High quality anaerobic fungal genome assemblies and annotation for study and optimization of lignocellulose conversion

Casey Hooker[1,2*] (hookerc@purdue.edu), Ethan Hillman[1,3], Javier Muñoz Briones[1,3], and **Kevin Solomon[1,2,3†]**

[1] Agricultural & Biological Engineering, Purdue University, West Lafayette, IN; [2]Laboratory of Renewable Resources Engineering (LORRE), Purdue University, West Lafayette, IN; [3] Purdue University Interdisciplinary Life Sciences (PULSe) Program, Purdue University, West Lafayette, IN

[†]Present Address: Chemical & Biomolecular Engineering, University of Delaware, Newark, DE

http://solomonlab.weebly.com

Project Goals: This project develops genetic and epigenetic tools for emerging model anaerobic fungi to identify the genomic determinants of their powerful biomass-degrading capabilities, facilitate their study, and enable direct fungal conversion of untreated lignocellulose to bioproducts.

Deconstruction of plant cell wall biomass is a significant bottleneck to the production of affordable biofuels and bioproducts. Anaerobic fungi (*Neocallimastigomycota*) from the digestive tracts of large herbivores, however, have evolved unique abilities to degrade untreated fiber-rich plant biomass by combining hydrolytic strategies from the bacterial and fungal kingdoms[1]. Anaerobic fungi secrete the largest known diversity of lignocellulolytic carbohydrate active enzymes (CAZymes) in the fungal kingdom (>300 CAZymes), which unaided can degrade up to 60% of the ingested plant material within the animal digestive tract[2,3]. Unlike many other fungal systems, these CAZymes are tightly regulated and assembled in fungal cellulosomes to synergistically degrade plant material, including untreated agricultural residues, bioenergy crops, and woody biomass, with comparable efficiency regardless of composition[1,4–6]. However, the specific role of individual enzymes in maintaining hydrolytic efficiency remains unknown due to a lack of genetic tools that facilitate testing of gene function in its natural context. Thus, there is a critical need to create methods that manipulate CAZyme expression and rapidly interrogate gene function in anaerobic fungi to identify targets that will advance biofuel and bioproducts production.

In this project, we study three novel specimens of anaerobic fungi. Species characterization confirm that they are unique and exhibit high enzymatic activity against a range of untreated lignocellulolytic substrates regardless of lignin composition (e.g. poplar, sorghum, alfalfa, corn stover)[6]. Anaerobic fungi tailor the secretion of CAZymes to adapt to differences in substrate composition and achieve consistently high-levels of synergistic activity. To better understand this response, we need to develop tools for genetic engineering of anaerobic fungi.

As a first step towards tool development, we assembled new fungal genomes to identify key regulatory sequences for future parts. Previous genomes were highly fragmented in to as many as 30,000 scaffolds. That is, many potential parts identified from these datasets could be non-functional due to truncations in the sequence due to poor assembly. However, by leveraging high quality genome isolations, long-read sequencing, and Hi-C (chromosomal conformation capture)

sequencing, we have improved genome assembly by an order of magnitude to generate the first genomes for this family of organisms with chromosomal resolution. For example, our assembly of *Piromyces* sp. UH3-1 incorporated 99% of the genome into 12 chromosomes, a 94% reduction in fragmentation relative to the historically best anaerobic fungal genome assemblies.

Annotation of these rich datasets for parts discovery is underway. We have developed a bioinformatic pipeline to identify conserved AT-rich promoters that may drive CAZyme expression and are currently synthesizing more than a dozen at the JGI for evaluation along with reporter proteins and other putative parts for gene expression. In parallel, we have begun to identify and test other regulatory sequences such as centromere binding sequences (CEN) and autonomous replicating sequences (ARS) needed to create self-replicating plasmids that may be used to deliver new genes and control cell phenotypes. We are also acquiring transcriptomes and proteomes of our fungal isolates across various substrates in partnership with the DOE-JGI and EMSL. Using these resources, we are identifying parts and environmental conditions that regulate them to develop novel tools for stable gene expression.

In addition to developing specific parts, we have also studied the genetics of anaerobic fungi to inform strategies for manipulation of CAZymes and other genes. Genome annotation and nuclei labelling suggest that all our isolates are haploid (only one chromosomal copy per species), which streamlines future genetic engineering testing and development. Genome architecture analysis has also identified several features such as long-terminal repeat sequences and horizontal gene transfer events, which may be mutational hotspots on the genome to avoid in future strain engineering for stable constructs.

In summary, the ongoing work has begun to identify parts for a genetic toolbox and provides an atlas of anaerobic fungal genomes complete with targets for study and genomic regions to avoid. Once validated, these parts will form foundational tools to generate a deeper systems-level understanding of anaerobic fungal physiology while establishing fundamental knowledge about regulation of gut fungal CAZymes. Ultimately, we enable predictive biology in anaerobic fungi and derive insight into microbial plant deconstruction to advance the development of economical biofuels and bioproducts.

References

1 Haitjema, C. H. *et al. Nature Microbiology* 2, 17087 (2017)
2 Seppälä, S. *et al. Metabolic Engineering* 44, 45–59 (2017)
3 Youssef, N. H. *et al. Appl. Environ. Microbiol.* 79, 4620–4634 (2013)
4 Solomon, K. V. *et al. Science* 351, 1192–1195 (2016)
5 Solomon, K. V. *et al. Fungal Genetics and Biology* 121, 1–9 (2018)
6 Hooker, C. A. *et al. Biotechnology for Biofuels* 11, 293 (2018)

This project is supported by the Office of Biological and Environmental Research through the DOE Office of Science (Early Career Program) Contract No DE-SC0020117. A portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative and used resources at the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory, which are DOE Office of Science User Facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under Contract Nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL).