Title

Catabolic repression in early-diverging anaerobic fungi is partially mediated by natural antisense transcripts

Authors

Solomon, Kevin V
Henske, John K
Gilmore, Sean P
et al.

Publication Date

2018-12-01

DOI

10.1016/j.fgb.2018.09.004

Peer reviewed

Catabolic repression in early-diverging anaerobic fungi is partially mediated by natural antisense transcripts

Kevin V. Solomon[1, †], John K. Henske[1], Sean P. Gilmore[1], Anna Lipzen[2], Igor V. Grigoriev[2-3], Dawn Thompson[4], Michelle A. O'Malley[1*]

[1]Department of Chemical Engineering, University of California, Santa Barbara, Santa Barbara, CA 93106, United States
[2]U.S. Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
[3]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[4]Ginkgo Bioworks, Boston, MA 02210, USA.

Author emails: kvs@purdue.edu, john.henske@gmail.com, seangilmore@umail.ucsb.edu, alipzen@lbl.gov, ivgrigoriev@lbl.gov, dawnt@ginkgobioworks.com, momalley@ucsb.edu

[†]Present Address: Department of Agricultural & Biological Engineering, Purdue University, West Lafayette, IN 47907

[*]Corresponding author: momalley@ucsb.edu

Running title: Catabolic regulation via antisense in gut fungi

Keywords: antisense; RNA; anaerobic fungi; catabolic regulation

Abstract

Early-diverging anaerobic fungi (order: Neocallimastigomycota), lignocelluolytic chytrid-like fungi central to fiber degradation in the digestive tracts of large herbivores, are attractive sources of cellulases and hemicellulases for biotechnology. Enzyme expression is tightly regulated and coordinated through mechanisms that remain unelucidated to optimize hydrolytic efficiency. Our analysis of anaerobic fungal transcriptomes reveals hundreds of *cis*-natural antisense transcripts (*cis*-NATs), which we hypothesize play an integral role in this regulation. Through integrated genomic and transcriptomic sequencing on a range of catabolic substrates, we validate these NATs in three species (*Anaeromyces robustus, Neocallimasix californiae*, and *Piromyces finnis*), and analyze their expression patterns and prevalence to gain insight into their function. NAT function was diverse and conserved across the three fungal genomes studied, with 10% of all metabolic process NATs associated with lignocellulose hydrolysis. Despite these similarities, however, only eleven gene targets were conserved orthologs. Several NATs were dynamically regulated by lignocellulosic substrates while their gene targets were unregulated. This observation is consistent with a hypothesized, but untested, regulatory mechanism where selected genes are exclusively regulated at the transcriptional/post-transcriptional level by NATs. However, only genes with high NAT relative expression levels displayed this phenomenon, suggesting a selection mechanism that favors larger dynamic ranges for more precise control of gene expression. In addition to this mode, we observed two other possible regulatory fates: canonical transcriptional regulation with no NAT response, and positive co-regulation of target mRNA and cognate NAT, which we hypothesize is a fine-tuning strategy to locally negate control outputs from global regulators. Our work reveals the complex contributions of antisense RNA to the catabolic response in anaerobic fungi, highlighting its importance in understanding lignocellulolytic activity for bioenergy applications. More importantly, the relative expression of NAT to target may form a critical determinant of transcriptional vs post-transcriptional (NAT) control of gene expression in primitive anaerobic fungi.

67

Background

Antisense transcription, or transcription of RNA from the noncoding strand of a gene, is a key mode of gene regulation in all domains of life (Britto-Kido et al., 2013; Donaldson and Saville, 2012; Katayama et al., 2005; Wagner and Simons, 1994). The resulting RNA exhibits extended sequence complementarity to a 'sense' RNA transcript or target mRNA that it interacts with to control gene expression. Transcripts that originate from the same genomic locus as the regulated target are known as *cis*-antisense transcripts, while those that arise from a distinct locus, with potentially imperfect complementarity, are known as *trans*-antisense transcripts (Donaldson and Saville, 2012). Both cis- and trans- antisense transcripts may regulate gene expression of the target mRNA through a number of mechanisms including induced chromatin remodeling, differential mRNA splicing/translation via sequence masking, and induced mRNA instability via RNA:RNA duplex formation (Donaldson and Saville, 2012; Faghihi and Wahlestedt, 2009; Hastings et al., 2000). Further, due to their positioning on the same locus as their target, cis- transcripts may also regulate gene expression transcriptionally through opposing RNA polymerase collision and transcriptional interference (Donaldson and Saville, 2012). While these mechanisms typically downregulate expression, they may at times, counterintuitively, enhance protein translation (Carrieri et al., 2012; Jabnoune et al., 2013). Regardless of mechanism, however, transcriptional/post-transcriptional regulation via antisense is nominally more responsive than traditional transcriptional control and adapts more rapidly to evolutionary pressures (Pelechano and Steinmetz, 2013; Yan and Wang, 2012). More importantly, antisense transcription can reduce noise in gene expression and achieve more switch-like behavior for precise control (Pelechano and Steinmetz, 2013). For these reasons, natural antisense transcripts (NATs) have formed a pervasive feature of fungal genomes with critical roles in cellular development (Ohm et al., 2010; Yassour et al., 2010), plant biomass degradation ("Complex regulation of hydrolytic enzyme genes for cellulosic biomass degradation in filamentous fungi - Springer," n.d.; Ries et al., 2013; Solomon et al., 2016a), and stress

adaptation (Donaldson and Saville, 2012) among other processes (Arthanari et al., 2014; Ni et al., 2010; Smith et al., 2008).

