DEVELOPING ENGINEERING TOOLS FOR ANAEROBIC FUNGI

by

Ethan Hillman

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Agricultural and Biological Engineering

West Lafayette, Indiana

May 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Kevin Solomon, Chair

School of Agricultural and Biological Engineering

Dr. Kari Clase

School of Agricultural and Biological Engineering

Dr. Cindy Nakatsu

Department of Agronomy

Dr. Michael Scharf

Department of Entomology

Approved by:

Dr. Nathan Mosier

Department of Agricultural & Biological Engineering

Dr. Jason Cannon

Interdisciplinary Life Sciences

Dedicated to my wife to whom I am forever indebted for immeasurable amounts of moral, familial, and culinary support. Without her sacrifices, I would not have been able to complete this work.

ACKNOWLEDGMENTS

I would like to thank Dr. Kevin Solomon for his guidance and support of me as a graduate student and researcher. I am extremely grateful for the freedom you granted during your mentorship that allowed me to explore my own ideas for better or for worse, as well as for your patience during the trials of science and writing. Thank you also for allowing me to be one of the founding members of the lab and for providing countless opportunities to become a better researcher, writer, teacher, and mentor. I am very honored to have had you as my advisor, and I would not be the same researcher without your guidance.

I am also grateful for the mentorship of Dr. Cindy Nakatsu and Dr. Steve Lindemann who helped me gain additional perspective in how to think like a microbe. The time and effort they invested in me despite not being their student has greatly shaped how I approach and understand microbiome research, and I am very grateful for their guidance. Similarly, I would like to express gratitude to all the Solomon Lab members and my Purdue University Interdisciplinary Life Science program cohort who have provided uncountable hours of feedback, moral support, and an extra set of hands in the lab. Finally, I would also like thank my undergraduate researchers who helped make this work possible, put up with my crazy ideas, and gave me the pleasure of guiding them through research in synthetic biology. It was such a pleasure to work with everyone, and I will always have you to thank for enhancing my time here at Purdue.

Lastly, I am grateful for the many people who have provided feedstocks, equipment, or expertise for this work: Dr. Cliff Weil, Dr. Richard Meilan, Dr. Nathan Mosier, Dr. Bernard Engel, Dr. Abby Engelberth, Dr. Klein Illeleji, C3Bio, Xingya Liu, LORRE, Blaine Brown, and the Indianapolis Zoo. Additionally, I owe many thanks to our collaborators at Phase Genomics and the Joint Genome Institute, as well as the Pioneer Oil Company and the Department of Energy for support my work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

%GC	nucleotide percentage of DNA that is made up of Gs + Cs
2-PE	2-phenylethanol
5-FOA/OA	5- fluoroorotic acid / orotic acid
5-FUMP/UMP	5-fluorinated / uridine monophosphate
5-OMP/OMP	5-fluorinated / orotidine 5'-monophosphate
AcCoA	acetyl-coenzyme A (CoA)
AGF	Anaerobic gut fungus
AmpR	ampicillin resistance
ARO10	phenylpyruvate decarbixylase
ARS	autonomously replicating sequence
atoB	acetyl-CoA acetyltransferase, also aacT,
BGC	biosynthetic gene cluster
BLAST	basic local alignment search tool
CAI	codon adaptation index
Cas9	CRISPR associated protein 9
CAZyme	Carbohydrate active enzyme
CBP	Carbohydrate binding protein
CDS	coding sequence
CEN	centromere binding sequence
CMC	Carboxy methyl cellulose
CO2	carbon dioxide
CRISPR	clustered regularly interspaced short palindromic repeats
DAPI	4’ 6 diamidino-2-phenylindole
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
DOE	Department of Energy
EAT1	ethanol acetyl transferase 1
EtAc	ethyl acetate
FAC	fungal artificial chromosome

G lignin	Guiacyl lignin
gDNA	genomic DNA
Gf-ma	*Neocallimastix sp. Gf-ma*
GFP	Green fluorescent protein
GH#	glycoside hydrolase family number
GI	gastrointestinal tract
GusA	beta-glucuronidase gene
H Lignin	Para-hydroxy phenyl lignin
H2	hydrogen gas
HAT	histone Acetyl Transferase.
HGT	horizontal gene transfer
HiC	genome-wide chromatin confirmation capture sequencing
HMGR	Hydroxy β-methylglutaryl-CoA reductase
HMGS	Hydroxy β-methylglutaryl-CoA synthase
HPLC	high performance/pressure liquid chromatography
HygR	hygromycin resistance gene
IGS	rRNA operon intergenic spacer
iLOV	light, oxygen, voltage domain flavin-binding fluorescent protein
IPTG	isopropyl β-D-1-thiogalactopyranoside
iRFP	near-infrared fluorescent protein
IsA	isoamyl acetate
ITS	Internal transcribed spacer
JGI	Joint Genome Institute
Kb/Kbp	kilo/thousand base pairs (10^3 bases)
L50	smallest number of contigs whose sum length equals half of genome size
LB	Luria-Bertani rich medium
LC	liquid chromatography
LSU	28s rRNA large ribosomal subunit
Mb/Mbp	mega/million base pairs (10^6 bases)
MB	Medium B (synthetic media for anaerobic fungus)
MC	Medium C (complex, rumen-fluid based media for anaerobic fungus)

mRNA	messenger RNA
MS/MS	tandem mass spectroscopy
MUG	4-Methylumbelliferyl-β-D-galactopyranoside
N50	shortest contig length needed to cover half of the genome size.
NAD+	nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NADP+	nicotinamide adenine dinucleotide phosphate
NADPH	reduced nicotinamide adenine phosphate dinucleotide
NLS	nuclear localization sequence
NMR	nuclear magnetic resonance
NP	natural product
NRPS	non-ribosomal peptide synthetase
NTS	rRNA operon non-transcribed spacer
OD, OD600	optical density
PacBio	Pacific Biosciences sequencing platform
PCR	polymerase chain reaction
Penol	enolase promoter
PI	propidium iodide
PKS	polyketide synthase
PPi	inorganic pyrophosphate
PRPP	phosphoribosyl pyrophosphate
PTM	post-translational modification
RNA	ribonucleic acid
RNAi	RNA interference
S lignin	Syringyl lignin
SAW	sterile anaerobic water
SC / CS / P / A	Sigmacell/ corn stove / poplar / alfalfa
SSU	18s rRNA small ribosomal subunit
tRNA	transfer RNA
tTEF	translation elongation factor 1a terminator
UH3-1	*Piromyces sp. UH3-1*

Ura3	orotidine 5'-phosphate decarboxylase
Ura5	orotate phosphoribosyltransferase
UTR	untranslated region
WI3-B	*Neocallimastix sp. WI3B*
YPD	yeast peptone dextrose medium

ABSTRACT

Renewable plant biomass represents a rich source of fixed carbon that is poised to accelerate the growth of the bioeconomy because it is widely available, underused, and inexpensive. Similarly, because it is a ubiquitous, carbohydrate-rich feedstock that can be used in a broad range of bioprocesses, current and emerging technologies are being designed to transform these feedstocks into a variety of products including surfactants, food additives, pigments, plastics, and biofuels. However, current strategies to deconstruct recalcitrant plant materials rely on expensive enzymes with inefficient and harsh pretreatment steps. However, anaerobic fungi degrade a variety of crude, untreated biomass materials into fermentable sugars that can be converted into various products making them an appealing, low-cost solution to this problem. Although there are potential applications in industry for anaerobic fungi, it remains untapped because of the difficulties in cultivating them, sequencing their genomes, and genetically engineering them.

In this work, three novel anaerobic fungi were isolated, and their genomes were sequenced to identify their genomic potential that was then leveraged to develop bioprocesses and engineering tools. Specifically, I developed methods to acquire the first gapless genomes for anaerobic fungi to provide more comprehensive insight into their capabilities. The biomass hydrolyzing abilities of one strain were characterized and leveraged as a pretreatment system for plant biomass; by partnering these anaerobic fungi with *K. marxianus* yeast, higher carbon conversion to fine and commodity chemicals was achieved as part of a two-stage bioproduction system. Similarly, the genomes were leveraged to identify novel genes for mevalonate production. My analysis of codon utilization due to the unusual GC composition of these genomes overcome one of the challenges with heterologous gene expression, leading to a hybrid pathway in *E. coli* with titers up to 2.5 g/L of mevalonate. Finally, a basic set of genetic tools was created including promoters, reporters, selection markers, and a gene-editing system that are still in development but form the fundamental toolbox for genetic engineering in anaerobic fungi. Together, this work provides a foundation for future genetic and metabolic engineering approaches that can enhance the efficiency and production of chemicals and fuels from renewable plant biomass.

1. INTRODUCTION

1.1 Motivation and Problem Statement

Renewable plant biomass represents a rich source of fixed carbon that is poised to accelerate the growth of the bioeconomy because it is underused and widely available [1], [2]. With approximately 150-170 billion tons of plant biomass being produced annually, it is an inexpensive, ubiquitous, and carbohydrate-rich (up to 55-75% of their composition) feedstock that can be used in a broad range of bioprocesses [3], [4]. Using current and emerging technologies, these feedstocks can be transformed into a variety of products from surfactants, paints, and food additives to pigments, plastics, and biofuels making up a global market of over $48 billion each year [1], [5]–[7]. However, current strategies to degrade recalcitrant plant material rely on expensive enzyme cocktails and/or harsh pretreatment techniques [7]–[10]. Therefore, there is a critical need to develop lignocellulose degrading systems that are high efficiency and low cost, and will ultimately enable us to better exploit the stockpile of renewable plant biomass for bioproduction of fuels and chemicals.

