diverse enzymes. Nonetheless, the additional enzymes necessary to degrade crude reed canary grass are independently regulated. Our analyses also revealed shifts between enzyme types for similar reactions (e.g., GH5 to GH9 as a β-glucosidase) as a function of substrate, demonstrating a highly tailored catabolic response.

Among the gene sets we tested were clusters identified in the glucose perturbation experiment (Fig. 4B). Protein expression clusters (Fig. 3B) regulated by glucose were enriched on insoluble substrates, reaffirming their role in mediating expression of lignocellulolytic enzymes. Another regulon encoding diverse hemicellulases and a handful of cellulases (cluster 2: hemicellulases) was central to all growth phenotypes other than glucose. This enzyme prevalence, even on non-polymeric carbohydrates, suggests that enzymes play an integral role in the recognition of insoluble substrates (26): In the absence of glucose, these enzymes are expressed at low levels to partially solubilize available cellulosic materials that can be recognized to trigger a more specific catabolic response. Consistent with this hypothesis is the sixfold up-regulation ($P \sim 0.02$, negative binomial test) of the conserved transcription factor XlnR on reed canary grass and Avicel to better recognize solubilized sugars and induce fungal xylan degradation. This response is further regulated by asRNA targeting CAZyme domains, as evidenced by their functional enrichment on Avicel [$P = 0.003$, false discovery rate (FDR) = 0.03] and reed canary grass cultures ($P \sim 0$, FDR = 0.003). An independent analysis using a hypergeometric statistical test confirms that antisense transcripts targeting CAZyme domains (cellulose catabolic process GO annotation) are functionally enriched among the regulated transcripts ($P \approx 0.01$) (data S10). Identities of the expressed asRNA, however, are substrate-specific to modify the catabolic response through a number of mechanisms (27) to conserve cellular resources (table S6).

Overall, our results show that anaerobic gut fungi tailor their hydrolytic response to lignocellulose, implying a coordination in catalysis between all expressed enzymes that may inform industrial hydrolysis strategies. The clear transcriptional signatures of these biomass-degrading enzymes provide a route to identify hundreds of sequence-divergent enzyme candidates with commercial potential from anaerobic microbial communities (28).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES AND NOTES

1. Sanderson K. Nature. 2011; 474:S12–S14.

2. Dodds DR, Gross RA. Science. 2007; 318:1250–1251. [PubMed: 18033870]

3. Berlin A, et al. J. Biotechnol. 2006; 125:198–209. [PubMed: 16621087]

4. Schülein M. Methods Enzymol. 1988; 160:234–242.

5. Hess M, et al. Science. 2011; 331:463–467. [PubMed: 21273488]

6. Nicholson MJ, Theodorou MK, Brookman JL. Microbiology. 2005; 151:121–133. [PubMed: 15632432]

7. Orpin CG, Gen J. Microbiol. 1977; 99:107–117.

8. Theodorou MK, et al. Proc. Nutr. Soc. 1996; 55:913–926. [PubMed: 9004333]

9. Youssef NH, et al. Appl. Environ. Microbiol. 2013; 79:4620–4634. [PubMed: 23709508]

10. Haitjema CH, Solomon KV, Henske JK, Theodorou MK, O'Malley MA. Biotechnol. Bioeng. 2014; 111:1471–1482. [PubMed: 24788404]

11. Tuckwell DS, Nicholson MJ, McSweeney CS, Theodorou MK, Brookman JL. Microbiology. 2005; 151:1557–1567. [PubMed: 15870465]

12. Lavergne S, Molofsky J. Crit. Rev. Plant Sci. 2004; 23:415–429.

13. Lombard V, Ramulu H. Golaconda, Drula E, Coutinho PM, Henrissat B. Nucleic Acids Res. 2014; 42:D490–D495. [PubMed: 24270786]

