www.sciencemag.org/cgi/content/full/science.aad1431/DC1

Supplementary Materials for

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

Kevin V. Solomon, Charles H. Haitjema, John K. Henske, Sean P. Gilmore, Diego Borges-Rivera, Anna Lipzen, Heather M. Brewer, Samuel O. Purvine, Aaron T. Wright, Michael K. Theodorou, Igor V. Grigoriev, Aviv Regev, Dawn A. Thompson, Michelle A. O'Malley*

*Corresponding author. E-mail: momalley@engineering.ucsb.edu

Published 18 February 2016 on *Science* First Release
DOI: 10.1126/science.aad1431

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/science.aad1431/DC1)

MATERIALS AND METHODS

Fungal Isolation

Gut fungal strains were isolated from fecal material harvested from a horse (*Piromyces*) from Verril Farms in Concord, MA, and a goat (*Neocallimastix*) and sheep (*Anaeromyces*) located at the Santa Barbara Zoo. Fecal samples were dispersed and mixed well in anaerobic Medium C (*30*)and serially diluted up to 10000 fold. These solutions were then used to inoculate 10 mL Medium C culture tubes containing reed canary grass and chloramphenicol. Cultures (from the higher dilutions) showing signs of growth (gas production, clumping of plant material and no signs of bacterial contamination [turbidity]) were then used to isolate individual strains. This was accomplished by inoculating roll tubes – sealed test tubes coated with a thin layer of agar-containing Medium C plus cellobiose – with 0.1 mL of growing liquid culture. Single colonies were picked from roll tubes and used to inoculate liquid cultures. The process was completed three times in order to ensure an isolated fungal culture was obtained. The first two tybe passages were supplemented with chloramphenicol. Isolated cultures were typed through amplification and sequencing of the internal transcribed spacer (ITS) region of the ribosomal DNA (*11*). These cultures are maintained through routine passaging and cryopreservation for long term storage (*31*).

Strains and culture methods

Gut fungi were grown anaerobically under a headspace of 100% CO_2 at 39^0C in anaerobic Medium C (*30*) supplemented with various carbon sources. Cultures were maintained primarily on reed canary grass in butyl rubber stoppered 15 mL Hungate tubes containing 10 mL of media. The fungus was also routinely grown on soluble sugar substrates, glucose and cellobiose; and cellulosic Avicel (PH 101, 50 µm particle size, Sigma Aldrich, St. Louis, MO). Particulate substrates were added at 10 g/L (0.1 g in 10 mL culture) and soluble substrates added at 5.0 g/L. To maintain viable cell populations, fungal cultures were passaged every 3-7 days by adding 1 mL of a growing culture to a fresh culture tube using a 1 mL syringe to prevent oxygen exposure.

RNA Isolation

RNA was isolated from mid-log cultures (P ~3-8 psig) using a Qiagen RNEasy Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions for Plants and Fungi. Sample integrity was confirmed by RIN score with a BioAnalyzer (Agilent Technologies, Santa Clara, CA). For the *Piromyces de novo* transcriptome assembly, RNA samples from cultures grown on glucose and reed canary grass were prepared. For *Neocallimastix* and *Anaeromyces de novo* assemblies, RNA was collected from samples grown on glucose, reed canary grass, Avicel™, cellobiose, and filter paper.

Piromyces cDNA Library Construction

Total RNA was quantified using the Quant-iT™ RiboGreen® RNA Assay Kit (Life Technologies, Carlsbad, CA) and normalized. An aliquot of each sample was transferred into the dUTP library preparation process (*32*) to generate a strand specific cDNA library. The resultant cDNA progressed through library preparation (end repair, base 'A' addition, adapter ligation, and enrichment) using Broad designed indexed adapters substituted in for multiplexing. After

enrichment, the libraries were quantified with qPCR using the KAPA Library Quantification Kit for Illumina Sequencing Platforms (Kapa Biosystems, Boston, MA) and then equimolar pooled.

Piromyces Transcriptome Acquisition

Pooled libraries were normalized and denatured using 0.2 N NaOH prior to sequencing. Flowcell cluster amplification and sequencing were performed according to the manufacturer's protocols using the HiSeq 2500. Each run was a 76bp paired-end with an eight-base index barcode read. Data was analyzed using the Broad Picard Pipeline which includes de-multiplexing and data aggregation. More than 10^8 total reads were acquired. These reads were aggregated and assembled using Trinity (r2013-02-25) (*33*) into a *de novo* transcriptome of more than 27,000 transcripts with an average sequence depth of more than 400x. The accuracy of this assembly has been validated by cloning and Sanger sequencing more than 5 kb of sequence from transcripts of interest, including putative biomass degrading genes, carbohydrate sensing-GPCRs and regulatory antisense transcripts, in full or in part. Discrepancies were primarily fungal introns (*6*) that were not present in the cDNA library and isolated point mutations, possibly a consequence of errors during the amplification process itself. cDNA libraries from subsequent experiments were sequenced using a MiSeq (Illumina, San Diego, CA). Reads from all conditions were aligned to the *de novo* assembly and expression estimated using RSEM (v 1.2.6) using default settings. Transcripts were grouped into gene families as determined by their component and subcomponent (compXX_c##) grouping within the Trinity platform (*33*).

Anaeromyces and *Neocallimastix* Transcriptome Acquisition

Stranded cDNA libraries were generated using the Illumina TruSeq Stranded RNA LT kit (Illumina, San Diego, CA). mRNA was purified from 1µg of total RNA using magnetic beads containing poly-T oligos. mRNA was fragmented using divalent cations and high temperature. The fragmented RNA was reversed transcribed using random hexamers and Superscript II (Invitrogen, Carlsbad, CA) followed by second strand synthesis. The fragmented cDNA was treated with end-pair, A-tailing, adapter ligation, and 10 cycles of PCR. qPCR was used to determine the concentration of the libraries. Libraries were sequenced on the Illumina HiSeq. Illumina reads of stranded RNA-seq data were used as input for de novo assembly of RNA contigs. Reads were assembled into consensus sequences using Rnnotator v. 2.5.6 or later (*34*).

Transcriptome annotation

The transcriptome was annotated using the automated BLAST2GO package (*35*). First, transcripts were analyzed for homology using BLAST's blastx program against the non-redundant database with an E-value cutoff of 10^{-3}. Transcripts were subsequently analyzed for protein domains using the InterPro database before gene ontology terms and E.C. numbers were assigned. Due to the strand specificity of the library, transcripts with BLAST hits in a reverse orientation (reading frames -1, -2, -3) were non-coding and flagged as antisense transcripts (asRNA). All transcripts were examined for orthology by comparing all possible positive open reading frames with the BLAST-based OrthoMCL against 150 genomes from all domains of life (*36*). Sequences with significant reciprocal best hits within the genome are identified as gene paralogs. Similarly, sequences with significant reciprocal best hits across taxa are assigned as orthologs and grouped into ortholog groups (families). The total paralog and ortholog families as well as remaining sequence component families estimate the number of unique genes within the *Piromyces finnis* genome at 18, 630 gene families. This estimate is consistent with known gut

fungal genome sizes of 14, 868 and 16, 347 genes, respectively, in Piromyces sp E2 (*37*) and Orpinomyces sp. C1A (*9*). Annotations were completed against the NCBI and EMBL databases in Dec. 2012 while orthology analysis was conducted on 31 Mar. 2013.

Antisense RNA analysis

Putative asRNA targets were identified by aligning the reverse complement of each asRNA against the transcriptome with the Smith-Waterman algorithm and a NUC44 scoring matrix. Targets were assigned to the highest scoring alignments. The imperfect complementarity of these target-asRNA pairs (Fig S1A) and independent cloning of a small subset confirm that these non-coding transcripts are not artifacts of library preparation and are indeed transcribed from a unique locus. GO term descriptions of asRNA and their CAZy targets are determined from annotations of the asRNA itself.

Isolation of the cellulosome-containing fraction

The vegetative growth was removed by centrifugation in a swinging bucket rotor at 3,200 g. Then, 0.4% (wt/vol) of SigmaCell 50 (Sigma-Aldrich, St. Louis, MO) was added to the supernatant and incubated with gentle agitation at 4˚C for 2 hours. The cellulose was removed by centrifugation and washed once with 100 mM Tris-HCl pH 7.5 with 100 mM NaCl. Proteins were eluted in water (10 ml per gram of cellulose) by agitation at room temperature for 1 hr. The cellulose was removed by centrifugation and the supernatant (cellulosome-containing fraction) was stored at -80˚C until further analysis. Proteins in the cellulose-precipitated fraction were analyzed by SDS-PAGE and visualized using stain-free imaging technology (Bio-Rad) or SYPRO Ruby total protein stain.