Because anaerobic fungi can degrade crude, untreated biomass into fermentable sugars that can be converted into a wide variety of products, they are an appealing low-cost solution to this problem [11]–[13]. These fungi can degrade a variety of feedstocks, including food waste and lignin-rich poplar [14] due to the fact that they have the largest repertoire of biomass degrading enzymes in the fungal kingdom[15] – even larger than the well-known biomass degraders *Trichoderma reesei*,[10] white rot fungi [16], and wood decay fungi [17]. More importantly, enzymes produced during cultivation of anaerobic fungi were competitive with commercial *Aspergillus* and *Trichoderma* cocktails on untreated biomass [18]. However, adapting anaerobic fungi in biotechnology platforms is currently hindered because their genomes remain poorly understood and even a set of basic genetic engineering tools have not been developed.

While there is a wealth of potential in anaerobic fungi, the difficulties in cultivating them and sequencing their genomes have hindered their utilization in industrial applications. Additionally, obtaining genomic information for them is challenging because their genomes are more AT-rich (18% GC) than any other fungi [19], making both sequencing and assembly difficult. As a result, less than a dozen genomes are available for anaerobic fungi and further hinders our

understanding of their physiology, metabolism, and biosynthetic potential. Similarly, the poor resolution of the existing genomes, which are fragmented into hundreds of pieces, limits our ability to mine them for genetic parts like promoters that can be used to enhance expression of biomass degrading enzymes or DNA editing enzymes such as Cas9.

Without genetic tools, identification, characterization, and isolation of their highly effective enzymes is limited to heterologous expression in model organisms such as *E. coli* or *S. cerevisiae*. Yet, these model organisms have had little success with genes from anaerobic fungi [20], [21] and have limited utility until they are better understood [22]. Expression in the anaerobic fungi, however, provides many advantages over heterologous expression because their genes can be evaluated in their native context where interactions with other proteins, cellular localization, and regulatory modifications are conserved. Therefore, the development of a set of genetic engineering tools is critical for identifying, characterizing, and harnessing the genes underlying their extensive biomass degrading abilities and biosynthetic capabilities. Ultimately, the combination of these tools and high-quality genomes will allow metabolic engineering strategies to probe and manipulate their physiology, leading to strains that can enhance biofuel and bioproduct production for lignocellulosic feedstocks.

1.2 Scope and Objectives

In order to overcome the barriers that restrict realization of the potential of anaerobic fungi, this dissertation aims to improve the quantity, quality, and accuracy of their genomes, to leverage their genomic and metabolic potential for bioproduction, and to develop the first set of genetic engineering tools for anaerobic fungi. Specifically, the work in this dissertation will:

- Isolate novel anaerobic fungus species and characterize their feedstock utilization
- Optimize the extraction, sequencing, and assembly of anaerobic fungi to better resolve their genetic potential and organization
- Produce fine and commodity chemicals from plant biomass by partnering anaerobic fungi with engineered yeast
- Demonstrate how model microbes can be leveraged to express anaerobic fungus genes
- Develop basic genetic parts that allow heterologous expression in anaerobic fungi

1.3 Thesis Organization

This dissertation is structured around developing multiple parallel strategies to realize the potential of anaerobic fungi. Chapter 2 lays out the framework of these strategies that include partnering microorganisms, leveraging model microbes for heterologous expression, and using genetic tools in the host organism. While Chapter 2 focuses specifically on the production of natural products, these same strategies can be adapted to exploit the biomass-degrading potential of anaerobic fungi in addition to the biosynthetic potential. In Chapter 3, the isolation and characterization of a novel anaerobic species is described. Here, this isolate's ability to use various waste streams including lignin-rich substrates that are problematic for many biomass degradation platforms. The isolation of a high-quality genome for this species and two other novel isolates is the focus of Chapter 4. Strategies are explored to optimize DNA extraction and pair it with new sequencing technologies to resolve their genomes to chromosomes for the first time. In the absence of genetic tools for anaerobic fungi, the degradative abilities (Chapter 5) and biosynthetic genes (Chapter 6) are leveraged for bioprocesses that exploit the potential of one isolate. In Chapter 5, anaerobic fungal pretreatment of biomass was evaluated as a strategy to sustain growth and chemical production in a downstream organism (*Kluyveromyces marxianus*). The yield of fragrances and next generation biofuels from this process such as ethyl acetate, 2-phenylethanol, and isoamyl alcohol, was evaluated from various lignocellulosic feedstocks. In Chapter 6, the focus was on identifying challenges associated with expressing genes from anaerobic fungi in *E. coli*. Insight from genomic data is gained and used to design optimized genes that alleviate the challenges of heterologously expressing these genes. Production of mevalonate, a valuable terpenoid precursor, is evaluated for the pathway homologs from one anaerobic fungus isolate. Chapter 7 describes the development of basic genetic tools for anaerobic fungi and discusses ongoing work to optimize them for genetic engineering. The combination of promoters, reporters, resistance genes, and DNA repair cassettes to form a gene editing system is shown and the potential for metabolic engineering is discussed. Finally, Chapter 8 summarizes the findings of Chapters 3-7 and provides a perspective on the impact and future directions.

2. EXPLOITING THE NATURAL PRODUCT POTENTISL OF FUNGI WITH INTEGRATED -OMICS AND SYNTHETIC BIOLOGY APPROACHES

This chapter is adapted from a paper by Ethan T Hillman, Logan R Readnour, and Kevin V. Solomon. *Current Opinions in Systems Biology* 5 (2017):50-56. [23].

2.1 Abstract

Fungi are rich, underexploited reservoirs for natural products that may serve as medicines, commodity chemicals, insecticides, pesticides and other valuable chemicals. Moreover, the biochemistry of natural product formation may be repurposed with emerging synthetic biology tools to make valuable non-natural compounds such as biofuels. However, the pathways that lead to these products are poorly understood and frequently inactive under lab conditions making discovery challenging. Recent advances in –omics approaches and synthetic biology tools provide powerful new methods to elucidate and tap this wealth of novel chemistry. In this review, we describe cutting-edge approaches to activate and characterize natural product formation and discuss the potential of established and emerging fungal systems for natural product discovery.

2.2 Introduction

Fungal secondary metabolites, or natural products (NPs), form an increasingly important portion of the pharmaceutical, agricultural, and industrial sectors (Figure 2-1). In the pharmaceutical industry alone, single molecules such as the powerful fungal NP-derived antibiotic amoxicillin are each estimated to top $1 billion USD in annual sales worldwide [24]. Among fungal NPs are cholesterol-lowering statins (e.g. the blockbuster drug lovastatin[25]), common antibiotics (e.g penicillin and cephalosporin[25]), insecticides (e.g. nodulisporic acids [25]), and pigments (e.g. carotenoids [26]). NPs are frequently bioactive and chemically diverse due to their role in conferring a selective advantage to the producing fungus in competitive environments [27]. Bolstered by these compounds, fungi are able to proliferate and thrive in diverse ecological niches including competitive soil environments [28] and the digestive tracts of humans and other mammals [29], [30]. However, of the estimated five million fungal species, less than 2% have been described and fewer than a 1,000 have been sequenced and studied in detail [31]–[33]. Thus, many

unclassified fungi with rich reservoirs for new NPs remain to be discovered within plain sight in locations such as urban parks and farm animals [34], [35].

Fungal NPs are generally classified into one of five categories based on their biosynthetic origin: polyketides, non-ribosomal peptides, terpenoids, prenylated tryptophan derivatives or hybrid compounds such as a polyketide/non-ribosomal peptide hybrids that incorporate components from several pathways (Box 1) [36]. From these basic pathways, NPs diversify into a wide range of carbon lengths (typically C6 – C30), potentially containing heterocyclic rings, and several stereocenters [37], [38]. Of the different classes, polyketides represent the majority of the fungal natural products [39]. Well-known polyketides include statins, the most profitable class of drugs worldwide, and many antibiotics [40]. Common to these NP pathways, however, is a requirement for several enzymes of diverse catalytic function that must be coordinately expressed.

Box 1: Biosynthetic Origin of Fungal Natural Products

Polyketides are synthesized from large multidomain polyketide synthases (PKSs) that iteratively construct the molecule with an acyl extender unit, commonly malonyl-CoA, through a decarboxylative Claisen condensation. The incorporated ketide unit may be subsequently functionalized and/or undergo redox reactions to generate chemical diversity before the next extender unit is added. Fungal PKSs may be further classified as either Type I or Type III. Type I PKSs use an acyl carrier protein to activate and tether the growing polyketide molecule to the PKS while Type III PKSs remain untethered from the molecule [41]. Modular multidomain non-ribosomal peptide synthetases (NRPSs) create non-ribosomal peptides in a similar fashion to Type I PKSs. However, each module of the NRPS adds an amino acid in a non-template driven synthesis reaction before functionalization [39]. Fungal terpenoid synthesis creates isoprenoid starter units via the mevalonate pathway, rather than the bacterial methylerythritol phosphate pathway [39]. These starter units (isopentyl pyrophosphate and dimethylallyl pyrophosphate; IPP and DMAPP) are condensed and cyclized by terpene synthases and cyclases to form diverse terpenoids through allylic carbocation addition [39]. Dimethylallyl tryptophan synthases may also transfer the dimethyallyl moiety from DMAPP to a tryptophan starter unit, which is then modified by methyltransferases and some tailoring enzymes to form prenylated tryptophan derivatives [39].

Although fungi do not use operons or polycistronic mRNA, they frequently organize NP biosynthetic genes into co-localized clusters (biosynthetic gene clusters; BGCs) to more easily coordinate pathway expression [39], [42]. The first identified fungal BGC was a three-gene, 56.9 kb penicillin cluster found in the fungi *A. nidulans* and *P. chrysogenum* [43], [44]. However, these BGCs need not be contiguous. Due to the size and number of proteins needed (typically >40 kb), NP formation is tightly regulated and expressed only at times when they are necessary for survival. Thus, many potential NPs evade detection as the majority of putative NP gene clusters are cryptic and remain silent under laboratory conditions.

This review covers recent advances in fungal NP discovery from the prediction of BGCs with bioinformatic tools, to their activation and validation with synthetic biology and –omics approaches. We discuss the success of this developing pipeline for NP discovery and highlight emerging fungal systems with potential for NP discovery along with new synthetic biology tools to repurpose their NP biosynthetic potential.

2.3 Prediction and validation of fungal natural products

Despite challenges in the detection of NPs, BGCs that catalyze their formation may be readily identified and mined from genomic sequences. The characteristic catalytic domains of BGCs are strongly conserved and recognized in routine homology searches of the genome with Hidden Markov Models (HMMs) and local alignments [45]. Popular bioinformatics tools SMURF [46] and antiSMASH [47] based on these principles annotate fungal genomes and have been used to identify BGCs including more than 80 non-redundant NP BGCs in *Aspergillus fumigatus* alone [48]. Similarly, analysis of *Penicillium* genomes revealed 89 putative BGCs from which NPs products were subsequently detected [49]. Next generation genomes-to-natural-products (GNP) platforms are building on this foundation to propose the chemical structure of an NP directly from its BGC including a putative LC-MS/MS (liquid chromatography, tandem mass spectrometry) spectra for detection of polyketides and nonribosomal peptides [50]. Similarly, Magarvey and coworkers have developed retro-biosynthetic tools (GRAPE & GARLIC) that solve the inverse problem of mapping an NP back to a specific BGC [51].

2.4 Activating gene expression with synthetic biology tools

Due to the cryptic nature of many NP genes in their host system, validation of novel BGCs predictions can be challenging. However, two approaches—heterologous expression and host engineering—take advantage of cutting-edge synthetic biology practices to activate expression of the otherwise silent genes. The first approach involves the transfer of these silent genes into another, more genetically tractable host, while the latter aims to express cryptic genes within the native host itself.

With the decreasing cost of DNA synthesis and growing synthetic biology toolkit [52], heterologous expression is an attractive method to activate silent BGCs from non-model organisms. For example, a silent gene cluster in the fungus *A. terreus* was revealed to produce asperfuranone, an anticancer agent candidate, after heterologous expression in the more genetically tractable *A. nidulans* [53]. When choosing a host and cloning technique, it is important to consider the advantages and challenges associated with each option. While prokaryotic hosts have a wide range of expression tools (e.g promoters, RBS), the expression of eukaryotic genes in a prokaryotic host can be problematic for several reasons: introns must be removed, significant codon bias can affect expression, protein misfolding may occur, important post-translational modifications may be missing, and/or the prokaryotic host may lack important precursors needed for biosynthesis [54]. It may be ideal then to transfer the BGC into another eukaryotic system, such as the well-studied *S. cerevisiae* or another fungal species. However, competition for important precursors or high production of similar endogenous compounds may interfere with target NP detection unless the native NP pathways are eliminated [53]. The large size of BGCs also poses a unique challenge for conventional DNA assembly. However, methods have recently been developed to solve this problem including the development of fungal artificial chromosomes (FAC) [55] and PCR-based USER cloning [56].

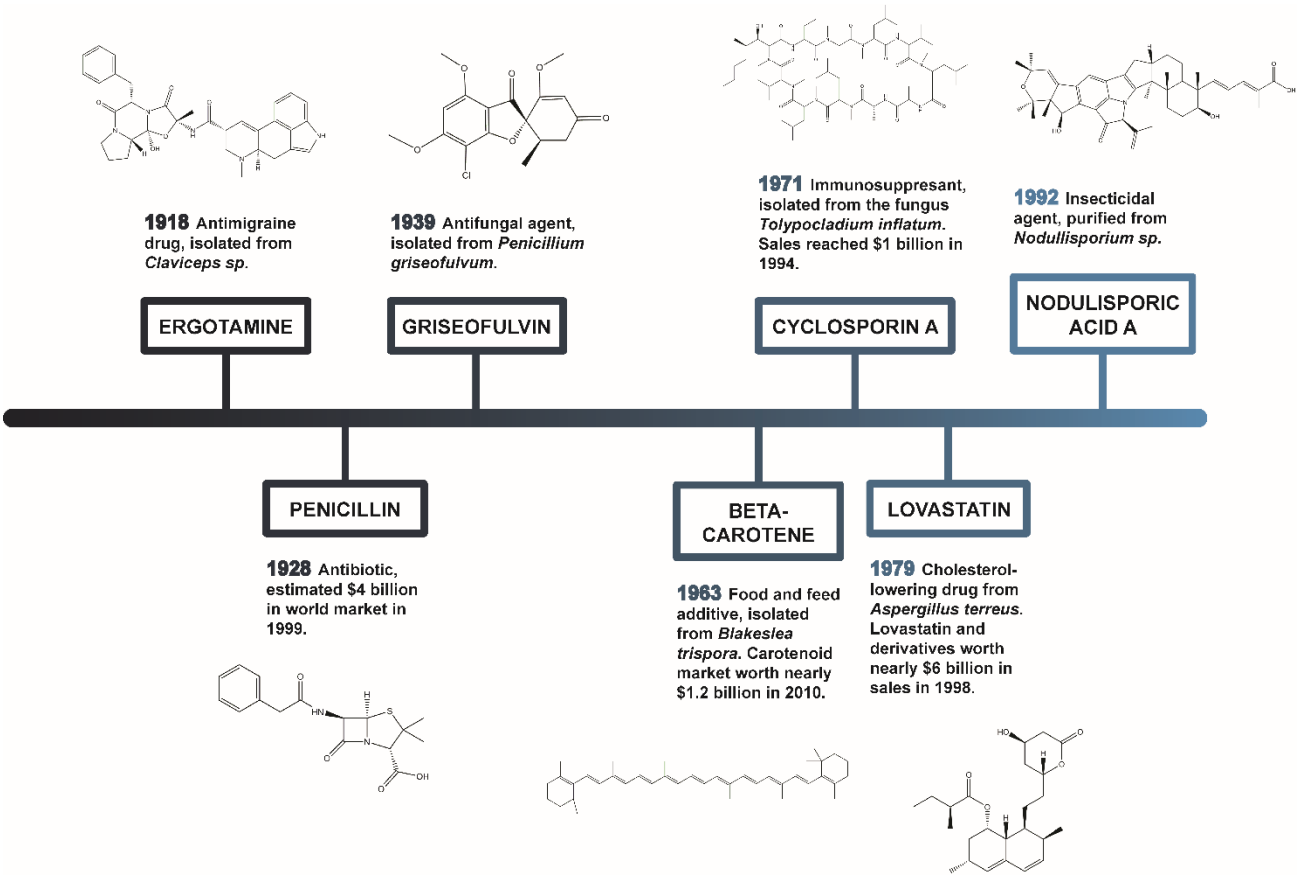


Figure 2-1: Value of natural products isolated from fungi. Fungal natural products represent a multi-billion dollar industry with diverse applications. Presented are example molecules along with their isolation year, function, and estimated market value [57]–[63]

Host engineering, on the other hand, takes advantage of the native physiology and protein expression system to correctly express the BGC [22]. However, the lack of molecular tools and well-defined regulatory elements in the host present challenges for NP discovery. Addressing this bottleneck, researchers are now adapting CRISPR/Cas9 technologies to directly manipulate the genomes of non-model fungi; recently, CRISPR systems have been developed for several filamentous fungi including *Trichoderma* [64], *Penicillium* [65], and *Neurospora* [66]. Similarly, hybrid bacterial-fungal promoters have been engineered that activated NP production in *Aspergillus* [67], which allowed for a 950-fold increase in NP production. Approaches that control host physiology (e.g. histone deacetylation [68], peroxisome number [69], and regulatory knockouts [37]) have also enabled greater NP expression. Finally, non-genetic approaches are also used to activate NP production in unsequenced organisms; co-culture of fungus with competing microbes may be used to induce NP production to confer a competitive advantage [70], [71].

Although the mechanisms are not completely understood, epigenetics may play a role in this interaction as bacteria have been shown to act through histone modification to induce NP expression [72]–[75].

2.5 Elucidating natural product biosynthetic pathways with Integrated –omics approaches

Regardless of the route of NP expression, integrated "-omics" and systems biology approaches are increasingly important in NP discovery. Untargeted metabolomic profiling is routine in the detection of new NPs [76], and can identify these compounds with high confidence even before the metabolic pathway is revealed. For example, LC-MS/MS was used to investigate the secondary metabolomes of several marine-derived strains of *Penicillium* and *Furcatum* to discover 32 secondary metabolites, including the mycotoxin patulin [77]. These approaches use powerful tandem mass spectrometry (MS/MS) to generate characteristic spectra that are compared to comprehensive compound libraries to ensure novelty from known chemical entities (dereplication) and identify known metabolites [78]. To facilitate dereplication and identify truly new compounds, there has been a shift towards crowdsourced and community curated databases that serve as a repository of all published and unpublished spectra under a wide range of analytical conditions [79]. Similarly, there has been a push for technical advancements to improve detection including spatial MS imaging of individual microbes that promises to simplify sample preparation and improve the detection sensitivity of new NPs [80].