Early-diverging anaerobic gut fungi were recently shown to have a diverse array of cellulases that are tightly programmed for synergistic degradation of lignocellulosic biomass (Solomon et al., 2016a). The regulatory program of these enzymes is essential for function. However, the role of noncoding putative NATs, in tuning this regulatory response has not been established. Many outstanding questions remain, including how conserved these NATs are among anaerobic gut fungi and their specific loci of origin. More importantly, it is unknown whether biomass-degrading enzymes in early-branching fungi are solely regulated by NATs in the absence of other transcriptional changes, as has been seen in later-diverging fungi (Ni et al., 2010). Thus, characterizing the behavior and properties of gut fungal NATs is needed to fully understand the regulatory program of these enzymes, and their resulting impact on cellular behavior.

Given the importance of catabolic repression for efficient biomass degradation (Solomon et al., 2016a), we sought to investigate the NAT contributions to this transcriptional response. Here, we integrate genomic sequencing and strand-specific RNA-Seq to characterize the role of NATs in anaerobic fungal biology. We annotate NAT candidates, target genes, and characterize their properties to validate and identify the function of more than 2000 NATs. We also measure the differential expression of these NATs and their cognate targets on various substrates to evaluate their function in fungal catabolic regulation. Our results reveal dozens of genes that are solely regulated post-transcriptionally by NATs, and reveal conditions where post-transcriptional regulation via NAT is preferred over direct transcriptional control.

Results and Discussion

Gut fungal genomes contain a diverse landscape of cis-natural antisense transcripts

Recent strand-specific RNA-Seq of anaerobic fungi revealed a number of ncRNA hypothesized to be antisense RNA, as determined by their gene annotations (Solomon et al., 2016a). Transcripts that are unannotated by this pipeline cannot be assigned a directional sense and are ignored in these analyses. To confirm the antisense predictions, we aligned these transcripts to their respective genomes (Haitjema et al., 2017) (Table 1) and determined that their loci were indeed likely to be from the non-template or antisense strand of a coding gene. We restricted our NAT candidates to those with perfect complementarity for at least 80% of the overlap with putative targets to eliminate partially mapped transcripts, poor matches, and potential *trans*- NATs as these fungal genomes are unclosed and the relative positions of many loci remain uncertain. This high degree of complementarity also excludes RNAi transcripts as it filters out the sequences needed for processing via Dicer encoded within these fungal genomes (Bernstein et al., 2001; Solomon et al., 2016a). Additionally, we filtered potential transcript assembly fusion artifacts while still accommodating potential introns by restricting the length of NAT overlap to no more than 3 times the length of its target. The strict filters still resulted in >93% of all *de novo* transcripts mapping to the genome in at least one location (Table 1). While only 20-40% of the antisense transcript candidates could be validated as true NATs (Figure 1A, Additional Files S1-S3, see Solomon et al. 2016a for annotations), they represented 3-9% of all observed transcripts (Table 1). This relative NAT abundance, while significant, is relatively low for fungi (Goodman et al., 2013; Ni et al., 2010; Ohm et al., 2010), and inconsistent with the well-established ‘gene number paradox’ of non-coding DNA increasing with organismal complexity or genome size in eukaryotes (Taft et al., 2007). However, this may be partially explained by the unique evolutionary history of these fungi, which includes significant horizontal gene transfer with rumen bacteria (Haitjema et al., 2017; Rosewich and Kistler, 2000). The gut fungi exhibit an inverse correlation between genome size and number of non-coding transcripts. Smaller, more gene-dense genomes contain a higher relative abundance of NATs as seen in *Piromyces* (56 MB - 9% NATs) vs *Neocallimastix* (193 MB - 2.5% NATs). Given the regulatory role of NATs, increasing the density of NATs to

complement a reduced number of regulatory proteins may be a mechanism to increase genome density while maintaining organismal complexity.

Gut fungal NATs are fairly long, ranging in size from 100 bp to 6 kb with a median length of ~250 bp (~10% of sense transcript, Figure 1B). These NATs are transcribed from independent promoters leading to one of several possible binding configurations: head-to-head overlap; tail-to-tail overlap; embedded NATs; and full complementarity (Figure 1A). However, the distribution of transcripts between these forms is non-random; like later diverging ascomycetes, earlier branching gut fungi favor tail-to-tail NATs that arise from convergent transcription (Figure 1C) (Donaldson and Saville, 2012). This bias in hybridization target suggests that transcriptional interference via polymerase collision may be a primary mode of action for these transcripts. Other possible mechanisms such as differential splicing due to RNA masking are also supported by the data. The presence of NATs increases the number of transcriptional isoforms for a predicted gene locus by at least 50% (p-value $< 10^{-20}$, unpaired t-test with unequal variance – Additional File S23). However, transcriptional regulation via chromatin remodeling is unlikely as the NATs do not hybridize near to the transcriptional start site in a manner that might directly alter the local chromatin structure around the promoter and affect gene transcriptional efficiency (Nocetti and Whitehouse, 2016).