Identification of cellulosome proteins by mass spectroscopy

Cellulosome fractions were precipitated by the addition of ice-cold pure methanol at a ratio of 4:1 (methanol:sample) followed by incubation at -20˚C for at least 1 h. The supernatant was removed and the protein pellet was separated by SDS-PAGE. Proteins bands were excised from the gel and subjected to tryptic digestion followed by tandem MS/MS analysis. The resulting peptide sequences were matched to a database of all possible positive open reading frames generated from transcriptomic sequencing using MASCOT (*38*).

Substrate range of gut fungi

Growth of gut fungi was measured through pressure accumulation of fermentation gases in the head-space of culture tubes (*39*). A variety of substrates were tested including C3 and C4 grasses: reed canary grass, corn stover, alfalfa stems, and switch grass; crystalline cellulose: Avicel, SigmaCell, carboxymethylcellulose (CMC); hemicellulose: xylan (from Beechwood); and soluble sugars: glucose, cellobiose, xylose. Growth curves were plotted as accumulated pressure versus time. The linear portion of the log-linear plot of this data (representing exponential growth) was used to calculate net specific growth rate. These calculated net specific growth rates were then used to compare growth of the fungus across various substrates.

Glucose perturbation

Parallel anaerobic cultures (10 ml) were grown to mid log phase on reed canary grass (~ 2 days) before they were pulsed with 5 mg of glucose. 4 tubes were set aside as an untreated control prior to sugar addition and harvested for RNA isolation and transcriptome quantification.

Samples were taken at regular intervals post glucose addition (20 min, 40 mins, 1 h, 3.5 h, 7 h, and 28 h) until all the glucose was consumed. For each sample, 3-4 tubes were sacrificed and the RNA isolated for transcriptome quantification. Glucose levels were tracked by assaying culture supernatant with a glucose hexokinase – based assay (Megazyme, Bray, Ireland). The remaining supernatant was reserved at -80 °C for later isolation of the cellulosome fraction. RNA samples were stored at -80°C until cDNA prep and analysis.

Cellulase activity assays

Activity on solubilized CMC in the cellulosome fraction was measured by a microplate activity assay essentially as described elsewhere (*21*) with the exception that hydrolysis was performed at 37˚C for 5 h. To determine the activity of gut fungi relative to *Trichoderma* and *Aspergillus*, activity on CMC, phosphoric acid swollen cellulose (PASC), and xylan were measured essentially as described elsewhere (*21*). PASC was prepared as previously described (*40*). Briefly, 30 µL of a 2% substrate solution in PBS (pH 7.4) was combined with 30 µL of the cellulosome fraction suspended in PBS (pH 7.4) or 30 µL of commercial cellulolytic enzyme mixtures from Trichoderma reesei ATCC 26921 (Sigma-Aldrich) and from Aspergillus spp. (Viscozyme® L from Sigma-Aldrich). The commercial enzyme mixtures were diluted 1000-fold in PBS (pH 7.4) to achieve a total protein concentration equal to the *Piromyces finnis* cellulosome fraction as determined by a BCA assay. Activity on Avicel was measured by combining 6 µL of a 10% slurry of Avicel in PBS with 54 µL of the cellulosome fraction or each enzyme mixture. Following hydrolysis at 39˚C for 20 h (Xylan) or 43 h (Avicel, CMC, PASC), the reducing sugar concentration was measured by adding 120 µL of DNS and then heating the solution at 95˚C for 5 minutes. 36 µL of the completed DNS reaction were transferred to 160 µL of water and the absorbance was measured at 540 nm. Rates were calculated by comparing to a standard curve constructed from glucose, and by subtracting a blank measurement where PBS (pH 7.4) was added to the substrate. β-glucosidase activity was determined by adding 30 µL of the cellulosome fraction or enzyme to 970 µL of a reaction mix containing a final concentration of 5 mM pNPG in 50 mM phosphate buffer (pH 7.0) with 2% (w/v) bovine serum albumin. Rates were calculated by tracking the absorbance at 405 nm. In all cases, samples were performed in triplicate, and all values were normalized by total protein as measured by a BCA protein assay kit (Pierce Biotechnology, Rockford, IL).

Catabolic profiling

Triplicate 10 ml anaerobic cultures were grown to mid-log phase (P ~5 psig) on 100 mg of substrate (glucose, cellobiose, Avicel, reed canary grass) and ~1 cm^2 Whatman #1 Filter Paper before samples were harvested and the RNA isolated. Cellulosome fractions were isolated from culture supernatants in a similar experiment.

Differential expression analysis and expression clustering

Differential expression was determined from estimated count data (determined using RSEM) using the Bioconductor DESeq2 package in the R programming language and default parameters (*41*). Results were filtered for statistical significance by adjusted p-values ≤0.01 and $|\log_2 \text{fold change}| \geq 1$. Expression data was subsequently clustered using complete hierarchical clustering based on a Pearson correlation-derived distance metric (*1-r*) of the filtered $\log_2$ fold changes. Clusters were defined at h = 0.5 to form 21 regulons that were manually curated based

on the most frequently occurring functions as determined by protein domain (or BLAST hit if domains were missing).

Gene Set Enrichment Analysis

Enrichment for up-or down regulation of specified gene sets was computed with the GSEA Preranked tool in GSEA v2.0.14 (*25*) against a ranked list of genes. Ranking was based on the $\log_2$ fold change relative to glucose as determined with DESeq2. Gene sets between 15 and 500 members were specified based on protein domains or regulon membership, as indicated. Statistical significance was estimated from 1000 permutations of the dataset gene names.

Phylogenetic Analysis

Sequenced CAZyme families were downloaded from Mycocosm (*37*) and aligned against the corresponding *Piromyces* transcripts with Clustal Omega (*42*) in a multiple alignment with default parameters. The results were used to generate a neighbor-joining phylogenetic tree without distance corrections and plotted with the interactive Tree of Life (*43*).

asRNA functional Enrichment Analysis

To establish the functional roles of antisense RNA, we calculated the fraction of antisense with a given GO annotation and used the hypergeometric distribution to calculate a P-value for this fraction (compared with the null hypothesis of choosing the same number of antisense RNA at random). Similarly, functional enrichment of the substrate-regulated antisense RNA was calculated based on the fraction of regulated antisense ($|\log_2$ fold change$| \geq 1$, $p_{val} \leq 0.01$) with a given GO annotation using the hypergeometric distribution.