Once a novel NP is identified, transcriptomics and proteomics may be used to elucidate its biosynthesis. Transcriptomics is used to identify active BGCs and the conditions that activate them for NP production [81]. Current advances in next generation sequencing with its low cost and sequence-agnostic nature make RNASeq a popular alternative to classical microarray approaches. For example, RNASeq was used to demonstrate that *A. niger*, *P. chrysogenum*, and *T. reesei* respond to competitive, co-cultured environments by activating BGCs of uncharacterized NPs [82]. Mass spectrometry-based proteomics, while less common in NP discovery, validates predicted BGC enzymes and identifies critical protein-protein interactions [83] and post-translational modifications (PTMs) needed for activity [84]. It may also be used to directly reveal a BGC itself without the need for a genome *a priori*. One method, known as Proteomic Investigation of Secondary Metabolism (PrISM), was developed to detect PKSs and NRPSs whose primary amino

acid sequences were then used to design primers that would amplify the corresponding gene cluster from the genome [85]. While the clusters identified in this study were from *Bacillus* and *Streptomyces*, this emerging method for NP discovery may also prove beneficial in detecting fungal BGCs as well. Currently, efforts to integrate metabolomic, transcriptomic, and proteomic data are elucidating NP pathways and revealing strategies on how to optimize yields [86], [87]. Ultimately, once a product has been successfully identified, it must be characterized with analytical biochemical techniques including HPLC or column chromatography for purification followed by NMR structural characterization [88].

With the rapidly expanding sequence data of the post-genomics era, advances in synthetic biology approaches for gene activation, and new inexpensive –omics tools to detect NP formation and gene activation, researchers have begun to create systematic analytical pipelines to discover fungal natural products (Figure 2-2). These new pipelines offer the advantage of being sensitive, and can potentially elucidate the NP biosynthetic pathway in a single growth experiment by integrating genomic, transcriptomic, proteomic, and metabolomics analyses. One example of this pipeline is the European QuantFung program [89] launched in 2013 to discover and characterize fungal secondary metabolites. While relatively new to NP discovery, these integrated –omics approaches have been demonstrated in unconventional fungal systems to elucidate new fungal enzymes [18].

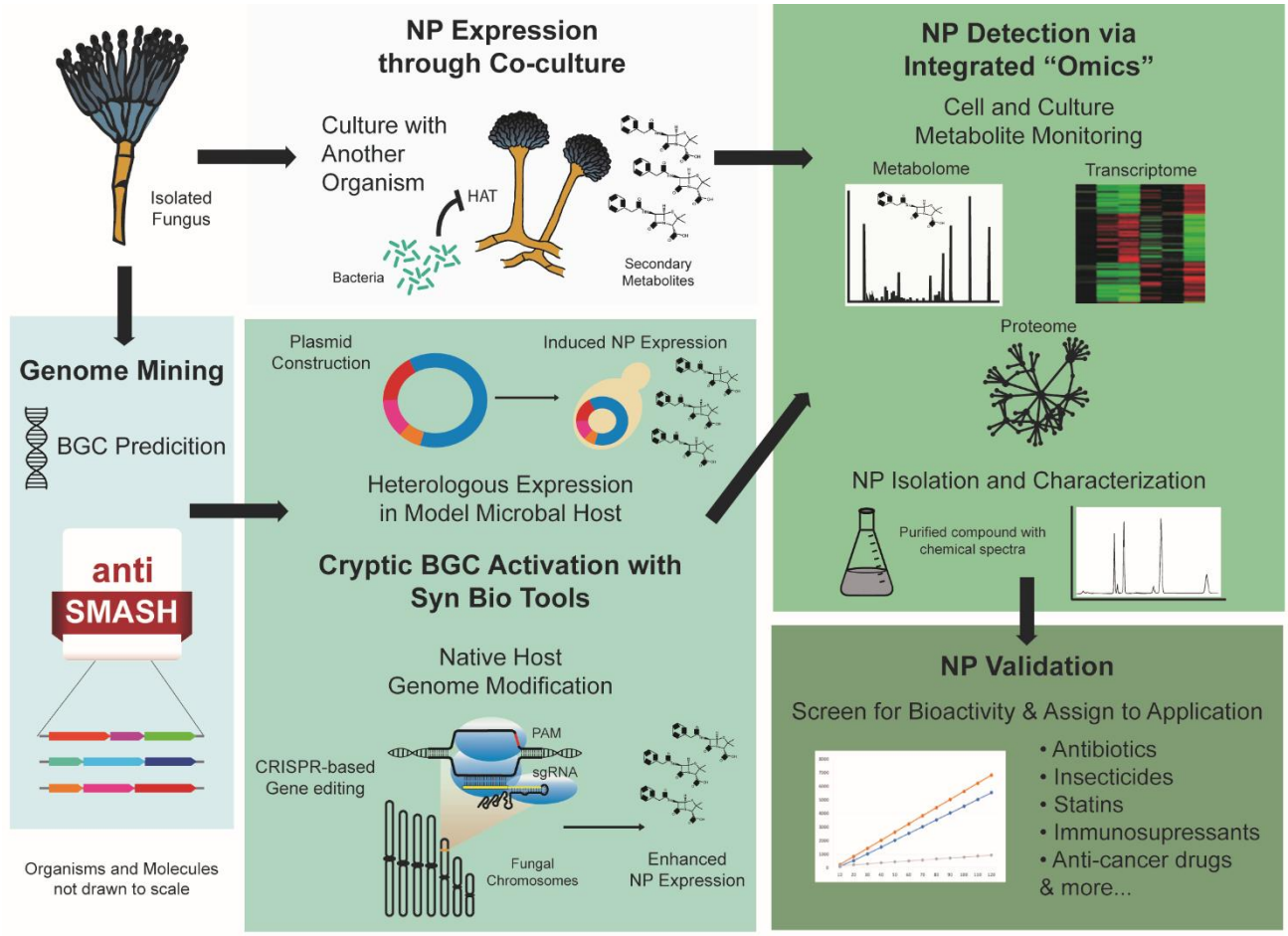


Figure 2-2: Fungal Natural Product Discovery Pipeline. Natural products (NPs) from isolated fungi have two common pipelines through which they are expressed. The first pipeline is to express through co-culture with bacteria which dysregulates fungal histone acetylation to enhance NP expression [74]. The second pipeline sequences genomes and predicts biosynthetic gene clusters (BGCs) with bioinformatics. After BGC prediction, CRISPR-based tools or heterologous expression are used to activate expression of the BGC the native or model organism, respectively. Ultimately, the metabolites, transcripts, and proteins from both pathways can be detected and quantified with the same integrated omics pipeline. Finally, novel natural products of interest are screened for biological activity; HAT: Histone Acetyl Transferase.

2.6 Conclusions

Fungal systems remain a vastly underrepresented, and untapped reservoir for NP discovery. Although fungal natural products have been studied heavily over the past century, only recently has fungal biotechnology diversified beyond filamentous fungi and yeasts. However, relatively few nonconventional fungi have been cultivated and even fewer examined for NPs. Yet, they remain rich targets with more than 700 natural products identified from 2005 to 2010 in

uncultivated marine fungi alone [90]. Similarly, Neocallimastigomycota [91], [92], a phylum of biomass degrading anaerobic gut fungi found in herbivore GI tracts [18], [35], [93], has been proposed as an exciting new source for NPs due to an abundance of BGCs identified in recent genomic sequencing surveys [35], [94]. Similar discoveries in biomass deconstruction enzymes for biofuels and advances in sequencing technologies have launched a 'gold rush' to sequence new and unconventional fungi. At the forefront of this effort is the U.S. Department of Energy's (DOE) Joint Genome Institute (JGI) who produces high-quality DNA sequencing, synthesis, and analysis to support DOE missions in bioenergy and environmental processes. Due to the critical role of fungi as natural recyclers and decomposers, the JGI is completing the 1000 Fungal Genome Project, which promises to reveal exciting new enzymes and BGCs [95], [96]. As the potential of fungi becomes increasingly recognized and the number of studied fungal systems continues to grow, so will the opportunities for NP discovery.

Fungal natural products are attractive molecules for biotechnology. Their rich chemical structure and bioactivity have historically given rise to many new drugs, and may hold the key to addressing emerging antimicrobial resistance. Similarly, the biochemistry that gives rise to this chemical diversity is equally promising as it creates tailored hydrocarbon rich compounds. The modularity of these biosynthetic pathways, and new synthetic biology tools to repurpose these enzymes have enabled the biological synthesis of powerful non-natural compounds including advanced biofuels [97]–[99]. As these technologies mature and new fungi are isolated and characterized, newly discovered fungal natural products and their derivatives are certain to form integral portions of the bioecnonomy.