14. Raghothama S, et al. Nat. Struct. Mol. Biol. 2001; 8:775–778.

15. Himmel ME, et al. Science. 2007; 315:804–807. [PubMed: 17289988]

16. Lionetti V, et al. Proc. Natl. Acad. Sci. U.S.A. 2010; 107:616–621. [PubMed: 20080727]

17. Finehout EJ, Lee KH. Electrophoresis. 2003; 24:3508–3516. [PubMed: 14595698]

18. Kramer C, Loros JJ, Dunlap JC, Crosthwaite SK. Nature. 2003; 421:948–952. [PubMed: 12607002]

19. Yassour M, et al. Genome Biol. 2010; 11:R87. [PubMed: 20796282]

20. Donaldson ME, Saville BJ. Mol. Microbiol. 2012; 85:405–417. [PubMed: 22703386]

21. King BC, Donnelly MK, Bergstrom GC, Walker LP, Gibson DM. Biotechnol. Bioeng. 2009; 102:1033–1044. [PubMed: 18973283]

22. Eisen MB, Spellman PT, Brown PO, Botstein D. Proc. Natl. Acad. Sci. U.S.A. 1998; 95:14863–14868. [PubMed: 9843981]

23. Bakir U, Yavascaoglu S, Guvenc F, Ersayin A. Enzyme Microb. Technol. 2001; 29:328–334.

24. Todero Ritter CE, Camassola M, Zampieri D, Silveira MM, Dillon AJP. Enzyme Res. 2013; 2013:e240219.

25. Subramanian A, et al. Proc. Natl. Acad. Sci. U.S.A. 2005; 102:15545–15550. [PubMed: 16199517]

26. Glass NL, Schmoll M, Cate JHD, Coradetti S. Annu. Rev. Microbiol. 2013; 67:477–498. [PubMed: 23808333]

27. Pelechano V, Steinmetz LM. Nat. Rev. Genet. 2013; 14:880–893. [PubMed: 24217315]

28. Solomon KV, Haitjema CH, Thompson DA, O'Malley MA. Curr. Opin. Biotechnol. 2014; 28:103–110. [PubMed: 24503479]

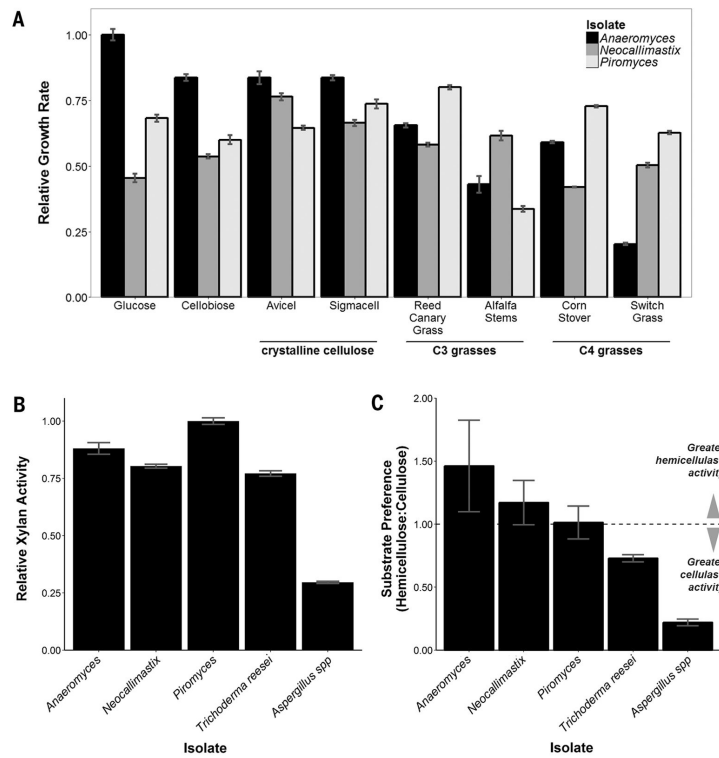29. Edgar R, Domrachev M, Lash AE. Nucleic Acids Res. 2002; 30:207–210. [PubMed: 11752295]