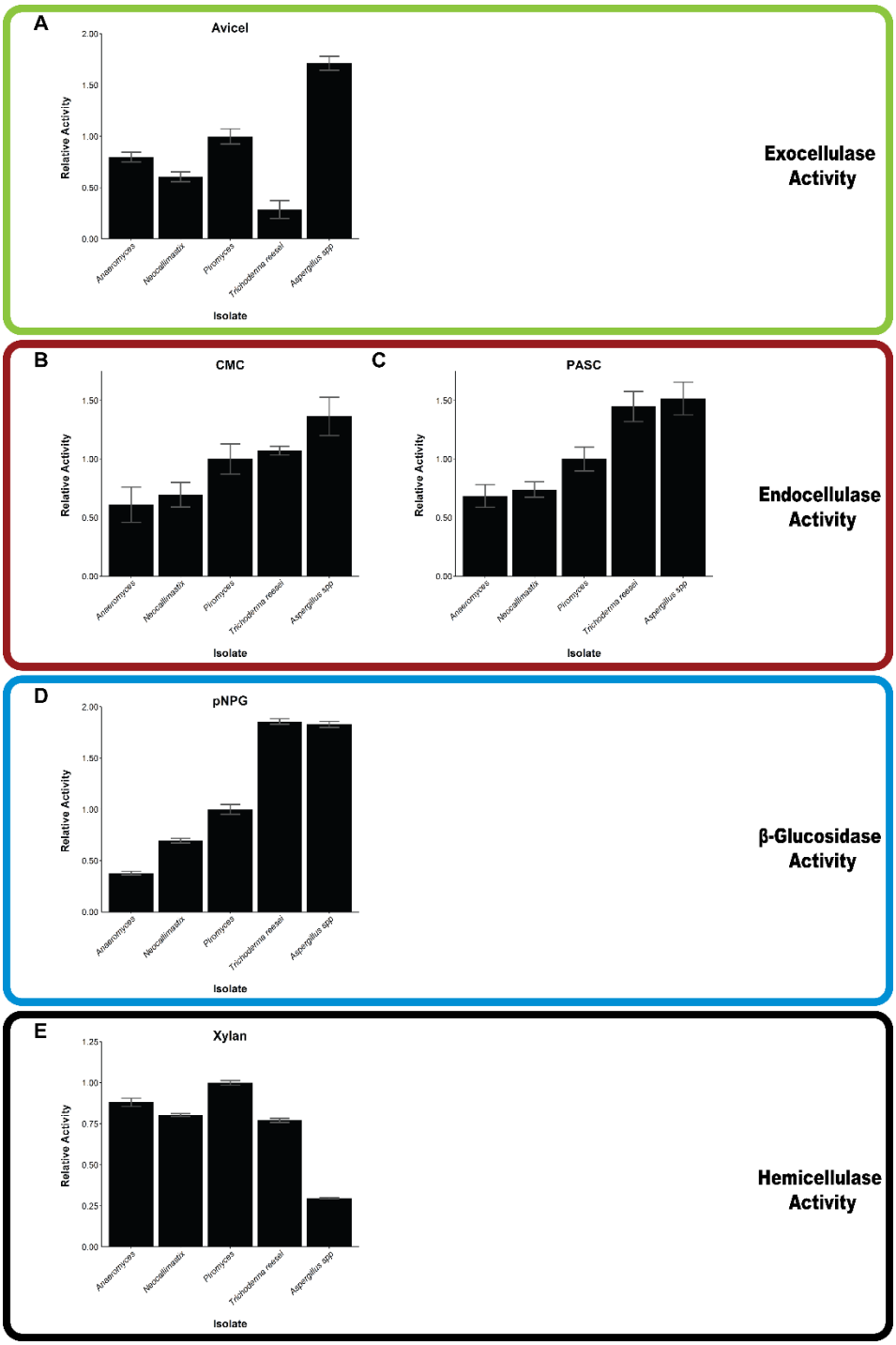

Fig S1|CAZyme Activity of Gut Fungi

Activity of cellulose precipitated gut fungal secretions and commercial *Trichoderma* (Celluclast™) and *Aspergillus* (Viscozyme™) preparations on cellulosic substrates relative to that of *Piromyces*. Data represent mean ± SEM of at least 3 samples. pNPG (4-Nitrophenyl-β-D-glucopyranoside); PASC (phosphoric acid swollen cellulose); CMC (carboxymethyl cellulose)

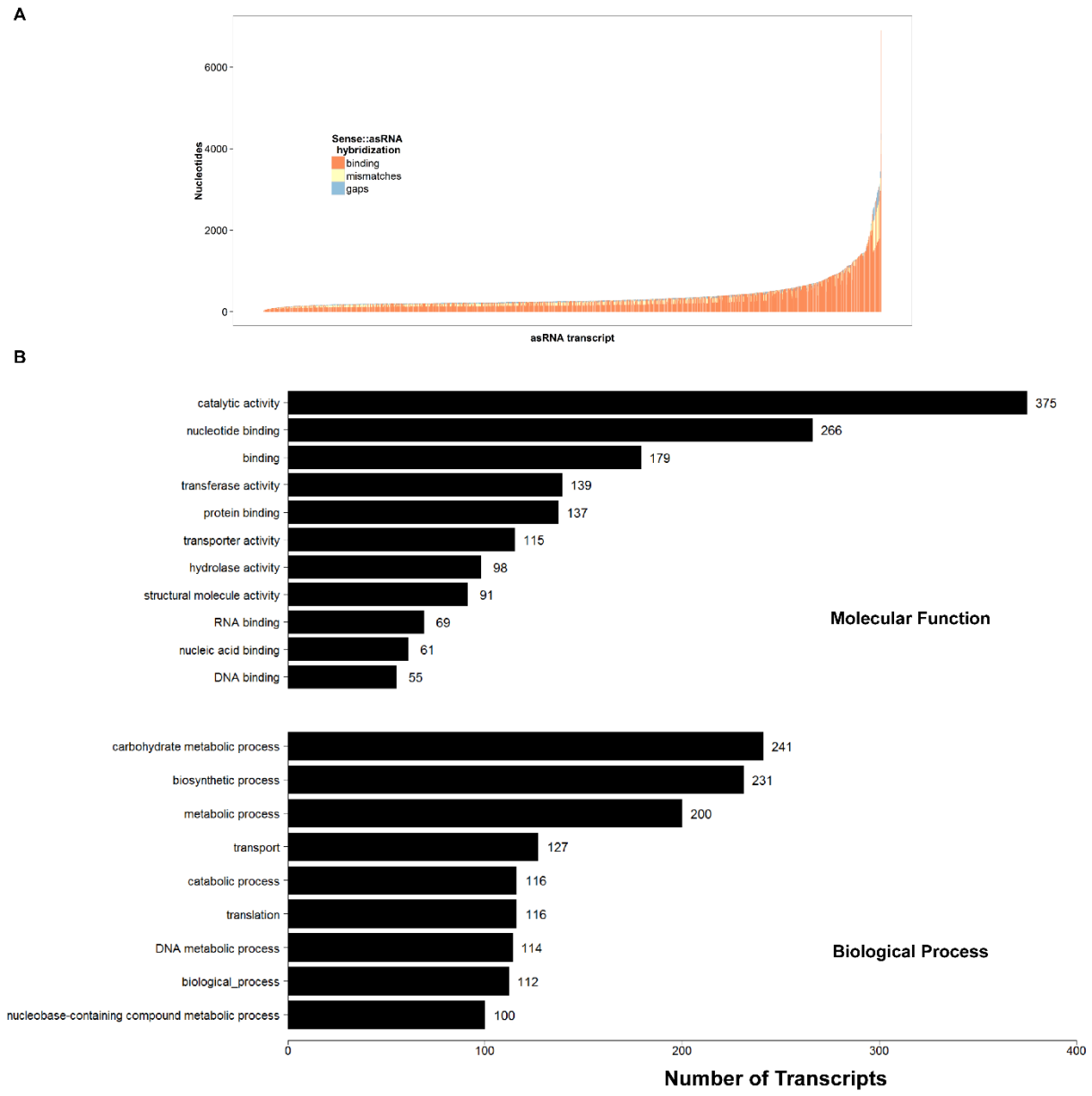

Fig S2|Gut fungal noncoding RNA

(A) Complementarity of sense::asRNA pairs. Putative targets of the asRNA transcripts (sense transcripts) are the highest scoring alignments of the reverse complement of the asRNA against the transcriptome. The presence of gaps and mismatches in hybridized sense::asRNA pairs precludes the possibility of leaky bidirectional expression of a common gene locus or lack of strand specificity in the cDNA library prep and suggests directed expression from specific asRNA loci and/or RNAi-like processing of regulatory non-coding transcripts. (B) Broad functional role of asRNA. GO Term annotation of asRNA protein domain targets in the Biological Process and Molecular Function branches of the Generic GO-Slim subset.

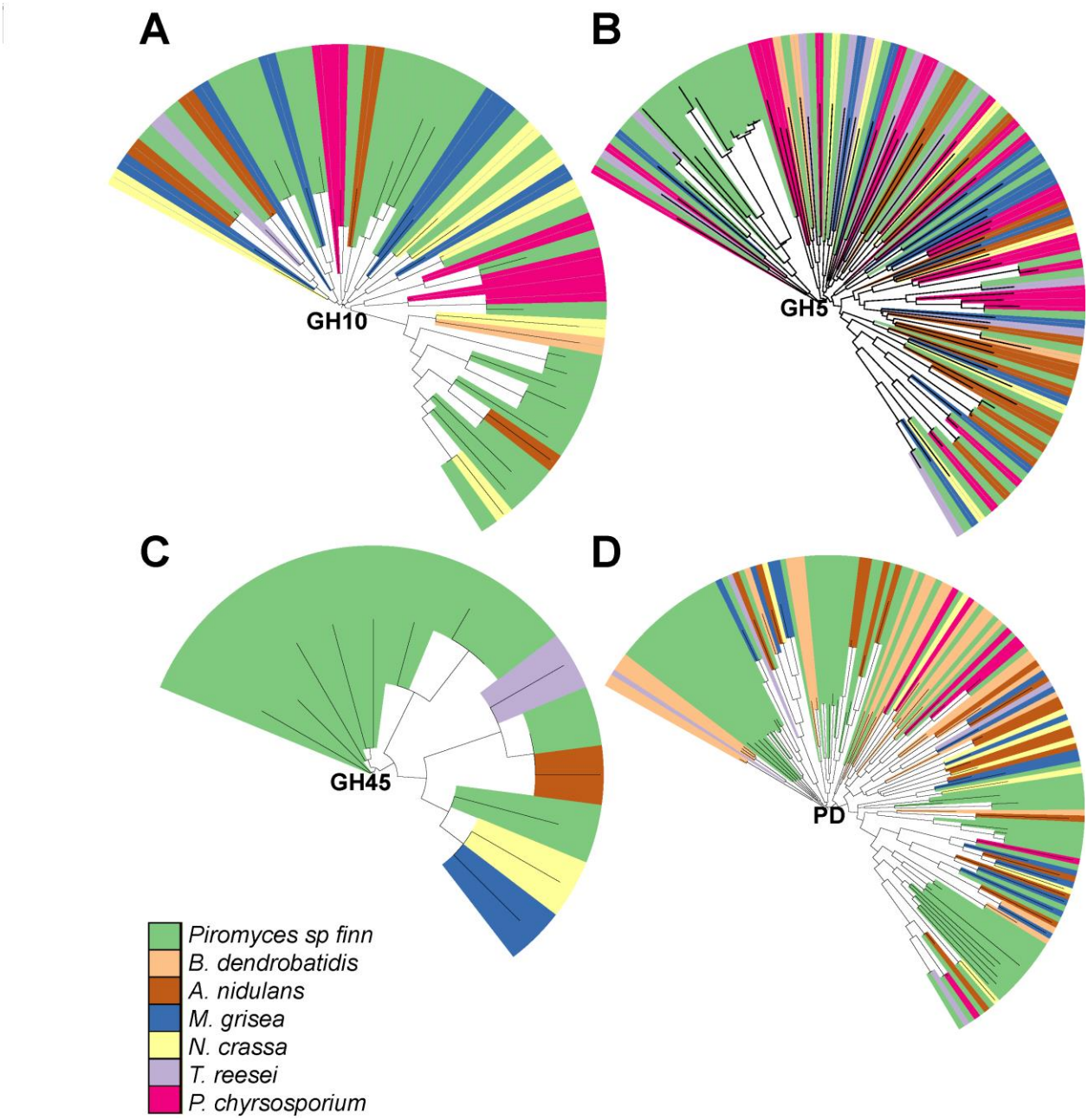

Fig. S3| Amplification of CAZyme functionality within anaerobic gut fungi

Phylogenetic analysis of key cellulase families within sequenced cellulase producers and phytopathogens. (A) Hemicellulase - GH10 family (B) β-glucosidase – GH5 family (C) Endoglucanase - GH45(D) Polysaccharide deacetylases (PD). The amplification of CAZyme functionality in *P. finnis* results from a sequence diversity not seen in later diverging fungi, giving rise to anticipated novel catalytic functionality due to the selective pressures arising from a fibrous diet.

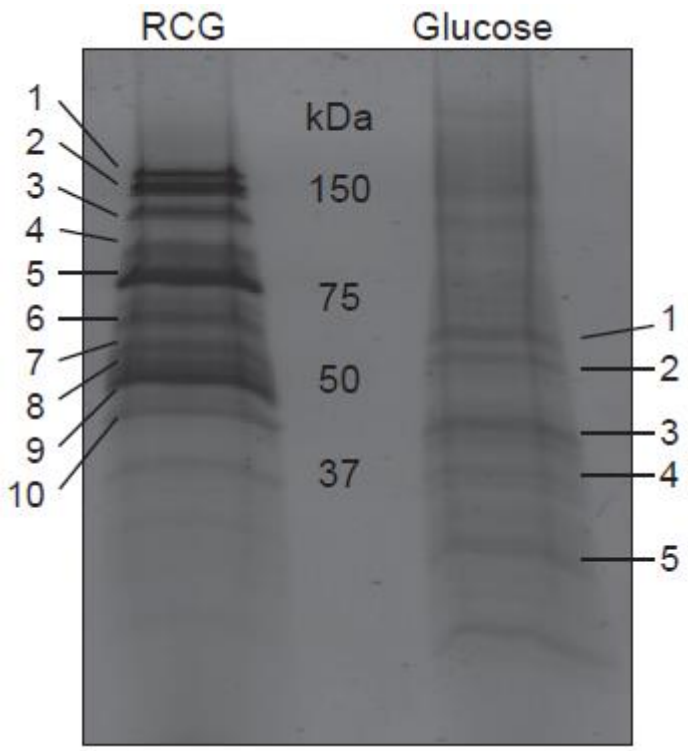

Fig. S4| Cellulose-precipitated fraction of gut fungal cultures grown on either reed canary grass (RCG) or glucose

The indicated bands were excised and subjected to MS analysis (Table S2).

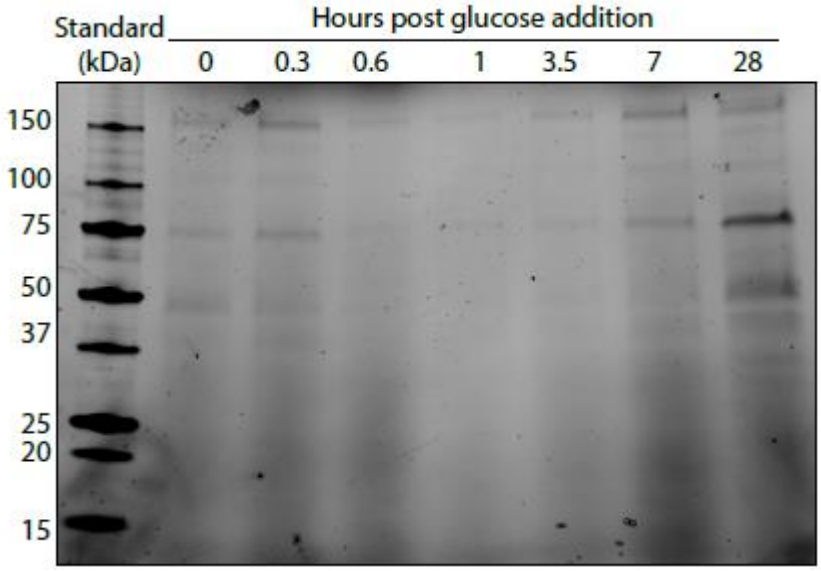

Fig. S5|Repression of biomass degrading machinery in response to glucose

SDS-PAGE gel showing total protein of cellulose-precipitated fraction of the secretome at regular intervals after addition of glucose to a fungal culture growing on reed canary grass.

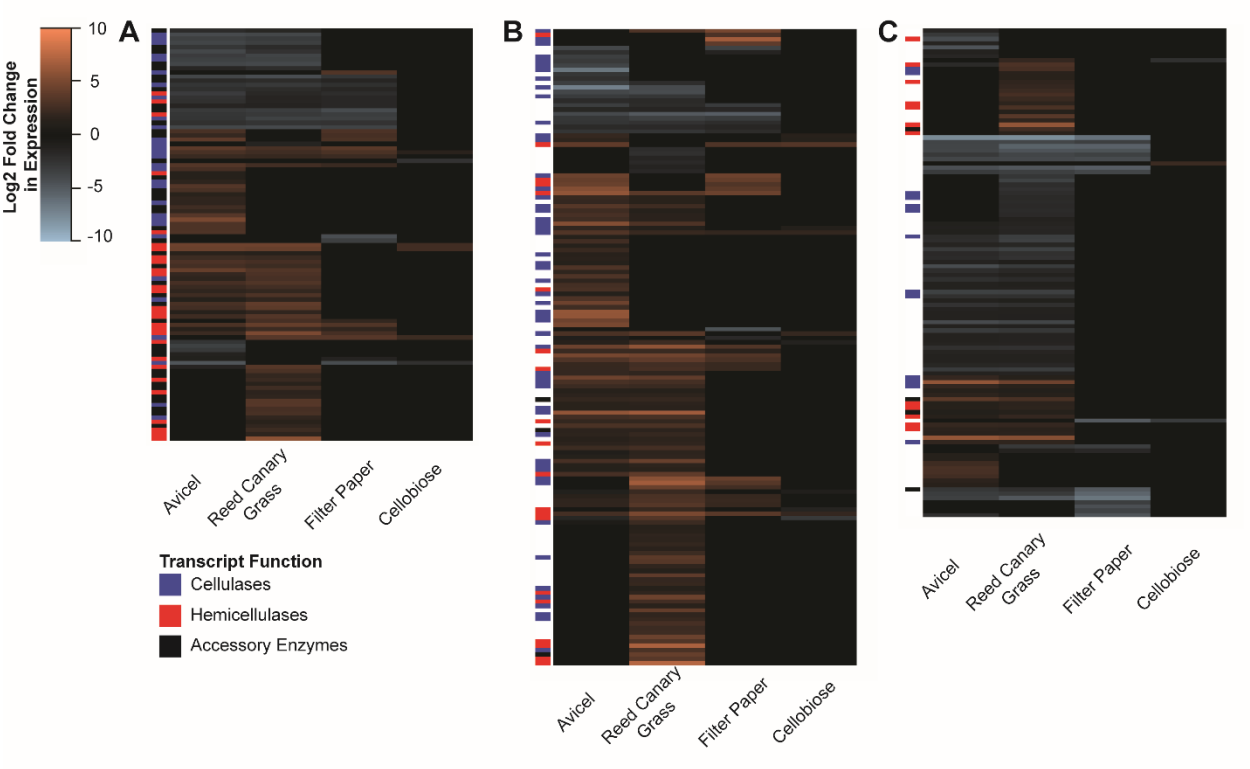

Fig. S6|Differential expression of *Piromyces* transcriptome on various substrates relative to glucose

Expression fold change was filtered for statistical significance ($p \leq 0.01$- Negative binomial distribution) and only transcripts with significant regulation (≥ 2 fold change) are displayed. A) Secreted biomass degrading transcripts (98 transcripts) B) Cellulosome associated transcripts (145 transcripts) C) Antisense RNA transcripts (114 transcripts)

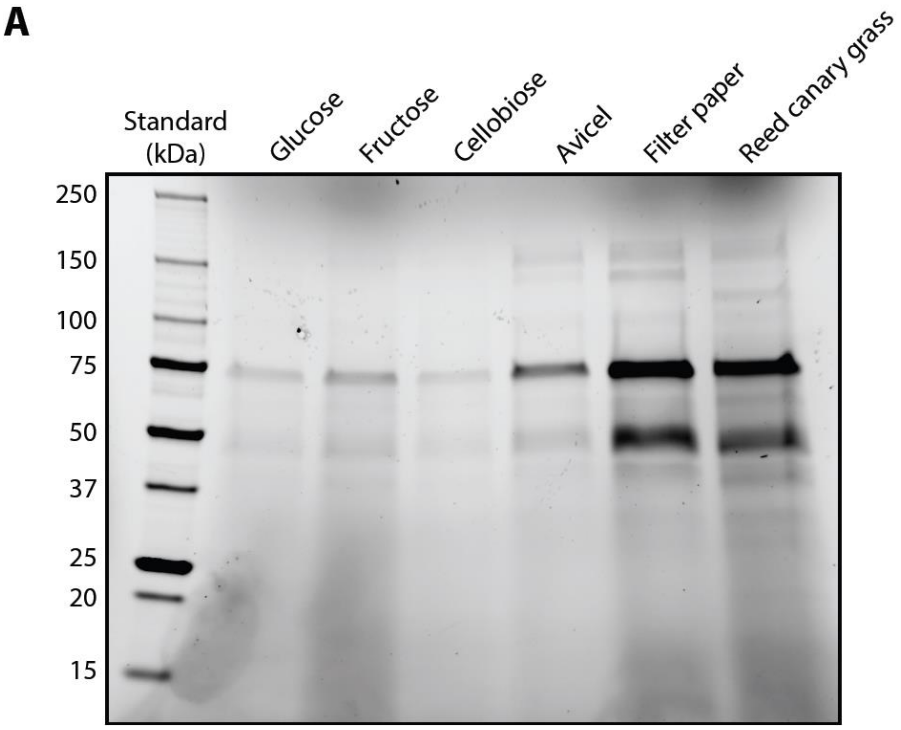

Fig. S7|Substrate specificity of biomass degrading machinery

SDS-PAGE gel showing total protein of cellulosome fraction after growth on the indicated substrates.

Table S1| Biomass degrading genes encoded by Neocallimastigomycota as compared to representative fungal members in JGI's Mycocosm (*37*) **(http://jgi.doe.gov/fungi).**

Catalytic activities enriched in or unique to gut fungi among fungal lineages are highlighted (in grey). Differences between the number of genes and conserved orthologs (indicated in parentheses) for sequenced transcriptomes (highlighted in yellow) represent an estimate of the orthologs unique to gut fungi.

File: *Table S1 – Fungal Tree of Life Biomass Degradation Enzymes.xlsx*

[a]Estimated unique genes as determined by sequence similarity.

[b]PD (CE4) = polysaccharide deacetylase (acetylxylan esterase), CE = carbohydrate esterases, PL = polysaccharide lyase such as pectinases/rhamnogalacturonate lyase

Genome References

Rozella allomycis CSF55	T. Y. James et al., Current Biology. **23**, 1548–1553 (2013).
Encephalitozoon intestinalis ATCC 50506	N. Corradi, J.-F. Pombert, L. Farinelli, E. S. Didier, P. J. Keeling, Nat Commun. **1**, 77 (2010).
Piromyces finnis	This paper
Neocallimastix californiae	This paper
Anaeromyces robustus	This paper
Orpinomyces sp. C1A	N. H. Youssef et al., Appl. Environ. Microbiol. **79**, 4620–4634 (2013).
Rhizopus microsporus var. chinensis CCTCC M201021	D. Wang, R. Wu, Y. Xu, M. Li, Genome Announc. **1**, e00195–12 (2013).
Rhizophagus irregularis DAOM 181602	E. Tisserant et al., PNAS. **110**, 20117–20122 (2013).
Schizosaccharomyces japonicus yFS275	N. Rhind et al., Science. **332**, 930–936 (2011).
Saccharomyces cerevisiae S288C	A. Goffeau et al., Science. **274**, 546, 563–567 (1996).
Neurospora crassa OR74A	J. E. Galagan et al., Nature. **422**, 859–868 (2003).
Magnaporthe grisea	R. A. Dean et al., Nature. **434**, 980–986 (2005).
Trichoderma reesei	D. Martinez et al., Nat Biotech. **26**, 553–560 (2008).
Ascocoryne sarcoides NRRL50072	T. A. Gianoulis et al., PLoS Genet. **8**, e1002558 (2012).
Cochliobolus sativus ND90Pr	R. A. Ohm et al., PLoS Pathog. **8**, e1003037 (2012).
Aspergillus nidulans	M. B. Arnaud et al., Nucl. Acids Res. **38**, D420–D427 (2010).
Monacrosporium haptotylum CBS 200.50	T. Meerupati et al., PLoS Genet. **9**, e1003909 (2013).
Pyronema confluens CBS100304	S. Traeger et al., PLoS Genet. **9**, e1003820 (2013).
Phanerochaete chrysosporium RP-78	R. A. Ohm et al., Fungal Genetics and Biology. **72**, 82–90 (2014).
Ustilago maydis	J. Kämper et al., Nature. **444**, 97–101 (2006).
Puccinia graminis	S. Duplessis et al., PNAS. **108**, 9166–9171 (2011).

Table S2| Identification of cellulose precipitated secretome of *Piromyces* by MS

Mapping of gel plugs from Figure S3 to transcripts within the transcriptome

Growth on reed canary grass			**Growth on glucose**		
Gel plug	**Catalytic domain**	**Transcript Isoform Family**	**Gel plug**	**Catalytic domain**	**Transcript Isoform Family**
1	GH10 GH5	comp11926_c1 comp11637_c1	1	Phosphoenolpyruvate carboxykinase	comp10926_c2
2	GH5 GH6	comp11637_c1 comp11872_c1	2	Malic oxidoreductase GH6*	comp12133_c0 comp12016_c0
3	GH5 GH11	comp11848_c2 comp12025_c4	3	Enolase Xylose isomerase*	comp12158_c0 comp12136_c0
4	GH9	comp11928_c3, comp11930_c4	4	Keto acid reductoisomerase	comp12196_c0
5	GH48	comp12026_c0	5	Glyceraldehyde-3-phosphate dehydrogenase Malate dehydrogenase	comp12123_c0 comp6530_c0
6	GH45	comp10070_c2, comp11629_c3			
7	GH6*	comp12016_c0			
8	GH6 *	comp12016_c0			
9	GH6 *	comp12016_c0			
10	Xylose isomerase *	comp12136_c0			

*Present in both the RCG and glucose samples.