3. HYDROLYSIS OF UNTREATED LIGNOCELLULOSIC FEEDSTOCK IS INDEPENDENT OF S-LIGNIN COMPOSITION IN NEWLY CLASSIFIED ANAEROBIC FUNGAL ISOLATE, *PIROMYCES SP. UH3-1*

This chapter was adapted from Hooker, Hillman et al. *Biotechnology for Biofuels* 11(1) (2018):293 [14] and was incorporated in part in the MS thesis of C Hooker [100]. My unique contributions to this work are sections 3.6.1. - 3.6.4. However, I reproduce the text in its entirety here for full context and ease of interpretation

3.1 Abstract

Plant biomass is an abundant but underused feedstock for bioenergy production due to its complex and variable composition, which resists breakdown into fermentable sugars. These feedstocks, however, are routinely degraded by many uncommercialized microbes such as anaerobic gut fungi. These gut fungi express a broad range of carbohydrate active enzymes and are native to the digestive tracts of ruminants and hindgut fermenters. In this study, we examine gut fungal performance on these substrates as a function of composition, and the ability of this isolate to degrade inhibitory high syringyl lignin-containing forestry residues. We isolated a novel fungal specimen from a donkey in Independence, Indiana, United States. Phylogenetic analysis of the Internal Transcribed Spacer 1 sequence classified the isolate as a member of the genus Piromyces within the phylum Neocallimastigomycota (*Piromyces sp. UH3-1*, strain UH3-1). The isolate penetrates the substrate with an extensive rhizomycelial network and secretes many cellulose-binding enzymes, which are active on various components of lignocellulose. These activities enable the fungus to hydrolyze at least 58% of the glucan and 28% of the available xylan in untreated corn stover within 168 h and support growth on crude agricultural residues, food waste, and energy crops. Importantly, UH3-1 hydrolyzes high syringyl lignin-containing poplar that is inhibitory to many fungi with efficiencies equal to that of low syringyl lignin-containing poplar with no reduction in fungal growth. This behavior is correlated with slight remodeling of the fungal secretome whose composition adapts with substrate to express an enzyme cocktail optimized to degrade the available biomass. Piromyces sp. UH3-1, a newly isolated anaerobic gut fungus, grows on diverse untreated substrates through production of a broad range of carbohydrate active enzymes that are robust to variations in substrate composition. Additionally, UH3-1 and

potentially other anaerobic fungi are resistant to inhibitory lignin composition possibly due to changes in enzyme secretion with substrate. Thus, anaerobic fungi are an attractive platform for the production of enzymes that efficiently use mixed feedstocks of variable composition for second generation biofuels. More importantly, our work suggests that the study of anaerobic fungi may reveal naturally evolved strategies to circumvent common hydrolytic inhibitors that hinder biomass usage.

3.2 Background

Lignocellulosic material is an inexpensive and abundant source of carbon that remains underexploited for biofuel production due to its complex heteropolymeric structure that hinders release of fermentable sugars by lignocellulolytic enzymes [101]. Available plant biomass for bioenergy is greatly dependent on geographic location and climate variability, leading to large differences in the types and compositions of the potential substrates [102]. More importantly, the biomass composition strongly affects the performance of a given enzyme cocktail [103]. As a result, the enzyme cocktails that are used to hydrolyze these feedstocks are optimized for individual substrates, and are not suitable for more economically-viable feedstock streams whose composition fluctuate greatly with market availability [104]. As enzyme cost is a significant bottleneck to the development of economical biofuels, enzyme systems that display superior performance on diverse feedstocks would advance the economic feasibility of bioenergy [9], [10], [105].

Current lignocellulolytic enzymes systems are based on well-known fungi such as *Trichoderma reesei* and *Aspergillus spp.* due to their oversecretion of many glycoside hydrolyases (CAZymes), which are active on the glycosidic bonds of lignocellulosic materials [106]. However, these species do not naturally express all of the enzymes needed to fully hydrolyze the sugars contained in plant biomass [107]. For example β-glucosidases in *T. reesei,* an enzyme essential to release the free glucose, form less than 1% of all secreted CAZymes [106]. Thus, enzyme cocktails based on *T. reesei* must be supplemented with enzymes from other species for sufficient activity [108]. The need for cocktail supplementation with enzymes from various species greatly increases enzyme production costs due to capital-intensive parallel enzyme production processes [105],

[109]. Therefore, a single species enzyme platform would simplify enzyme production and reduce cost.

Degradation of untreated biomass is common in many underexplored environments that may harbor efficient microbial enzymes for biofuels. One example is the rumen and hindgut of large herbivores where grasses, shrubs, and other untreated fiber-rich plant biomass are processed daily by a consortium of microbes including early-divergent Neocallimastigomycota (anaerobic fungi) [110]. While anaerobic fungi are known to harness powerful biomass-degrading enzymes, the ability of these enzymes to hydrolyze diverse plant biomass remain poorly characterized [18]. To date, only five specimens in this phylum have been sequenced and studied in any detail [33]. The fungi of Neocallimastigomycota thrive under mild conditions (pH ≈ 7, 39 ºC), and possess large arrays of CAZymes that efficiently degrade untreated plant biomass [18], [91]. However, there is little data on the extent of the cellulosic and xylanosic degradation by these enzymes across a range of lignin compositions.

Given the potential for anaerobic fungi to reduce enzyme production costs, we sought to characterize their enzymatic performance as a function of substrate composition. Here, we report the isolation and taxonomic placement of a novel species of anaerobic gut fungi (*Piromyces sp. UH3-1*) in the Neocallimastigaceae family. We characterize the ability of *Piromyces sp. UH3-1* to degrade and grow on an array of untreated substrates (e.g. corn stover, switchgrass, orange peel, and sorghum) under mild conditions. Additionally, we measure the free sugars released from untreated poplar across a range of lignin compositions to estimate fungal enzyme performance with feedstock composition. This work suggests that anaerobic fungal enzymes are robust for hydrolysis of diverse untreated lignocellulose and are promising new candidates for lignocellulosic enzyme production.

3.3 Results

3.3.1 Isolation of a biomass degrading anaerobic gut fungus from a donkey

To identify more robust and efficient CAZymes and microbial systems that may be used for bioenergy applications, we isolated a previously uncharacterized microbe from the fecal samples of a donkey. Light microscopy revealed the presence of non-planktonic microorganisms that grew invasively into the plant substrates, reminiscent of a mature fungal sporangium (Figure

3-1A), after three to four days of growth with a simultaneous increase in headspace pressure. This isolatedorganism was cultured to axenic purity by repeated passage through roll tubes containing multipleantibiotics (see Methods, Figure 3-1B). Further microscopic analysis revealed that this organism produces zoospores with a single flagellum (~30 μm long) (Figure 3-1C), another key characteristic of the genus *Piromyces* of the fungal phylum Neocallimastigomycota. Additionally, this isolate exhibits endogenous zoosporangial development, where the zoosporangium retains its nuclei. The slow growth, zoospore presence, and well-differentiated stages of a life-cycle (Figure 3-1C-E) suggested a fungal specimen. DAPI staining of the nucleic acid in the developing fungal sporangium revealed that this isolate was monocentric (has nuclei only within the zoosporangium) (Figure 3-1E), which is also consistent with the morphology of the fungal genus *Piromyces*

Taxonomic classification of our novel fungal isolate was confirmed via phylogenetic analysis [111], [112]. Amplification of the 16s rRNA genes failed, while amplification of the ITS1, ITS2, and the 28s rRNA large ribosomal subunit (LSU) were all successful (Appendix B: Figure B.1) [113]. Therefore, the isolate was definitively fungal in origin, rather than bacterial or archaeal, which agrees with our morphological assessment. We aligned these amplicons against 51 Genbank deposited anaerobic fungal sequences (division Neocallimastigomycota) and confirmed that our isolate formed a distinct branch within the fungal *Piromyces* genus (Figure 3-2A-B) [111], [114]. The monocentric thallus and uniflagellated zoospore, both characteristic of *Piromyces* fungi, further support this placement (Figure 3-1E-F) [49].

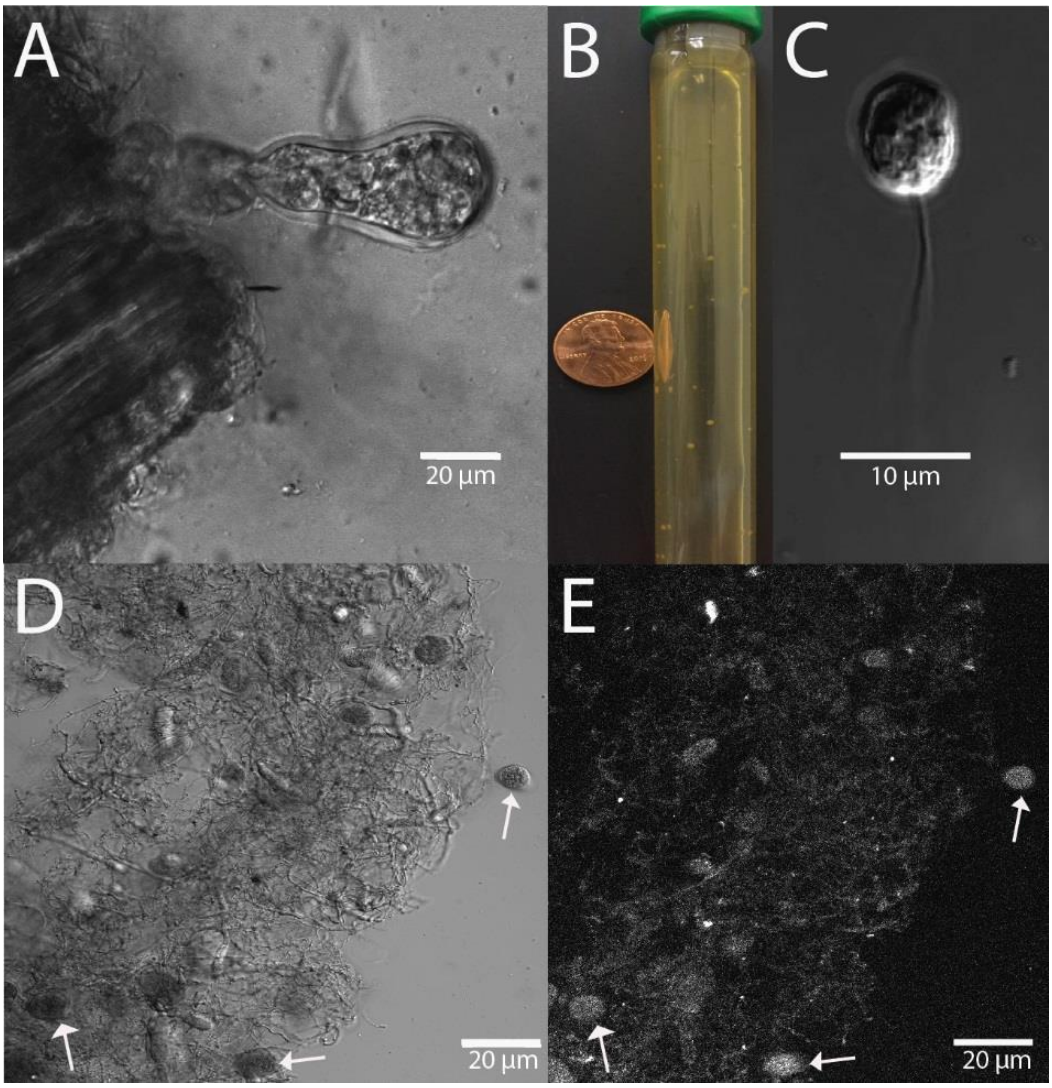


Figure 3-1: The host and life cycle of *Piromyces sp. UH3-1*. A) Individual mature sporangia on corn stover (left) displaying ovoid structure. B) Roll tube used to isolate individual axenic cultures of *Piromyces sp. UH3-1*. C) Uniflagellated zoospore of *Piromyces sp. UH3-1* imaged after zoospore death D) Multiple sporangia, demonstrating the predominantly spherical to ovoid structure; arrows indicate individual sporangia in rhizomycelial network. E) DAPI stain indicating the monocentric nature as zoosporatic nuclei are contained with the sporangia [14].

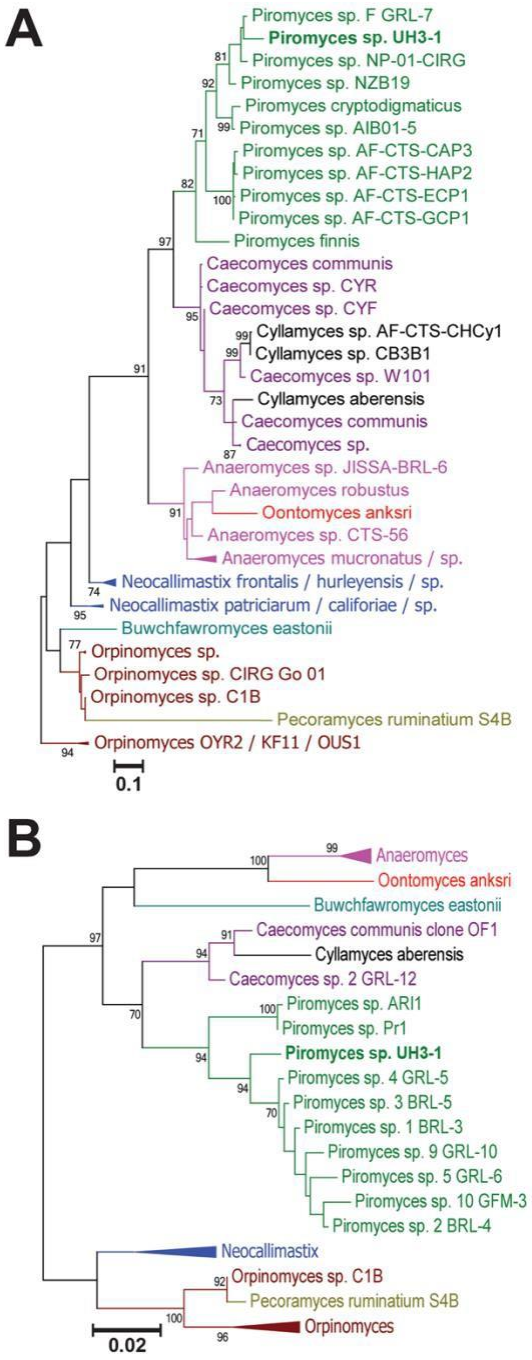


Figure 3-2: Phylogenetic analyses place our isolate within the genus *Piromyces*. A) Collapsed ITS1 phylogenetic tree and B) Collapsed LSU phylogenetic tree. Fully expanded phylogenetic trees displaying the Genbank accession numbers are in Appendix B: Figures B.2-B.3. Significantbootstrap values from 1000 iterations are indicated to the left of each branch [14].

This organism represents a novel cultured species as it has less than 90% BLAST similarity to known cultured species of anaerobic fungi. Therefore, we classify this organism as the species *Piromyces sp. UH3-1* (NCBI Taxon ID: KY494854, JRMC: SF:012426, Index

Fungorum: IF554555) and (Appendix B: Formal Species Description), in honor of the state in which it wasisolated (Indiana, United States).

3.3.2 Anaerobic fungi degrade complex substrates with efficiencies comparable to glucose

Untreated lignocellulosic substrates are rich in sugars that can sustain fungal growth; however, the degradation rate of these substrates into free sugars is frequently limiting for growth. Thus, to estimate hydrolysis efficiency, we assessed the ability of *Piromyces sp. UH3-1* to grow on agricultural residues, bioenergy crops, food wastes, and forestry products that had not undergone pretreatment (Appendix A: Tables A.1-A.5). Anaerobic fungi secrete an array of CAZymes that break down diverse lignocellulosic material into fermentable sugars that the fungus metabolize to CO_2 and H_2, among other fermentation products such as lactate, formate, acetic acid, and ethanol [115]. Anaerobic gut fungi grow invasively into plant substrates forming plugs that trap the fermentation gasses leading to more buoyant floating cultures (Figure 3-3A). However, when grown on soluble substrates, the fungi grow into themselves to form a mat of biomass (Figure 3-3A). Gas accumulation is proportional to fungal biomass production and may be used as a convenient indicator of growth (Figure 3-3B) [116]. Both pressure accumulation, and visual analysis were used to assess the ability of *Piromyces sp. UH3-1* to grow on these feedstocks. While the growth rates for these substrates varied significantly, the total pressure accumulations were comparable for lignocellulosic substrates (Figure 3-3C-D). Therefore, these results suggest that this organism secretes an array of CAZymes that liberate sufficient sugars, regardless of feedstock composition, to sustain fungal growth into stationary phase.

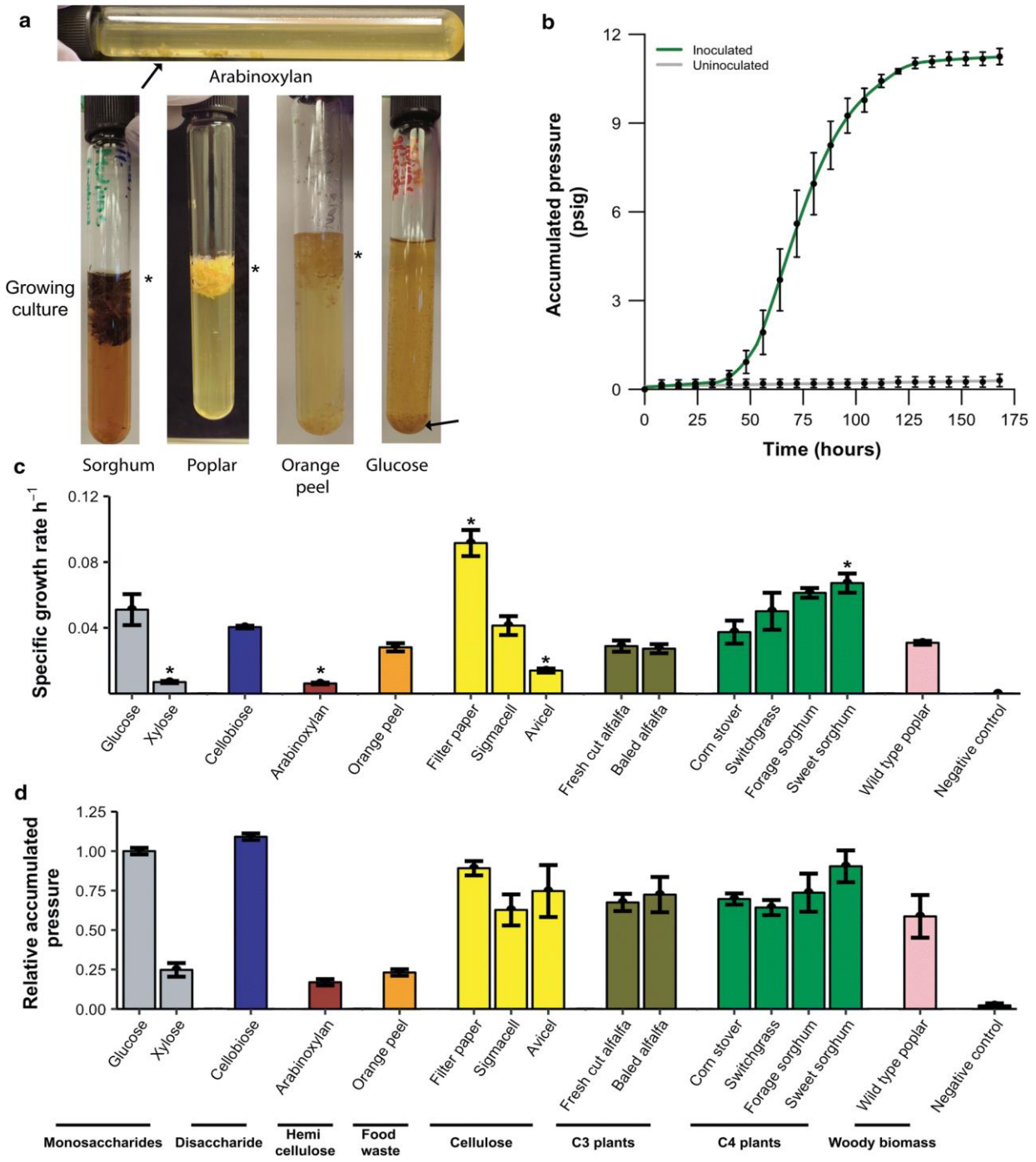


Figure 3-3: *Piromyces sp. UH3-1* grows on diverse feedstocks. A) Growth of *Piromyces sp. UH3-1* on soluble substrates leads to colony formation on the walls of the tubes (arrows indicating colony formation). Fungal cultures growing on lignocellulosic substrates float up during fermentation. B) A representative growth curve of *Piromyces sp. UH3-1* on corn stover. C-D) *Piromyces sp. UH3-1* degrade and proliferate on a wide array of untreated agricultural wastes, bioenergy feedstocks, and forestry wastes. All accumulated pressures are normalized toglucose. Asterisks denote statistically significant differences in specific growth rate relative to glucose ($p < 0.05$, unpaired t-test) [14].