NATs are implicated in many cellular processes such as fungal development

Having established the prevalence of NATs within the gut fungal genomes, we then investigated their functional roles. We identified the complementary target transcript for each NAT from our genome maps and used the more complete target annotations to describe NAT function. With this approach, we assigned a putative function to just under half the NATs from each species. These Gene Ontology (GO) terms were binned into unique functional processes ranging from metabolism to cellular transport to assign functional roles to the NATs. Despite the differences in abundance, the distribution of NAT function is

conserved at the process level across all three genera (Figure 2). Unsurprisingly, the distribution of NAT functionality reflected that of the transcriptome as a whole (Figure S1) with regulation of metabolism and protein expression as the dominant roles for NATs. 13.9 – 21.0% of NATs targeting metabolic processes were implicated in lignocellulose hydrolysis. However, the relative abundance of lignocellulose hydrolysis NATs relative to metabolic process NATs was enriched by at least 25%, suggesting the importance of antisense regulation to the cellulolytic lifestyle of anaerobic fungi. These NATs were unbiased toward specific enzyme family targets, putatively hybridizing with transcripts for secreted cellulosomal and non-cellulosomal components with functions as cellulases, hemicellulases, acetylxylan esterases, and pectinases.

Despite the broad conservation of NAT functional roles at the process level across the 3 fungi examined here (Figure 2), each fungus was uniquely enriched in selected GO terms (Figure 3). A hypergeometric test of NAT abundance relative to the transcriptome as a whole revealed 10-14 GO terms per fungus enriched with NATs at $p \leq 0.01$ [Table S1 & Additional Files S4-S6]. None of these GO terms is similarly enriched across all three fungi and may reflect the relative importance of key cellular processes to fungal lifestyles across different genera. However, conserved functions that define the gut fungal division may emerge as more species are isolated and sequenced. As NATs are typically regulatory in nature, their presence and activity are important predictors of cellular phenotype. Coupled with their evolutionary plasticity, or ability to evolve quickly, these key differences in NAT functional enrichment may, in part, drive the differentiation of gut fungal genera. For example, NATs that target sporulation processes (GO Term: sporulation resulting in formation of a cellular spore, (Borneman et al., 1989)) are absent in both monocentric *Neocallimastix* and *Piromyces*, while being present and enriched in high zoospore producing polycentric *Anaeromyces* ($p = 0.015$). This regulation of cellular development by antisense is a conserved feature across many fungi (Kramer et al., 2003; Ohm et al., 2010; Yassour et al.,

2010) and other eukaryotes (Piatek et al., 2016; Serobyan et al., 2016), suggesting that such mechanisms are simple yet effective at regulating complex life cycles.

In spite of the challenges in computationally identifying gene orthologs (Koonin, 2005), we were able to identify more than a dozen shared, possibly ancestral, ortholog targets conserved among the gut fungi. *Piromyces* and *Neocalliomastix*, both monocentric fungi (Additional File S7) (Li et al., 2016), share the most NAT ortholog targets (8) despite the large differences in their genome sizes. Among these targets are spore coat proteins (CotH), which have been associated with an extracellular cellulolytic complex or cellulosome (Dassa et al., 2012; Solomon et al., 2016a), motor protein dynein, and a FMP27 transcription factor (O'Sullivan et al., 2004). In contrast, monocentric *Piromyces* and polycentric *Anaeromyces*, and *Neocallimastix* and *Anaeromyces* share only 5 and 1 targets, respectively. These targets regulate more essential processes such as ribosomal biogenesis (LSM proteins) and environmental sensing (DUF221) (Hou et al., 2014, p. 221; Wilusz and Wilusz, 2013). The degree of ortholog conservation between species is inversely correlated with phylogenetic distance (Li et al., 2016), increasing confidence in our analysis, and hints at the evolutionary history of these regulatory transcripts. However, this dataset may be subject to undersampling that limits the amount of observable sequence observation, due to difficulties associated with low genomic GC content, tightly associated macromolecules of DNA packaging, and complex life cycles, which all limit the yield of high-quality intact DNA needed for genomic analysis (Solomon et al., 2016b)

NATs exclusively regulate genes in response to changes in carbon source

Given the abundance of NAT transcripts, we then sought to quantify their expression profiles to assess their role in the regulation of lignocellulose hydrolysis – a primary function of anaerobic fungi. Exploiting catabolite repression in anaerobic fungi (Solomon et al., 2016a), we identified NATs and/or their targets that exhibited differential expression when grown on cellobiose, Avicel, or reed canary grass as compared to expression on glucose (Additional File S8-16). At least 30 unique NATs and 69 NAT targets for each species displayed

differential regulation (≥2-fold change in expression, $p_{val} \leq 0.01$) under the growth substrates tested (Figure 4A-C, Additional Files S17-S22 – expression fold change relative to glucose) with a functional distribution similar to that of all NATs. Among the regulated transcripts were a small subset (8-17%) where both a NAT and its target transcript were positively co-regulated with similar fold-changes in expression. These increases in target and NAT transcription, which could functionally oppose each other through polymerase collision (Figure 1C) (Pelechano and Steinmetz, 2013), may effectively cancel out and are anticipated to only minimally change protein expression (Figure 4D). Most regulated transcripts, however, displayed only direct differential expression of the target or the corresponding NAT (Figure 4A-C). This result is consistent with the possibility that most genes were exclusively regulated by either NAT expression or target transcriptional regulation. While sole transcriptional regulation is far more common, as is expected, the independent NAT expression patterns highlight the unique regulatory roles of these NATs that respond to stimuli distinct to that of their target transcript. More importantly, it underscores a need to capture and study NAT regulation to fully understand phenotypes. This exclusivity has been similarly observed in other fungi (Ni et al., 2010) and may have arisen as a method to fine tune the programming of individual transcripts without disrupting coordinated expression via global transcription factors.