Table S3| Specific growth rates of gut fungal isolates on soluble sugars, crystalline cellulose and a range of C3/C4 grasses.

Isolate	Substrate Type	Substrate	Specific Growth Rate (h^{-1})	Standard error of the mean (h^{-1})
Piromyces finnis	Soluble Sugars	Glucose	0.075	0.013
		Cellobiose	0.066	0.017
	Crystalline Cellulose	Avicel	0.071	0.009
		Sigmacell	0.081	0.017
	C3 Grasses	Reed Canary Grass	0.088	0.008
		Alfalfa Stems	0.037	0.011
	C4 Grasses	Corn Stover	0.080	0.004
		Switch Grass	0.069	0.007
Neocallimastix californiae	Soluble Sugars	Glucose	0.050	0.016
		Cellobiose	0.059	0.008
	Crystalline Cellulose	Avicel	0.084	0.014
		Sigmacell	0.073	0.011
	C3 Grasses	Reed Canary Grass	0.064	0.007
		Alfalfa Stems	0.068	0.018
	C4 Grasses	Corn Stover	0.046	0.001
		Switch Grass	0.055	0.009
Anaeromyces robustus	Soluble Sugars	Glucose	0.110	0.022
		Cellobiose	0.092	0.013
	Crystalline Cellulose	Avicel	0.092	0.024
		Sigmacell	0.092	0.010
	C3 Grasses	Reed Canary Grass	0.072	0.008
		Alfalfa Stems	0.047	0.032
	C4 Grasses	Corn Stover	0.065	0.005
		Switch Grass	0.022	0.004

Table S4|Novel biomass degrading enzyme candidates

Transcripts of unknown function that co-regulated with biomass degrading enzymes

Transcript	Cluster/Regulon
comp12262_c0_seq1	5 – Biomass degrading
comp12026_c1_seq1	6 – Biomass degrading
comp12362_c0_seq1	6 – Biomass degrading
comp7503_c0_seq2	6 – Biomass degrading
comp11992_c0_seq2	6 – Biomass degrading
comp11882_c0_seq1	6 – Biomass degrading
comp11735_c0_seq1	6 – Biomass degrading
comp12028_c12_seq1	6 – Biomass degrading
comp7496_c0_seq1	6 – Biomass degrading
comp5143_c0_seq1	6 – Biomass degrading
comp10778_c1_seq1	6 – Biomass degrading
comp13233_c0_seq1	6 – Biomass degrading
comp6536_c0_seq1	6 – Biomass degrading
comp11012_c2_seq1	6 – Biomass degrading
comp7326_c0_seq1	6 – Biomass degrading
comp14924_c0_seq1	21 – Hemicellulose/Pectin Degrading
comp11723_c0_seq2	21 – Hemicellulose/Pectin Degrading

Table S5| Regulators of lignocellulytic enzymes in gut fungi

Conserved lignocellulolytic transcription factors were analyzed for homology against a direct translation of the gut fungal transcriptomes in all 3 positive open reading frames (*blastp*) to identify their corresponding ortholog *(37).* Highlighted transcripts are the corresponding best BLAST hits for the *Anaeromyces* transcripts.

				Piromyces			***Anaeromyces***			***Neocallimastix***			
Regulator Family	**Gene**	**Query Accession Number**	**Response**	**Best BLAST hit**	**E_{value}**	**Similarity**	**Best BLAST hit**	**E_{value}**	**Similarity**	**Best BLAST hit**	**E_{value}**	**Similarity**	**Orthologous to query?**
Cre-1/ CreABC	*cre-1*	589100213	Repress lignocellulytic enzymes, XlnR, Ace2 on glucose	comp9339_c1_seq1	10^{-16}	51%	Locus12200v2	10^{-15}	38%	Locus7561v1	10^{-14}	45%	N
	cre-1	67476474		comp9339_c1_seq1	10^{-16}	51%	Locus12200v1	10^{-16}	36%	Locus22410v1	10^{-15}	49%	N
	creA	544095		comp9339_c1_seq1	10^{-18}	54%	Locus12200v1	10^{-14}	52%	Locus7561v1	10^{-15}	48%	N
	creB	317025538		comp284583_c0_seq1	10^{-111}	42%	Locus5673v1	10^{-107}	39%	Locus9362v1	10^{-104}	39%	Y
	creC	300680900		comp7502_c0_seq1	10^{-88}	37%	Locus5906v1	10^{-86}	36%	Locus4513v1	10^{-93}	40%	Y
ACE1-2	*ace1*	32699313	• Represses cellulases	comp4011_c0_seq1	10^{-5}	28%	Locus5676v1	10^{-6}	31%	Locus9020v1	10^{-5}	28%	N
	ace2	340518224	• Induces cellulases	comp9196_c0_seq1	10^{-5}	44%	Locus7291v1	10^{-4}	42%	Locus15611v1	10^{-3}	40%	N
ClbR	*clbr2*	399769775	Induces lignocellulytic enzymes on cellulose/cellobiose	comp9196_c0_seq1	10^{-4}	42%	Locus8550v1	10^{-5}	31%	Locus15611v1	10^{-6}	41%	N
Xyr1/ XlnR	*xyr-1*	340517797	Induces hemicellulases on xylan	comp9196_c0_seq1	10^{-6}	43%	Locus8550v1	10^{-5}	35%	Locus15611v1	10^{-7}	48%	N
	xlnR	85108643		comp9196_c0_seq1	10^{-6}	43%	Locus7291v1	10^{-6}	27%	Locus15611v1	10^{-8}	28%	N
	xlnR	292495047		comp9196_c0_seq1	10^{-9}	33%	Locus3826v1	10^{-6}	39%	Locus15611v1	10^{-9}	28%	N
Clr1-2	*clr-1*	553136585	Induces lignocellulolytic enzymes on cellobiose	comp10783_c1_seq1	10^{-7}	54%	Locus8645v1	10^{-7}	35%	Locus11145v1	10^{-6}	45%	N
	clr-2	553136900		comp10199_c2_seq1	10^{-6}	30%	Locus8550v1	10^{-6}	30%	Locus15611v1	10^{-6}	29%	N

Table S6| asRNA fine tuning of biomass degrading machinery

asRNA targeting biomass degrading enzymes functionally enriched on Avicel and reed canary grass (RCG) classified by target and broad function(cellulase, hemicellulase, or other) with average expression in Fragments Per Kilobase of transcript per Million mapped reads (FPKM) +/- SEM indicated in parentheses. The median transcriptome expression level is ≤ 1.

Growth Condition	Antisense	asRNA Avicel Expression	asRNA RCG Expression	Target Domain	Target Avicel Expression	Target RCG Expression	Classification
Both	comp12028_c8_seq7	125.99 (51.95)	10.3 (10.23)	GH48F	20.96 (19)	582.39 (180.86)	cellulases
	comp12029_c3_seq10	676.18 (229.81)	1312.99 (581.43)	GH48F	1319.63 (206.17)	2292.43 (815.13)	cellulases
	comp12009_c1_seq2	5933.22 (633.91)	2008.27 (355.77)	GH43	11.76 (7.72)	0 (0)	hemicellulases
	comp12028_c8_seq6	100.5 (21.12)	9.48 (8.6)	GH48F	22.45 (15.98)	730.08 (141)	cellulases
	comp10490_c0_seq1	57.84 (5.52)	6.11 (1.19)	GH16	80.97 (8.26)	5.5 (3.76)	cellulases
	comp11673_c0_seq1	37.11 (5.46)	21.25 (4.27)	Polysaccharide deacetylase	0 (0)	0 (0)	others
	comp12029_c3_seq9	1434.75 (129.89)	1314.82 (401.32)	GH48F	1319.63 (206.17)	2292.43 (815.13)	cellulases
	comp12028_c8_seq8	1362.7 (574.54)	3803.57 (242.87)	GH48F	18.45 (17.96)	523.71 (90.11)	cellulases
	comp1236_c0_seq1	14.29 (15.6)	35.97 (10.25)	GH5	34.77 (9.66)	202.38 (36.83)	cellulases
	comp12022_c8_seq15	33.6 (4.39)	71.42 (22.47)	Polysaccharide deacetylase	11.46 (11.16)	69.46 (62.77)	others
	comp12022_c8_seq14	32.79 (29.34)	136.51 (61.69)	Polysaccharide deacetylase	6.28 (6.44)	0 (0)	others