To determine whether biomass hydrolysis was efficient or limiting for growth, we first established a baseline for growth on simple sugars. Glucose led to robust growth, (Figure 3-3, Appendix B: Figure B.4), and was used as a baseline to which all other substrates were compared. Similarly, the disaccharide cellobiose led to strong, robust fungal growth, suggesting that anaerobic fungi readily produce β-glucosidases that can cleave cellobiose to glucose at a rate in excess of glucose uptake and metabolism (Figure 3-3, Appendix B: B.4). In contrast, fungal growth on hemicellulosic components such as xylose and arabinose (Figure 3-3, Appendix B: Figure B.4, B.7-B.8) led to inconsistent pressure accumulation and a significantly reduced growth rate relative to glucose ($p = 0.0147$, unpaired t-test). Nonetheless, accumulation of fungal biomass on xylose was consistently observed (Appendix B: Figure B.8). Thus, xylose transport and incorporation into central metabolism likely occurs more slowly than six carbon sugars and may be limiting for growth. Taken together, these results suggest that this fungal isolate grows primarily on hexose sugars, and has robust β-glucosidase activity that is not a bottleneck for biomass hydrolysis, unlike *T. reesei* [117]. While fungal growth on hemicellulose components is poor, it must still remove hemicellulose and other carbohydrate polymers to access the glucose-rich cellulosic portions of lignocellulose. Arabinoxylan, a form of hemicellulose, contains fermentable arabinose and xylose sugars, and is highly abundant in the cell walls of cereals and grasses used as bioenergy crops [118]. Similarly, pectin is a complex and variable component in the middle lamella between the plant cell walls. As this surrounds the energy rich cellulosic and hemicellulosic polymers, pectin removal or deconstruction is advantageous for efficient lignocellulose hydrolysis [119]–[121]. The growth of *Piromyces sp. UH3-1* on wheat arabinoxylan and pectin rich feedstocks such as orange peel, while consistent, was unlike typical microbial growth and non-sigmoidal in nature (Figure 3-3, Appendix B: Figure B.4). However, when the sugar monomers rhamnose and galacturonic acid were tested, pressure accumulation was irregular and suggested that these substrates could not be metabolized fast enough to support fungal growth (data not shown). Thus, the degradation products of pectin, and to a lesser extent arabinoxylan are unlikely to sustain robust growth. Given the poor growth on these polymeric substrates, we directly analyzed their hydrolysis by collecting the fungal secretome and testing for CAZyme activity.

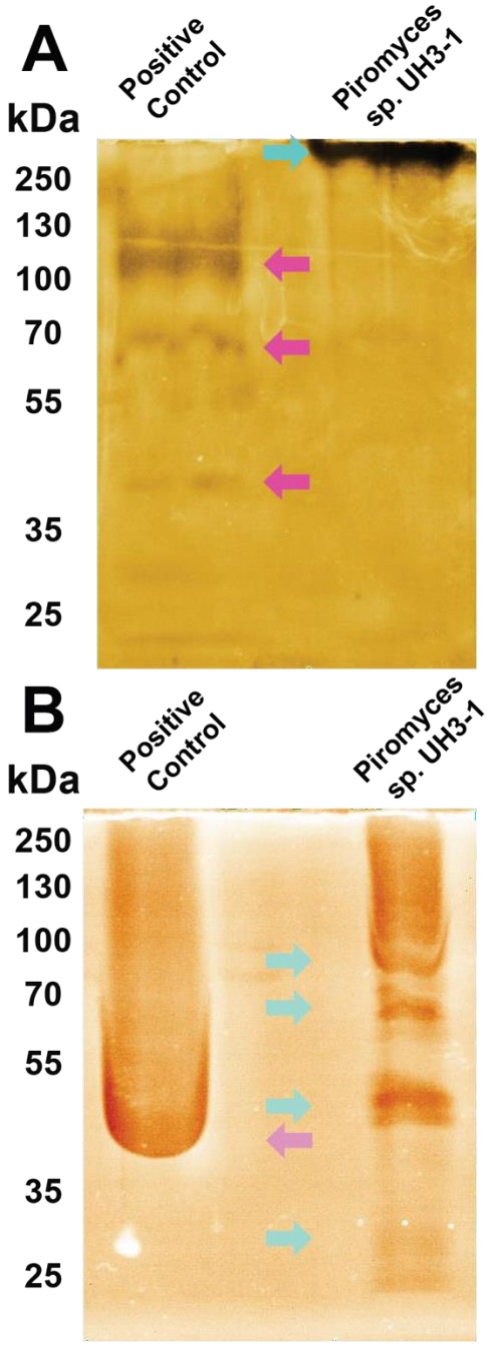


Figure 3-4: *Piromyces sp. UH3-1* secretes diverse CAZymes for degrading the polymers of lignocellulose. A) A pectin zymogram shows strong pectinolytic activity for *Piromyces sp. UH3-1* at the top of the gel (teal arrow), while *Aspergillus* shows multiple bands having pectinolytic activity (pink arrows). B) A carboxy methyl cellulose zymogram shows distinct cellulolytic activity for multiple proteins of *Piromyces sp. UH3-1* (teal arrows), while *Aspergillus* (Viscozyme, positive control) shows high cellulolytic activity (Pink arrow). Controls and experimental samples were loaded with the same total protein mass as measured by a Bradford assay [14]

By isolating the fungal enzymes we were able to test their activity via zymography, which exploits the ability of some stains to preferentially bind to polysaccharides (Figure 3-4) [122]. Differential staining around individual protein bands results from the consumption of substrate and is positive for hydrolytic activity. Pectin zymograms show a high molecular weighthydrolysis zone indicating that this fungal isolate can degrade this complex polymer (Figure 3-4).Similarly, reducing sugar assays reveal strong xylanolytic activity from anaerobic fungal secretedproteins (Appendix B: Figure B.9). Thus, while *Piromyces sp. UH3-1* is unable to efficiently metabolize these substrates, it still expresses an array of CAZymes that break down the pectin and hemicellulose components of lignocellulose under mild conditions.

Readily degrading cellulose is critical to efficiently producing energy from renewable plant biomass [3], [123]. Given the variability in cellulose structure between plant sources and preprocessing before enzymatic hydrolysis occurs (e.g. degree of crystallinity, porosity, and specific surface area), we evaluated the efficiency of cellulose hydrolysis by testing three different substrates, which all yielded robust fungal growth (Figure 3-3, Appendix B: Figure B.4) [124]–[126]. Sigmacell from cotton linters and filter paper yielded growth rates that were equal to or in excess of growth on glucose suggesting that cellulase activity is not limiting for growth on lower crystallinity substrates. In contrast, growth on Avicel, a highly crystalline cellulose produced by acid hydrolysis of wood pulp, was reduced by 65% ($p = 0.0268$, unpaired t test), likely due to inhibition from the high crystallinity and reduced surface area caused by settling and packing of the substrate in these stationary fermentations [125]. Counterintuitively, growth on filter paper was faster than on glucose ($p = 0.0023$, unpaired t test). Despite these differences in growth rate, the total accumulated pressures were comparable, suggesting similar levels of carbon use, and thus sugar release, by the fungus independent of substrate crystallinity (Figure 3-3D). We sought to further characterize these cellulases by testing their activity through zymography (Figure 3-4B). Through this analysis, we identified multiple cellulose-binding proteins having cellulolytic activity. Taken together, these results suggest that this fungal isolate efficiently degrades cellulose by expressing multiple cellulases that have high activity in excess of glucose uptake and metabolism. *Piromyces sp. UH3-1* robustly grew on untreated lignocellulosic feedstocks, regardless of composition or photosynthetic type (Figure 3-3). Photosynthetic type (C3 or C4) leads to significant differences in cell wall structure, and thus the CAZymes needed to degrade the lignocellulose [127]. For C3 plants, we tested untreated alfalfa (*Medicago sativa)*, which resulted

in strong fungal growth (Figure 3-3, Appendix B: Figure B.4). Commonly available C4 feedstocks for biofuel production such as corn stover (*Zea mays*), switchgrass (*Panicum virgatum*), and sorghum (*Sorghum bicolor*) were consistently degraded by *Piromyces sp. UH3-1* (Figure 3-3B-D, Appendix B: Figure B.5). Several varieties of sorghum, with differing cell wall compositions, were tested as they thrive in different climates and are planted in specific regions, unlike the other tested C4 feedstocks [128]–[130]. Notably, sweet sorghum was the only lignocellulosic substrate that yielded a significantly higher growth rate when compared to glucose ($p = 0.0212$, unpaired t test), possibly due to the excess free sugars common in sweet sorghum [128]. Thus, these results suggest that cell wall composition of untreated lignocellulose does not significantly reduce fungal growth rate, implying that the CAZymes of *Piromyces sp. UH3-1* efficiently degrade these substrates.