If the observed changes in NAT expression directly affect protein expression, the decision to independently regulate protein expression via antisense or transcriptional control is determined by relative expression of the target and NAT transcripts (Figure 5). When NAT expression far exceeded that of its cognate target on both glucose and an alternative substrate, protein expression was regulated exclusively by the NAT. Conversely, transcriptional control occurred when target expression was dominant. Similarly, genes are co-regulated when NAT and target have similar expression levels. These trends are consistent with a model where NATs fine tune gene expression. Cells are able to maximize the dynamic range of possible expression outputs by varying the expression of the dominant

transcript (Pelechano and Steinmetz, 2013). This ability to effect more robust control supersedes the apparent cellular function of the transcript as a mechanism for selection as there was no observable correlation of regulatory fate and NAT function (Additional Files S24-S32 – GO term annotation of antisense targets for a given regulatory fate in each species) via enrichment analysis.

Conclusions

In this study, we characterize the role of antisense transcription in the understudied early-branching *Neocallimastigomycota* fungi. While pervasive throughout the three fungal genomes studied, overall NAT abundance was low compared to other fungi, potentially reflecting the early divergence of anaerobic fungi and their significant horizontal gene transfer with bacteria. These NATs were implicated in diverse processes and hypothesized to primarily regulate its target through a polymerase collision mechanism. NATs showed differential expression on growth substrates with several gene targets displaying no other form of transcriptional regulation. That is, genes were potentially independently regulated by their NATs alone. This fate between post-transcriptional (NAT) and transcriptional control of gene expression was correlated with NAT expression level suggesting a possible selection mechanism. Thus, NAT profiling is critical to fully recapitulate cellular phenotypes such as plant biomass degradation, and reveal efficient strategies of enzyme regulation needed for more economical biofuels. In sum, our work refines the understanding of gene regulation in lignocellulolytic anaerobic fungi and suggests potential determinants of transcriptional vs. post-transcriptional control.

Materials & Methods:

Strains and Culture Maintenance

Three early-branching fungal isolates of *Neocallimastigomycota* (Li et al., 2016), *Anaeromyces robustus, Neocallimastix californiae*, and *Piromyces finnis*, were grown and

continually passaged in 10 mL Medium C supplemented with 15% cow rumen fluid (Bar Diamond, Parma, ID) and up to 0.5% of a carbon substrate (Theodorou et al., 2005). Substrates included glucose (anhydrous, Thermo Fisher Scientific, Canoga Park, CA), cellobiose (Sigma-Aldrich, St Louis, MO), Avicel (PH-101, 50 µm particle size, Sigma-Aldrich), and reed canary grass gifted from the USDA-ARS Research Center (Madison, Wisconsin). We grew at least 3 cultures of each isolate per substrate before pelleting them at mid-exponential phase (P ~ 4 psig) (Haitjema et al., 2014). We then stored these cell pellets in 1 ml RNALater (Qiagen, Valencia, CA) at -80°C to preserve their transcriptomic profiles for later analysis. To sequence the genomes of these organisms, we grew multiple 40 mL cultures of each isolate supplemented with glucose as a carbon source. Cells were pelleted at mid-exponential phase and their DNA immediately extracted.

RNA Isolation and Transcriptome Acquisition

We isolated the RNA content from the preserved culture pellets within a week of harvesting. Pellets were thawed on ice before we extracted the total RNA using the Qiagen RNEasy Mini Kit (Qiagen) according to the manufacturer's instructions for Plants and Fungi. All samples were treated with an on-column DNaseI treatment (Qiagen) at room temperature for 15 minutes, per the manufacturer's recommendations, to remove any contaminating DNA.

We measured the total RNA content of each sample using a NanoDrop 2000 (ThermoScientific, Waltham MA) to ensure a minimum concentration of 60 ng/µL of RNA and validate sample purity ($A_{260}/A_{280} \geq 1.8$). We then verified their integrity via an RNA Integrity (RIN) score with a TapeStation (Agilent Technologies, Santa Clara, CA) or BioAnalyzer (Agilent Technologies, Santa Clara, CA). We submitted at least 3 µg RNA to the Dept. of Energy Joint Genome Institute (JGI, Walnut Creek, CA) for sequencing. To acquire a full catalog of all possible transcripts, we pooled equal masses of RNA obtained from growth on various substrates and submitted one sample per isolate for sequencing to the JGI (Solomon et al., 2016a). The transcriptional profile as function of substrate were captured from replicate samples and sequenced on a HiSeq2500 (Illumina, San Diego, CA) at the Broad

Institute of MIT and Harvard (*Piromyces*) (Solomon et al., 2016a), and on a NextSeq500 (Illumina) at the UCSB Biological Nanostructures Laboratory (*Anaeromyces* and *Neocallimastix*). Libraries were prepared either using polyA purification (*Piromyces*) (Solomon et al., 2016a) or rRNA depletion (*Anaeromyces,* and *Neocallimastix*) (Henske et al., 2018). Samples from cultures grown on all substrates were sequenced in triplicate with the exceptions of maltose and corn stover for *Anaeromyces,* and switchgrass for *Neocallimastix,* which were completed in duplicate due to sample loss in the processing pipeline. All transcriptional reads were mapped to the JGI assembled transcriptomes before being analyzed for gene expression profiles (Haitjema et al., 2017). In all cases, strand-specific transcriptional libraries were generated using a dUTP protocol (Levin et al., 2010).