Fig. 1. Anaerobic fungi degrade crude biomass

(**A**) Relative growth of gut fungal isolates on crystalline cellulose and crude C_3 and C_4 bioenergy crops (see table S3 for specific growth rates). (**B**) Relative xylan activity of cellulose-precipitated gut fungal secretions and commercial *Trichoderma* [Celluclast (Sigma-Aldrich)] and *Aspergillus* [Viscozyme (Sigma-Aldrich)]. (**C**) Relative hemicellulose:cellulose [xylan versus carboxymethylcellulose (CMC)] activity of cellulose-precipitated gut fungal secretions and commercial preparations. For all panels, data represent mean ± SEM (error bars) of more than three samples.
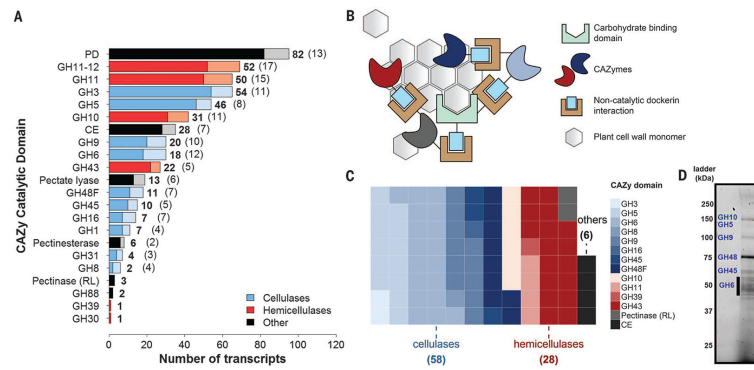
Fig. 2. Anaerobic fungi contain a wealth of biomass-degrading machinery