	comp12022_c8_seq9	139.33 (59.97)	132.27 (19.02)	Polysaccharide deacetylase	293.26 (10.9)	286.51 (50.55)	others
	comp12029_c3_seq6	793.56 (94.03)	569.84 (128.66)	GH48F	139.47 (139.59)	24.07 (21.38)	cellulases
	comp11637_c0_seq1	866.86 (138.66)	499.5 (25.77)	GH16	269.25 (5.75)	161.56 (46.47)	cellulases
	comp7852_c0_seq1	110.78 (43.85)	101.38 (24.15)	GH3	0 (0)	16.31 (13.99)	cellulases
	comp1479261_c0_seq1	1018.01 (72.42)	1061.63 (74.43)	NA	235.71 (30.59)	266.9 (39.72)	NA
	comp12022_c8_seq12	1408.92 (151.79)	1449.18 (82.74)	Polysaccharide deacetylase	23.79 (29.68)	81.52 (96.48)	others
	comp12028_c11_seq1	958.17 (132.66)	1172.3 (244.63)	Polysaccharide deacetylase	6.28 (6.44)	0 (0)	others
	comp12022_c8_seq6	15.09 (13.46)	12.11 (1.38)	Polysaccharide deacetylase	11.46 (11.16)	69.46 (62.77)	others
	comp12022_c8_seq8	966.19 (73.87)	1246.55 (81.39)	Polysaccharide deacetylase	11.46 (11.16)	69.46 (62.77)	others
	comp12022_c8_seq17	352.34 (48.17)	594.16 (66.39)	Polysaccharide deacetylase	293.26 (10.9)	286.51 (50.55)	Others
	comp11698_c1_seq2	2.24 (1.97)	8.38 (2.49)	GH16	4.8 (2.21)	11.23 (4.32)	cellulases
	comp12022_c8_seq19	138.19 (37.45)	192.23 (18.13)	Polysaccharide deacetylase	67.97 (60)	737.81 (186.79)	others
	comp12022_c8_seq10	745.72 (101.23)	1013.73 (144.47)	Polysaccharide deacetylase	6.28 (6.44)	0 (0)	others
RCG	comp5477_c0_seq1	0 (0)	321.09 (136.57)	GH43	90.08 (52.24)	416.86 (103.44)	hemicellulases
	comp12022_c8_seq11	18.68 (4.86)	620.49	Polysaccharid	189.79	46.99 (56.68)	others

		(72.26)	e deacetylase	(43.64)		
comp12025_c3_seq4	88.18 (19)	1162.02 (177.77)	Polysaccharide deacetylase	67.97 (60)	737.81 (186.79)	others
comp11852_c3_seq1	1.54 (2.67)	7.71 (2.93)	Polysaccharide deacetylase	14.2 (13.38)	12.12 (8.51)	others
comp12029_c3_seq7	429.77 (47.12)	1524.72 (147.06)	GH8	198.35 (27.05)	927.07 (59.17)	cellulases
comp12029_c3_seq8	401.46 (114.81)	1603.45 (293.5)	GH8	198.35 (27.05)	927.07 (59.17)	cellulases
comp9619_c1_seq1	17.87 (4.28)	31.18 (4.74)	Polysaccharide deacetylase	278.42 (150.04)	668.12 (177.21)	others
comp11698_c1_seq1	4.53 (1.4)	9.19 (6.35)	GH16	1.68 (1.54)	7.91 (2.36)	cellulases
comp12029_c3_seq11	718.66 (96.24)	1393.08 (98.38)	GH48F	139.47 (139.59)	24.07 (21.38)	cellulases
comp9995_c0_seq1	662.76 (75.77)	1965.66 (145.59)	GH43	1051.56 (223.18)	3661.05 (145.54)	hemicellulases
comp12028_c11_seq4	11.88 (12.66)	185.61 (70.38)	GH10	195.75 (51.84)	562.41 (139.17)	hemicellulases
comp12022_c8_seq18	2.87 (4.97)	192.18 (15.08)	Polysaccharide deacetylase	11.46 (11.16)	69.46 (62.77)	others
comp12022_c8_seq7	3.5 (6.06)	55.65 (51.54)	Polysaccharide deacetylase	0 (0)	0 (0)	others
comp2377741_c0_seq1	0.78 (1.36)	7.58 (7.95)	Pectate lyase†	1.91 (1.92)	2.77 (2.41)	others
comp11304_c0_seq1	0.87 (1.51)	5.16 (5.21)	Esterase	1.65 (2.86)	1.46 (2.53)	others
comp10020_c2_seq1	0 (0)	6.3 (6.24)	NA	235.71 (30.59)	266.9 (39.72)	NA
comp12025_c3_seq3	584.63 (84.65)	683.81 (35.73)	Polysaccharide deacetylase	0 (0)	0 (0)	others

Avicel	comp7811_c0_seq1	36.89 (10.81)	1.62 (2.8)	GH5	8.65 (5.88)	29.58 (10.25)	cellulases
	comp7839_c2_seq1	6.5 (1.67)	1.33 (2.3)	Pectate lyase‡	23.48 (8.37)	5.96 (5.22)	others
	comp1619147_c0_seq1	304.32 (105.67)	121.85 (65.09)	GH3	1862.49 (42.26)	805.66 (86.82)	cellulases
	comp11629_c2_seq1	26.31 (5.15)	7.49 (5.43)	GH45	44.67 (11.89)	12.74 (2.45)	cellulases
	comp12028_c8_seq21	4.61 (7.2)	0 (0)	GH48F	18.45 (17.96)	523.71 (90.11)	cellulases

†Low expression pectate lyase
‡High expression pectate lyase

Data S1
Annotated transcripts observed in *Piromyces finnis*

Data S2
Annotated transcripts observed in *Neocallimastix californiae*

Data S3
Annotated transcripts observed in *Anaeromyces robustus*

Data S4
Gene ontology terms significantly enriched with non-coding antisense transcripts as determined by a hypergeometric distribution ($p<0.05$)

Data S5
Proteome of *Anaeromyces robustus* mapped to transcripts

Data S6
Proteome of *Neocallimastix californiae* mapped to transcripts

Data S7
Proteome of *Piromyces finnis* mapped to transcripts

Data S8
Clusters of transcripts that are differentially expressed in *P. finnis* and their expression profile in response to a glucose pulse as determined by hierarchical clustering.

Data S9
Clusters of transcripts that are differentially expressed in *P. finnis* and their expression profile on different substrates as determined by hierarchical clustering.

Data S10
Statistical significance of gene ontology terms enriched in *P. finnis* under growth various substrates (hypergeometric distribution)

References and Notes

1. K. Sanderson, Lignocellulose: A chewy problem. *Nature* **474**, S12–S14 (2011). Medline doi:10.1038/474S012a

2. D. R. Dodds, R. A. Gross, Chemicals from biomass. *Science* **318**, 1250–1251 (2007). Medline doi:10.1126/science.1146356

3. A. Berlin, M. Balakshin, N. Gilkes, J. Kadla, V. Maximenko, S. Kubo, J. Saddler, Inhibition of cellulase, xylanase and beta-glucosidase activities by softwood lignin preparations. *J. Biotechnol.* **125**, 198–209 (2006). Medline doi:10.1016/j.jbiotec.2006.02.021

4. M. Schülein, in *Methods in Enzymology*, S. T. K. Willis, A. Wood, Ed. (Academic Press, 1988), vol. 160 of *Biomass Part A: Cellulose and Hemicellulose*, pp. 234–242.