Antisense RNA Identification and Annotation

We annotated all RNA transcripts with the BLAST2GO pipeline (Götz et al., 2008) and noted the orientation of annotation (Solomon et al., 2016a). As these transcriptomes are generated from a strand-specific library, any annotation in a negative reading frame (-1, -2, -3) flags a transcript as a candidate antisense RNA. The transcriptomes were then mapped to the corresponding genome using GMAP (Wu and Watanabe, 2005), with a strict cutoff of > 80% complementarity and a mapping length no greater than 3 times the transcript length. From this list of verified mappings, we then identified validated antisense candidates that mapped to a target transcript as cis-natural antisense transcripts (NATs) whose function are given by the annotations of their target mRNA (Solomon et al., 2016a). Included in these annotations are BLAST homologs, InterPro protein domain identities, Gene Ontology (GO) Term descriptions, Enzyme Commission Numbers, and mapping to ortholog groups (OrthoMCL (Chen et al., 2006)).

Differential Expression Analysis and Expression Clustering

We analyzed the transcriptomic profiles for each isolate as a function of substrate using RSEM and DESeq2. First, we estimated the counts of each transcript from the raw reads and

assembled transcriptomes (Solomon et al., 2016a) using the RSEM analysis script in the Trinity toolkit. These estimated counts were then analyzed using the Bioconductor DESeq2 package (version 1.8.2) in the R programming language with default parameters (Anders and Huber, 2010) to identify differential expression patterns relative to growth on glucose. We filtered these results for biological significance (absolute $\log_2$-fold change in expression ≥ 1) and statistical significance with adjusted p-values ≤ 0.01.

Functional binning of NATs

We assigned putative functions to NATs from the GO Term annotations of its putative target. Functional bins included GO terms that incorporated the following terms: Cytoskeleton & Mobility – 'dynein', 'actin', 'chitin,' or 'cilium'; Signaling & Sensing – 'receptor' or 'signal'; Protein Expression – 'ribosom', 'transcrip', 'translat', 'peptidyl-prolyl cis-trans isomerase', 'proteasome', 'mRNA', 'proteo', 'ubiquitin', 'folding', 'tRNA', 'protein dimerization', 'protein glycosylation', or 'rRNA'; Transport – 'transport'; Lignocellulose hydrolysis – 'cellulose', 'xylan', 'pectin' or 'cellulase'; Metabolism – 'metabolic' and 'catabolic'; Other catalytic processes –'ase' or 'catalytic'; Other – all remaining GO term annotations. Each NAT was assigned to single function in a precedence order. NATs that could be assigned to multiple functions were grouped with the function of highest precedence in the order: Cytoskeleton & Mobility, Signaling & Sensing, Protein Expression, Transport, Lignocellulose hydrolysis, Metabolism, Other catalytic processes, and finally Other.

Functional Enrichment Analysis of NATs

We tested for functional enrichment of NATs within GO Terms via a hypergeometric statistical test to establish putative roles where the presence and/or regulation of NATs is critical. We calculated the fraction of NATs with a given GO term and compared this fraction to that of the entire transcriptome with that GO term to establish a probability (p-value) that this distribution is random and reflects the innate transcriptome distribution (null

hypothesis). GO Terms with a p-value ≤ 0.01 are said to be enriched in NATs and used to infer biological significance.

List of Abbreviations

GO	Gene Ontology
ncRNA	non-coding RNA
NAT	natural antisense transcript

DECLARATIONS

Ethics approval and consent to participate. Not applicable

Consent for publication. Not applicable

Availability of data and materials

Transcriptomic short read and expression data have been uploaded to the NCBI Sequence Read Archive (SRA: http://www.ncbi.nlm.nih.gov/sra/) and Gene Expression Omnibus (GEO: https://www.ncbi.nlm.nih.gov/geo/) under accession numbers: GSE64834 (Solomon et al., 2016a) and GSE95479. Genomes are available at Mycocosm (Grigoriev et al., 2014) (http://genome.jgi.doe.gov/programs/fungi/index.jsf) under taxon IDs Anasp1, Neosp1, and Pirfi3.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors gratefully acknowledge funding support from the Office of Science (BER), U.S. Department of Energy (DE-SC0010352), the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office, and the National Science Foundation (MCB-1553721). SPG is supported by the NSF GRFP (DGE-1144085). A

portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) exploratory effort 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). Funding sources had no role in the design of the study and collection, analysis, and interpretation of data and in writing.

Authors' contributions

KVS and JKH extracted the RNA samples for analysis. AL sequenced and assembled the transcriptomes. KVS, JKH, and SPG annotated the transcriptome, analyzed expression data and mapped to corresponding genomic targets. KVS, IVG, DT, and MAO wrote the manuscript. MAO and KVS designed the study.

Acknowledgements

We thank Paul Weimer from the United States Department of Agriculture (USDA) for providing freshly milled reed canary grass biomass substrates.

References

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, 1–12. https://doi.org/10.1186/gb-2010-11-10-r106

Arthanari, Y., Heintzen, C., Griffiths-Jones, S., Crosthwaite, S.K., 2014. Natural Antisense Transcripts and Long Non-Coding RNA in Neurospora crassa. PLoS ONE 9, e91353. https://doi.org/10.1371/journal.pone.0091353

Bernstein, E., Caudy, A.A., Hammond, S.M., Hannon, G.J., 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409, 363–366. https://doi.org/10.1038/35053110

Borneman, W.S., Akin, D.E., Ljungdahl, L.G., 1989. Fermentation products and plant cell wall-degrading enzymes produced by monocentric and polycentric anaerobic ruminal fungi. Appl. Environ. Microbiol. 55, 1066–1073.

Britto-Kido, S. de A., Ferreira Neto, J.R.C., Pandolfi, V., Marcelino-Guimarães, F.C., Nepomuceno, A.L., Vilela Abdelnoor, R., Benko-Iseppon, A.M., Kido, E.A., 2013. Natural Antisense Transcripts in Plants: A Review and Identification in Soybean Infected with Phakopsora pachyrhizi SuperSAGE Library. Sci. World J. 2013, 219798. https://doi.org/10.1155/2013/219798

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., Forrest, A.R.R., Carninci, P., Biffo, S., Stupka, E., Gustincich, S., 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature 491, 454–457. https://doi.org/10.1038/nature11508

Chen, F., Mackey, A.J., Stoeckert, C.J., Roos, D.S., 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 34, D363–D368. https://doi.org/10.1093/nar/gkj123

Complex regulation of hydrolytic enzyme genes for cellulosic biomass degradation in filamentous fungi - Springer, n.d. . Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-014-5707-6

Dassa, B., Borovok, I., Lamed, R., Henrissat, B., Coutinho, P., Hemme, C.L., Huang, Y., Zhou, J., Bayer, E.A., 2012. Genome-wide analysis of Acetivibrio cellulolyticus provides a blueprint of an elaborate cellulosome system. BMC Genomics 13, 210. https://doi.org/10.1186/1471-2164-13-210

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

Faghihi, M.A., Wahlestedt, C., 2009. Regulatory roles of natural antisense transcripts. Nat. Rev. Mol. Cell Biol. 10, 637–643. https://doi.org/10.1038/nrm2738

Goodman, A.J., Daugharthy, E.R., Kim, J., 2013. Pervasive Antisense Transcription Is Evolutionarily Conserved in Budding Yeast. Mol. Biol. Evol. 30, 409–421. https://doi.org/10.1093/molbev/mss240

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

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I., 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42, D699-704. https://doi.org/10.1093/nar/gkt1183

Haitjema, C.H., Gilmore, S.P., Henske, J.K., Solomon, K.V., Groot, R. de, Kuo, A., Mondo, S.J., Salamov, A.A., LaButti, K., Zhao, Z., Chiniquy, J., Barry, K., Brewer, H.M., Purvine, S.O., Wright, A.T., Hainaut, M., Boxma, B., Alen, T. van, Hackstein, J.H.P., Henrissat, B., Baker, S.E., Grigoriev, I.V., O'Malley, M.A., 2017. A parts list for fungal cellulosomes revealed by comparative genomics. Nat. Microbiol. 2, 17087. https://doi.org/10.1038/nmicrobiol.2017.87

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

Hastings, M.L., Ingle, H.A., Lazar, M.A., Munroe, S.H., 2000. Post-transcriptional Regulation of Thyroid Hormone Receptor Expression by cis-Acting Sequences and a Naturally Occurring Antisense RNA. J. Biol. Chem. 275, 11507–11513. https://doi.org/10.1074/jbc.275.15.11507

Henske, J.K., Wilken, S.E., Solomon, K.V., Smallwood, C.R., Shutthanandan, V., Evans, J.E., Theodorou, M.K., O'Malley, M.A., 2018. Metabolic characterization of anaerobic fungi provides a path forward for bioprocessing of crude lignocellulose. Biotechnol. Bioeng. 115, 874–884. https://doi.org/10.1002/bit.26515

Hou, C., Tian, W., Kleist, T., He, K., Garcia, V., Bai, F., Hao, Y., Luan, S., Li, L., 2014. DUF221 proteins are a family of osmosensitive calcium-permeable cation channels conserved across eukaryotes. Cell Res. 24, 632–635. https://doi.org/10.1038/cr.2014.14

Jabnoune, M., Secco, D., Lecampion, C., Robaglia, C., Shu, Q., Poirier, Y., 2013. A Rice cis-Natural Antisense RNA Acts as a Translational Enhancer for Its Cognate mRNA and Contributes to Phosphate Homeostasis and Plant Fitness. Plant Cell 25, 4166–4182. https://doi.org/10.1105/tpc.113.116251

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engström, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B., Wahlestedt, C., 2005. Antisense Transcription in the Mammalian Transcriptome. Science 309, 1564–1566. https://doi.org/10.1126/science.1112009

Koonin, E.V., 2005. Orthologs, Paralogs, and Evolutionary Genomics1. Annu. Rev. Genet. 39, 309–338. https://doi.org/10.1146/annurev.genet.39.073003.114725

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

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., Regev, A., 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat. Methods 7, 709–715. https://doi.org/10.1038/nmeth.1491

Li, G.J., Hyde, K.D., Zhao, R.L., Hongsanan, S., Abdel-Aziz, F.A., Abdel-Wahab, M.A., Alvarado, P., Alves-Silva, G., Ammirati, J.F., Ariyawansa, H.A., Baghela, A., Bahkali, A.H., Beug, M., Bhat, D.J., Bojantchev, D., Boonpratuang, T., Bulgakov, T.S., Camporesi, E., Boro, M.C., Ceska, O., Chakraborty, D., Chen, J.J., Chethana, K.W.T., Chomnunti, P., Consiglio, G., Cui, B.K., Dai, D.Q., Dai,