(**A**) Distribution of cellulolytic carbohydrate-active enzyme (CAZy) transcripts and their regulatory antisense transcripts in *Piromyces*. CAZymes are shown in bold, whereas antisense transcripts are indicated in parentheses and plotted in a lighter shade. In the key, "Other" refers to pectinases and accessory enzymes that separate cellulose and hemicellulose from other cell wall constituents. PD, polysaccharide deacetylase (acetylxylan esterase); CE, carbohydrate esterase (excluding pectinesterases); RL, rhamnogalacturonate lyase. (**B**) Proposed model for an extracellular catalytic complex for cellulose degradation. (**C**) CAZyme composition of the putative extracellular complex. Each square represents a single enzyme that encodes a CAZyme fused to at least one dockerin domain. (**D**) Identity of predominant secreted gut fungal CAZymes in the cellulose-precipitated fraction. Bands were excised and mapped to the transcriptome by tandem MS (fig. S4).
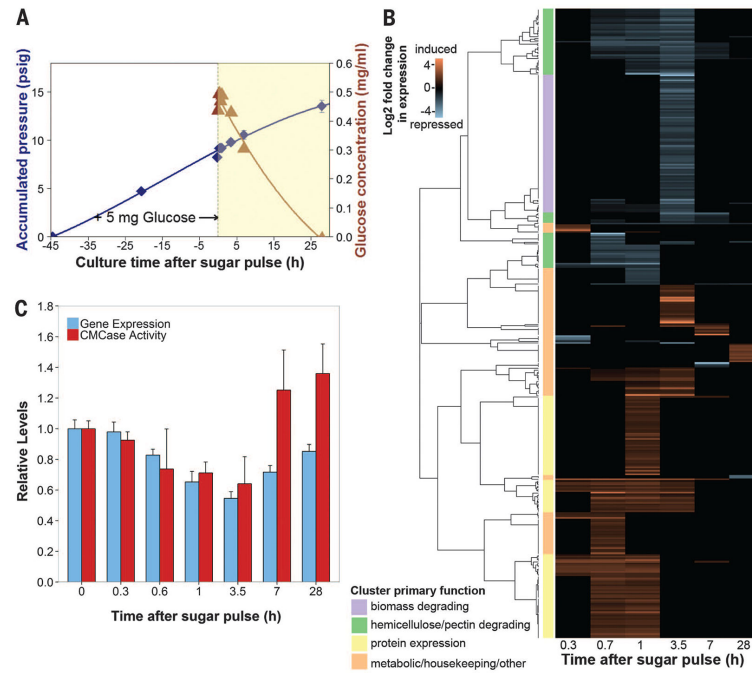
Fig. 3. Anaerobic fungal biomass-degrading machinery is catabolically repressed
(**A**) Glucose consumption in fungal cultures. Exponential cultures of *Piromyces* were pulsed with 5 mg of glucose, and mRNA and secretome samples were collected during glucose depletion (yellow region). Blue diamonds, accumulated pressure; brown triangles, glucose concentration. Error bars indicate SEM. (**B**) Cluster analysis of genes strongly regulated by glucose. Transcript abundance data were compared to uninduced samples at time $t = 0$ to calculate the $\log_2$ fold change in expression. Transcripts with large, significant regulation are displayed (P 0.01, negative binomial distribution, twofold change). Clusters were manually annotated based on the most common protein domains and/or BLAST (Basic Local Alignment Search Tool) hits. (**C**) Relative expression [fragments per kilobase of transcript per million mapped reads (FPKM)] of biomass-degrading enzymes (table S1) and their corresponding activity (cellulosome fraction) on CMC (21). Data represent the mean ± SEM (error bars) of at least two replicates.
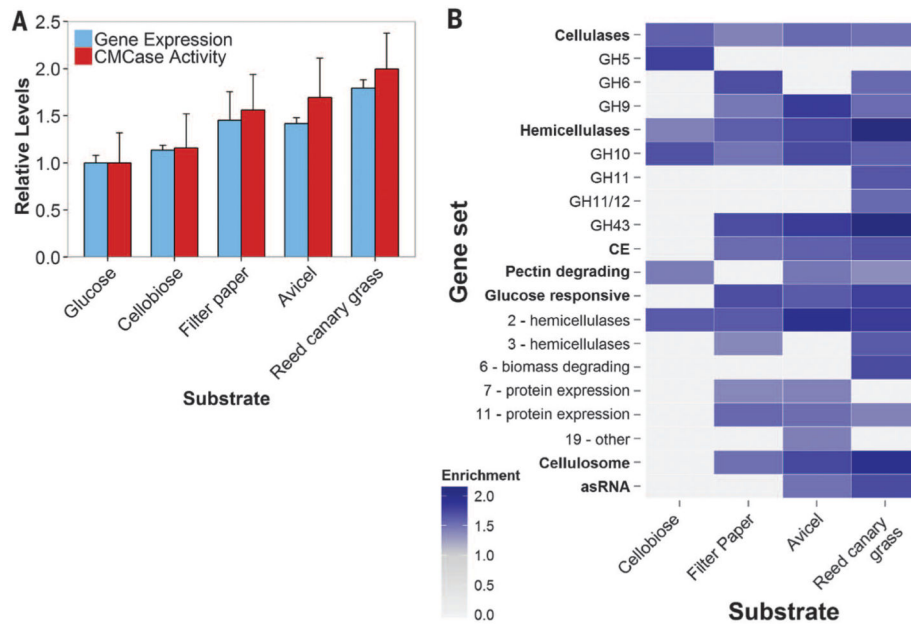
Fig. 4. Anaerobic fungi degrade complex substrates with increasingly diverse enzymes
(**A**) Relative expression (FPKM) of biomass-degrading enzymes (table S1) and their activity (cellulosome fraction) on CMC. Data represent the mean ± SEM (error bars) of at least two replicates. (**B**) Normalized enrichment scores of positively enriched specified gene sets relative to growth on glucose. Gene sets containing genes that are expressed more highly in a given substrate are indicated (FDR 10%; Kolmogorov-Smirnov distribution). Enrichment scores are directly proportional to expression level. Gene sets shown in bold were analyzed in aggregate and in subsets (unbolded sets below). asRNA, antisense RNA that targets CAZy domains (Fig. 2A); Cellulosome, dockerin tagged transcripts. Numbers in the "Glucose responsive" subset indicate clusters.