5. M. Hess, A. Sczyrba, R. Egan, T. W. Kim, H. Chokhawala, G. Schroth, S. Luo, D. S. Clark, F. Chen, T. Zhang, R. I. Mackie, L. A. Pennacchio, S. G. Tringe, A. Visel, T. Woyke, Z. Wang, E. M. Rubin, Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**, 463–467 (2011). Medline doi:10.1126/science.1200387

6. M. J. Nicholson, M. K. Theodorou, J. L. Brookman, Molecular analysis of the anaerobic rumen fungus *Orpinomyces* - insights into an AT-rich genome. *Microbiology* **151**, 121–133 (2005). Medline doi:10.1099/mic.0.27353-0

7. C. G. Orpin, The rumen flagellate *Piromonas communis*: Its life-history and invasion of plant material in the rumen. *J. Gen. Microbiol.* **99**, 107–117 (1977). Medline doi:10.1099/00221287-99-1-107

8. M. K. Theodorou, G. Mennim, D. R. Davies, W. Y. Zhu, A. P. Trinci, J. L. Brookman, Anaerobic fungi in the digestive tract of mammalian herbivores and their potential for exploitation. *Proc. Nutr. Soc.* **55**, 913–926 (1996). Medline doi:10.1079/PNS19960088

9. N. H. Youssef, M. B. Couger, C. G. Struchtemeyer, A. S. Liggenstoffer, R. A. Prade, F. Z. Najar, H. K. Atiyeh, M. R. Wilkins, M. S. Elshahed, The genome of the anaerobic fungus *Orpinomyces* sp. strain C1A reveals the unique evolutionary history of a remarkable plant biomass degrader. *Appl. Environ. Microbiol.* **79**, 4620–4634 (2013). Medline doi:10.1128/AEM.00821-13

10. C. H. Haitjema, K. V. Solomon, J. K. Henske, M. K. Theodorou, M. A. O'Malley, Anaerobic gut fungi: Advances in isolation, culture, and cellulolytic enzyme discovery for biofuel production. *Biotechnol. Bioeng.* **111**, 1471–1482 (2014). Medline doi:10.1002/bit.25264

11. D. S. Tuckwell, M. J. Nicholson, C. S. McSweeney, M. K. Theodorou, J. L. Brookman, The rapid assignment of ruminal fungi to presumptive genera using ITS1 and ITS2 RNA secondary structures to produce group-specific fingerprints. *Microbiology* **151**, 1557–1567 (2005). Medline doi:10.1099/mic.0.27689-0

12. S. Lavergne, J. Molofsky, Reed canary grass (*Phalaris arundinacea*) as a biological model in the study of plant invasions. *Crit. Rev. Plant Sci.* **23**, 415–429 (2004). doi:10.1080/07352680490505934

13. V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014). Medline doi:10.1093/nar/gkt1178

14. S. Raghothama, R. Y. Eberhardt, P. Simpson, D. Wigelsworth, P. White, G. P. Hazlewood, T. Nagy, H. J. Gilbert, M. P. Williamson, Characterization of a cellulosome dockerin domain from the anaerobic fungus Piromyces equi. *Nat. Struct. Mol. Biol.* **8**, 775–778 (2001). doi:10.1038/nsb0901-775

15. M. E. Himmel, S. Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady, T. D. Foust, Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* **315**, 804–807 (2007). Medline doi:10.1126/science.1137016

16. V. Lionetti, F. Francocci, S. Ferrari, C. Volpi, D. Bellincampi, R. Galletti, R. D'Ovidio, G. De Lorenzo, F. Cervone, Engineering the cell wall by reducing de-methyl-esterified homogalacturonan improves saccharification of plant tissues for bioconversion. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 616–621 (2010). Medline doi:10.1073/pnas.0907549107

17. E. J. Finehout, K. H. Lee, Comparison of automated in-gel digest methods for femtomole level samples. *Electrophoresis* **24**, 3508–3516 (2003). Medline doi:10.1002/elps.200305615

18. C. Kramer, J. J. Loros, J. C. Dunlap, S. K. Crosthwaite, Role for antisense RNA in regulating circadian clock function in Neurospora crassa. *Nature* **421**, 948–952 (2003). Medline doi:10.1038/nature01427

19. M. Yassour, J. Pfiffner, J. Z. Levin, X. Adiconis, A. Gnirke, C. Nusbaum, D. A. Thompson, N. Friedman, A. Regev, Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome Biol.* **11**, R87 (2010). Medline doi:10.1186/gb-2010-11-8-r87

20. M. E. Donaldson, B. J. Saville, Natural antisense transcripts in fungi. *Mol. Microbiol.* **85**, 405–417 (2012). Medline doi:10.1111/j.1365-2958.2012.08125.x

21. B. C. King, M. K. Donnelly, G. C. Bergstrom, L. P. Walker, D. M. Gibson, An optimized microplate assay system for quantitative evaluation of plant cell wall-degrading enzyme activity of fungal culture extracts. *Biotechnol. Bioeng.* **102**, 1033–1044 (2009). Medline doi:10.1002/bit.22151

22. M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863–14868 (1998). Medline doi:10.1073/pnas.95.25.14863

23. U. Bakir, S. Yavascaoglu, F. Guvenc, A. Ersayin, An endo-β-1,4-xylanase from *Rhizopus oryzae*: Production, partial purification and biochemical characterization. *Enzyme Microb. Technol.* **29**, 328–334 (2001). doi:10.1016/S0141-0229(01)00379-9

24. C. E. Todero Ritter, M. Camassola, D. Zampieri, M. M. Silveira, A. J. P. Dillon, Cellulase and xylanase production by *Penicillium echinulatum* in submerged media containing cellulose amended with sorbitol. *Enzyme Res.* **2013**, e240219 (2013). doi:10.1155/2013/240219

25. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). Medline doi:10.1073/pnas.0506580102

26. N. L. Glass, M. Schmoll, J. H. D. Cate, S. Coradetti, Plant cell wall deconstruction by ascomycete fungi. *Annu. Rev. Microbiol.* **67**, 477–498 (2013). Medline doi:10.1146/annurev-micro-092611-150044

27. V. Pelechano, L. M. Steinmetz, Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14**, 880–893 (2013). Medline doi:10.1038/nrg3594

28. K. V. Solomon, C. H. Haitjema, D. A. Thompson, M. A. O'Malley, Extracting data from the muck: Deriving biological insight from complex microbial communities and non-model organisms with next generation sequencing. *Curr. Opin. Biotechnol.* **28**, 103–110 (2014). Medline doi:10.1016/j.copbio.2014.01.007

29. R. Edgar, M. Domrachev, A. E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002). Medline doi:10.1093/nar/30.1.207

30. 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, Dordrecht, Netherlands, 2005), pp. 55–66.

31. K. V. Solomon, J. K. Henske, M. K. Theodorou, M. A. O'Malley, Robust and effective methodologies for cryopreservation and DNA extraction from anaerobic gut fungi. *Anaerobe* 10.1016/j.anaerobe.2015.11.008 (2015). Medline doi:10.1016/j.anaerobe.2015.11.008

32. D. Parkhomchuk, T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, H. Lehrach, A. Soldatov, Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* **37**, e123 (2009). Medline doi:10.1093/nar/gkp596

33. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011). Medline doi:10.1038/nbt.1883

34. J. Martin, V. M. Bruno, Z. Fang, X. Meng, M. Blow, T. Zhang, G. Sherlock, M. Snyder, Z. Wang, Rnnotator: An automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics* **11**, 663 (2010). Medline doi:10.1186/1471-2164-11-663

35. S. Götz, J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talón, J. Dopazo, A. Conesa, High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435 (2008). Medline doi:10.1093/nar/gkn176

36. F. Chen, A. J. Mackey, C. J. Stoeckert Jr., D. S. Roos, OrthoMCL-DB: Querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.* **34**, D363–D368 (2006). Medline doi:10.1093/nar/gkj123

37. I. V. Grigoriev, R. Nikitin, S. Haridas, A. Kuo, R. Ohm, R. Otillar, R. Riley, A. Salamov, X. Zhao, F. Korzeniewski, T. Smirnova, H. Nordberg, I. Dubchak, I. Shabalov, MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42**, D699–D704 (2014). Medline doi:10.1093/nar/gkt1183

38. M. W. Duncan, R. Aebersold, R. M. Caprioli, The pros and cons of peptide-centric proteomics. *Nat. Biotechnol.* **28**, 659–664 (2010). Medline doi:10.1038/nbt0710-659

39. M. K. Theodorou, B. A. Williams, M. S. Dhanoa, A. B. McAllan, J. France, A simple gas production method using a pressure transducer to determine the fermentation kinetics of ruminant feeds. *Anim. Feed Sci. Technol.* **48**, 185–197 (1994). doi:10.1016/0377-8401(94)90171-6

40. T. M. Wood, in *Methods in Enzymology*, S. T. K. Willis, A. Wood, Ed. (Academic Press, 1988; www.sciencedirect.com/science/article/pii/0076687988601030), vol. 160 of *Biomass Part A: Cellulose and Hemicellulose*, pp. 19–25.

41. S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010). doi:10.1186/gb-2010-11-10-r106

42. H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley, R. Lopez, Analysis tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41**, W597–W600 (2013). Medline doi:10.1093/nar/gkt376

43. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007). Medline doi:10.1093/bioinformatics/btl529