Y.C., Daranagama, D.A., Das, K., Dayarathne, M.C., Crop, E.D., Oliveira, R.J.V.D., Souza, C.A.F. de, Souza, J.I. de, Dentinger, B.T.M., Dissanayake, A.J., Doilom, M., Drechsler-Santos, E.R., Ghobad-Nejhad, M., Gilmore, S.P., Góes-Neto, A., Gorczak, M., Haitjema, C.H., Hapuarachchi, K.K., Hashimoto, A., He, M.Q., Henske, J.K., Hirayama, K., Iribarren, M.J., Jayasiri, S.C., Jayawardena, R.S., Jeon, S.J., Jerônimo, G.H., Jesus, A.L., Jones, E.B.G., Kang, J.C., Karunarathna, S.C., Kirk, P.M., Konta, S., Kuhnert, E., Langer, E., Lee, H.S., Lee, H.B., Li, W.J., Li, X.H., Liimatainen, K., Lima, D.X., Lin, C.G., Liu, J.K., Liu, X.Z., Liu, Z.Y., Luangsa-ard, J.J., Lücking, R., Lumbsch, H.T., Lumyong, S., Leaño, E.M., Marano, A.V., Matsumura, M., McKenzie, E.H.C., Mongkolsamrit, S., Mortimer, P.E., Nguyen, T.T.T., Niskanen, T., Norphanphoun, C., O'Malley, M.A., Parnmen, S., Pawłowska, J., Perera, R.H., Phookamsak, R., Phukhamsakda, C., Pires-Zottarelli, C.L.A., Raspé, O., Reck, M.A., Rocha, S.C.O., Santiago, A.L.C.M.A. de, Senanayake, I.C., Setti, L., Shang, Q.J., Singh, S.K., Sir, E.B., Solomon, K.V., Song, J., Srikitikulchai, P., Stadler, M., Suetrong, S., Takahashi, H., Takahashi, T., Tanaka, K., Tang, L.P., Thambugala, K.M., Thanakitpipattana, D., Theodorou, M.K., Thongbai, B., Thummarukcharoen, T., Tian, Q., Tibpromma, S., Verbeken, A., Vizzini, A., Vlasák, J., Voigt, K., Wanasinghe, D.N., Wang, Y., Weerakoon, G., Wen, H.A., Wen, T.C., Wijayawardene, N.N., Wongkanoun, S., Wrzosek, M., Xiao, Y.P., Xu, J.C., Yan, J.Y., Yang, J., Yang, S.D., Hu, Y., Zhang, J.F., Zhao, J., Zhou, L.W., Peršoh, D., Phillips, A.J.L., Maharachchikumbura, S.S.N., 2016. Fungal diversity notes 253–366: taxonomic and phylogenetic contributions to fungal taxa. Fungal Divers. 1–237. https://doi.org/10.1007/s13225-016-0366-9

Ni, T., Tu, K., Wang, Z., Song, S., Wu, H., Xie, B., Scott, K.C., Grewal, S.I., Gao, Y., Zhu, J., 2010. The Prevalence and Regulation of Antisense Transcripts in Schizosaccharomyces pombe. PLOS ONE 5, e15271. https://doi.org/10.1371/journal.pone.0015271

Nocetti, N., Whitehouse, I., 2016. Nucleosome repositioning underlies dynamic gene expression. Genes Dev. 30, 660–672. https://doi.org/10.1101/gad.274910.115

Ohm, R.A., de Jong, J.F., Lugones, L.G., Aerts, A., Kothe, E., Stajich, J.E., de Vries, R.P., Record, E., Levasseur, A., Baker, S.E., Bartholomew, K.A., Coutinho, P.M., Erdmann, S., Fowler, T.J., Gathman, A.C., Lombard, V., Henrissat, B., Knabe, N., Kües, U., Lilly, W.W., Lindquist, E., Lucas, S., Magnuson, J.K., Piumi, F., Raudaskoski, M., Salamov, A., Schmutz, J., Schwarze, F.W.M.R., vanKuyk, P.A., Horton, J.S., Grigoriev, I.V., Wösten, H.A.B., 2010. Genome sequence of the model mushroom Schizophyllum commune. Nat. Biotechnol. 28, 957–963. https://doi.org/10.1038/nbt.1643

O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J., Proudfoot, N.J., 2004. Gene loops juxtapose promoters and terminators in yeast. Nat. Genet. 36, 1014–1018. https://doi.org/10.1038/ng1411

Pelechano, V., Steinmetz, L.M., 2013. Gene regulation by antisense transcription. Nat. Rev. Genet. 14, 880–893. https://doi.org/10.1038/nrg3594

Piatek, M.J., Henderson, V., Zynad, H.S., Werner, A., 2016. Natural antisense transcription from a comparative perspective. Genomics 108, 56–63. https://doi.org/10.1016/j.ygeno.2016.05.004

Ries, L., Pullan, S.T., Delmas, S., Malla, S., Blythe, M.J., Archer, D.B., 2013. Genome-wide transcriptional response of Trichoderma reesei to lignocellulose using RNA sequencing and comparison with Aspergillus niger. BMC Genomics 14, 541. https://doi.org/10.1186/1471-2164-14-541

Rosewich, U.L., Kistler, and H.C., 2000. Role of Horizontal Gene Transfer in the Evolution of Fungi. Annu. Rev. Phytopathol. 38, 325–363. https://doi.org/10.1146/annurev.phyto.38.1.325

Serobyan, V., Xiao, H., Namdeo, S., Rödelsperger, C., Sieriebriennikov, B., Witte, H., Röseler, W., Sommer, R.J., 2016. Chromatin remodelling and antisense-mediated up-regulation of the developmental switch gene eud-1 control predatory feeding plasticity. Nat. Commun. 7, 12337. https://doi.org/10.1038/ncomms12337

Smith, C.A., Robertson, D., Yates, B., Nielsen, D.M., Brown, D., Dean, R.A., Payne, G.A., 2008. The effect of temperature on Natural Antisense Transcript (NAT) expression in Aspergillus flavus. Curr. Genet. 54, 241–269. https://doi.org/10.1007/s00294-008-0215-9

Solomon, K.V., Haitjema, C.H., Henske, J.K., Gilmore, S.P., Borges-Rivera, D., Lipzen, A., Brewer, H.M., Purvine, S.O., Wright, A.T., Theodorou, M.K., Grigoriev, I.V., Regev, A., Thompson, D.A., O'Malley, M.A., 2016a. Early-branching gut fungi possess a large, comprehensive array of biomass-degrading enzymes. Science 351, 1192–1195. https://doi.org/10.1126/science.aad1431

Solomon, K.V., Henske, J.K., Theodorou, M.K., O'Malley, M.A., 2016b. Robust and effective methodologies for cryopreservation and DNA extraction from anaerobic gut fungi. Anaerobe 38, 39–46. https://doi.org/10.1016/j.anaerobe.2015.11.008

Taft, R.J., Pheasant, M., Mattick, J.S., 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. BioEssays 29, 288–299. https://doi.org/10.1002/bies.20544

Theodorou, M.K., Brookman, J., Trinci, A.P.J., 2005. Anaerobic fungi, in: Makkar, H.P.S., McSweeney, C.S. (Eds.), Methods in Gut Microbial Ecology for Ruminants. Springer Netherlands, Dordrecht, pp. 55–66.

Wagner, E.G.H., Simons, R.W., 1994. Antisense RNA Control In Bacteria, Phages, And Plasmids. Annu. Rev. Microbiol. 48, 713–742.

Wilusz, C.J., Wilusz, J., 2013. Lsm proteins and Hfq. RNA Biol. 10, 592–601. https://doi.org/10.4161/rna.23695

Wu, T.D., Watanabe, C.K., 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21, 1859–1875. https://doi.org/10.1093/bioinformatics/bti310

Yan, B., Wang, Z., 2012. Long Noncoding RNA: Its Physiological and Pathological Roles. DNA Cell Biol. 31, S-34. https://doi.org/10.1089/dna.2011.1544

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

588 **Table 1: Overview of sequenced ananaerobic fungal genomes and transcriptomes**

	Anaeromyces robustus	***Neocallimastix californiae***	***Piromyces finnis***
Genome Size (MBp)[a]	72	193	56
Genome Scaffolds[a]	1,035	1,819	232
Transcriptome (transcripts)[a]	17,128	29,649	17,008
Mapped Transcripts	15,998 (93%)	27,842 (94%)	16,246 (96%)
Antisense RNA (transcripts/ % transcriptome)[b]	2,157 (13%)	4,832 (16%)	3,814 (22%)
Validated NATs (transcripts/ % transcriptome)	**439 (2.5%)**	**732 (2.5%)**	**1,586 (9.3%)**

589 a (Haitjema et al., 2017); b(Solomon et al., 2016a)

590

Figure Legends

Figure 1. Cis Natural Antisense Transcripts (NATs) are a significant fraction of the gut fungal mRNA landscape. A) Validated NATs are defined as mRNA that may be annotated in the reverse annotation and can be mapped to the genome complementary to a coding transcript. These transcripts assume a (i) head-to-head (ii) tail-to-tail, (iii) embedded or (iv) fully complementary binding configuration B) NATs are large ranging in size from 100 bp to 6 kb. C) NATs are highly complementary to the 3' end of their target mRNA. TSS – target gene transcriptional start site. TT – target gene transcriptional terminator.

Figure 2. NATs assume diverse regulatory roles within gut fungi. NAT function of categorized transcripts as described by the Gene Ontology (GO) Term annotations of their target RNA. Refer to Materials & Methods for GO Term definitions of the functional bins.

Figure 3. NATs are overrepresented in key functional roles. GO Terms enriched with antisense (hypergeometric test, $p \leq 0.01$) suggest processes (P), functions (F), and cellular compartments (C) where NAT regulation is critical.

Figure 4. NATs and their targets are independently regulated with antisense hypothesized to control expression of selected transcripts. Differentially expressed genes ((≥2-fold change in expression, $p_{val} \leq 0.01$) were binned by whether both the target and corresponding NAT were regulated (coregulated), NAT was regulated alone (antisense regulated), or target regulated alone (transcriptionally regulated). Due to the genome density, some NATs may control multiple targets. NATs common to multiple targets are indicated with a uniquely colored bar above the heatmap. A) *Anaeromyces robustus* B) *Neocallimastix californiae* C) *Piromyces finnis* D) Cartoon of potential impact on protein expression for an increase in NAT and target (positive coregulation), NAT only (antisense regulation), and target only (transcriptionally regulated). Target transcripts indicated in blue, and NAT in red/peach.

Figure 5. Independent antisense regulation of genes determined by relative expression level of NAT and target gene. Relative expression levels of target and NAT in transcripts per million (TPM) were measured on both glucose and reed canary grass. NAT:target pairs were color coded by method of gene regulation as observed in Figure 5.