3.3.3 Anaerobic fungal hydrolytic enzymes are robust to lignin composition

Woody biomass such as poplar has been proposed as a feedstock for second generation biofuel production as it is a fast-growing tree species capable of thriving in diverse geographic locations, has high biomass yields, and high glucan content (>40%) relative to other commonly used feedstocks (Appendix A: Tables A.2-A.5) [3], [131]–[134]. Furthermore, poplar can be grownon land that is marginally productive for most agricultural crops [135]. However, the lignin in poplar that has not undergone pretreatment is known to strongly affect cellulase and hemicellulaseactivity [136]. Despite this, *Piromyces sp. UH3-1* still showed strong growth on wild type poplar (Figure 3-3, Appendix B: Figures B.4-B.6). This result is consistent with published data as anaerobic fungi are known to degrade untreated woody biomass [137]. However, as lignin composition may change for diverse feedstocks, we tested fungal growth on transgenic lines of poplar containing varying ratios (5%-98%) of Syringyl (S)-lignin (Figure 3-5, Appendix A: TablesA.2-A.5, Appendix B: Figures B.5-B.6). [138], [139]. S-lignin content is known to reduce the growthof some fungi by as much as 80% [17].

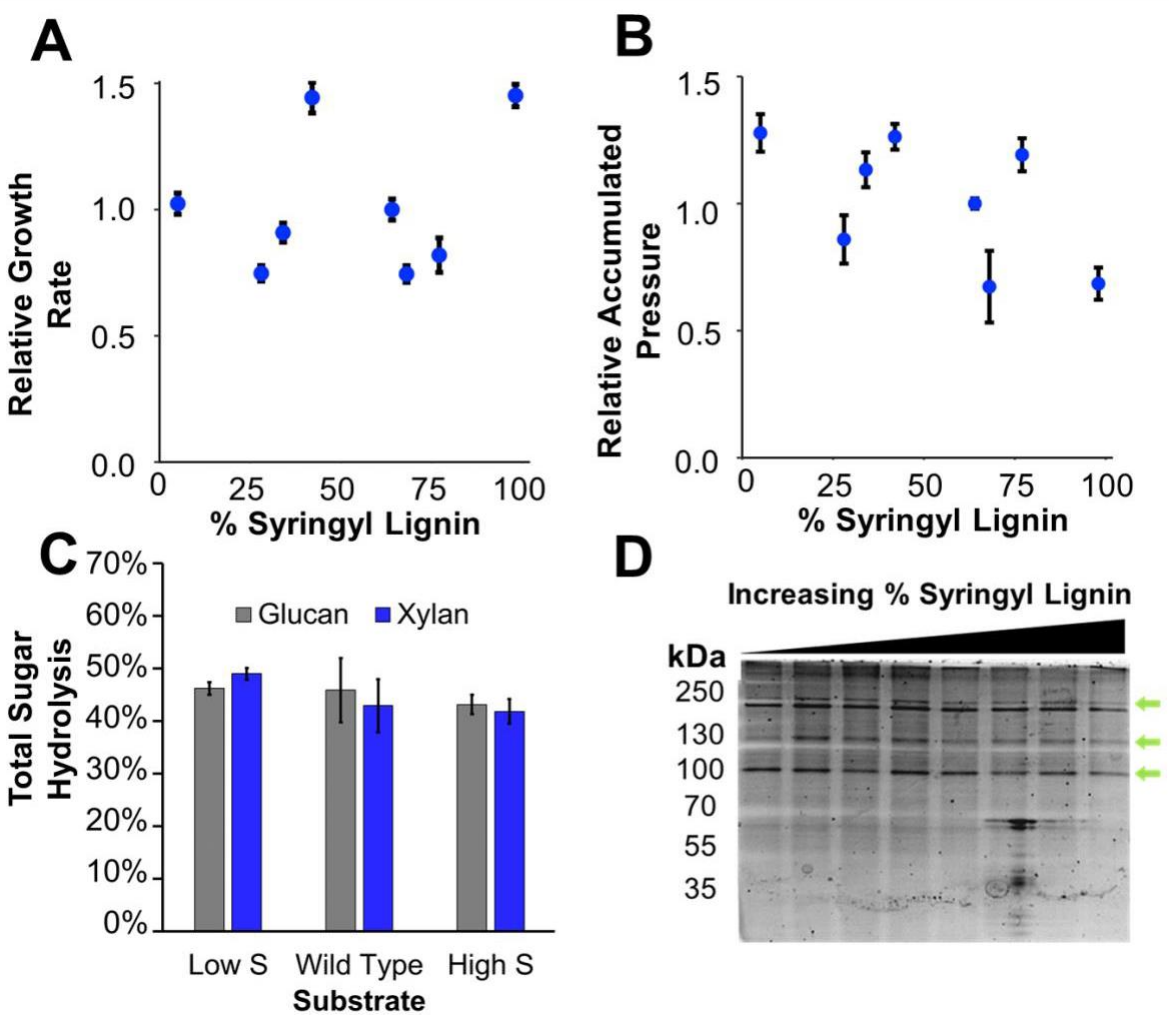


Figure 3-5: *Piromyces sp. UH3-1* growth and sugar degradation is robust against lignin composition with optimal enzyme expression. A) Relative growth rates of *Piromyces sp. UH3-1* on genetically modified lines of poplar relative to wild type INRA 717 (64% S-lignin), (p = 0.0317, R^2 = 0.1715). B) Relative fungal biomass accumulations of *Piromyces sp. UH3-1* on genetically modified lines of poplar relative to wild type INRA 717 (p = 0.0011, R^2 = 0.2991). C) Minimum hydrolysis percentages on three of the lines of poplar [72, 74] D) The carbohydrate binding portion of the fungal secretome shows changes in response to S-lignin composition (green arrows)[14].

Our fungal isolate was insensitive to S-lignin content and degraded both S-lignin rich andpoor substrates with high efficiency. Both growth rate and fungal biomass accumulation appearedto be independent of S-lignin content (Figure 3-5A-B). While an ANOVA analysis of this data yielded statistically significant trends for relative growth rate (p = 0.0317), and for relative fungal biomass accumulation (p = 0.0011), this correlation was weak with R^2 values of 0.1715 and 0.2991,respectively. To further test these results, we repeated this experiment with

another batch of poplar with S lignin molar percentages ranging from 20 to 93 percent (Appendix A: Tables A.4-A.5, Appendix B: Figure B.10). Further, this poplar was only milled to 20 mesh, while the above results were milled to 40 mesh (~0.420 mm). Notably, the total fungal biomass accumulation was insensitive to S lignin for these constructs despite having larger particle sizes (Appendix B: Figure B.10) R^2= 0.00817, p = 0.64723. To further evaluate the degradation of polymeric sugars in the presence of varying S-lignin compositions from the 2014 harvested poplar, we grew *Piromyces sp. UH3-1* on three different poplar constructs and measured the hydrolysis of polymeric sugars tomonomeric sugars (Figure 3-5C). As fungal rhizomycelia penetrate the plant material, it iscurrently not possible to distinguish fungal from plant biomass and accurately measure biomass loss, and thus total sugar consumption. However, an analysis of the glucan and xylan contents of spent and fresh poplar biomass, with the conservative assumption that total plant biomass is constant, suggests that *Piromyces sp. UH3-1* metabolizes at least 43% of the glucose sugars, and 42% of the pentose sugars within 168 hours (Appendix A: Table A.6). These results are consistent with those reported for an isolate of *Neocallimastix*, a different genera of fungi also within *Neocallimastigomycota* [137]. Specifically this isolate released glucan and xylan at efficiencies of47% and 34% respectively on untreated poplar after 11 days of growth [137]. Notably, glucan release was independent of S-lignin composition in the poplar constructs tested (wild type vs highS-lignin, p = 0.6499; wild type vs low S-lignin, p = 0.9951). There was also no significant difference in glucan release between low S-lignin and high S-lignin constructs (p = 0.5945). Similar trends were also observed for xylan release (wild type vs high S, p = 0.9105; wild-type vslow S-lignin, p = 0.1308; high S vs low S constructs, p = 0.0771). Taken together, these results suggest that these anaerobic fungal enzymes are robust against inhibitory syringyl lignin content and hydrolyze glucan and xylan in untreated lignocellulose with similar efficiency regardless of lignin composition. More importantly, while there are no known mechanisms by which anaerobicfungi can metabolize lignin constituents, our results suggest that fungal pathways may exist to recognize lignin composition to increase S lignin resistance. To analyze this, we collected the cellulose-binding portion of the fungal secretome after growth on modified poplar lines (Figure 3-5D). The relative concentration of several proteins changed non-linearly with S-lignin content suggesting a complex response to combat S-lignin recalcitrance. More importantly, while there are no known mechanisms by which anaerobic fungi can metabolize

lignin constituents, our results do suggest that fungal pathways exist to recognize lignin composition to regulate secretion of enzymes that perhaps trade hydrolytic activity for increased S-lignin resistance.

3.4 Discussion

Producing biofuels from lignocellulose that are competitive with current energy technologies requires more efficient use of existing biomass reserves in processes that incorporatemultiple feedstocks of variable composition. Increasing the number of potential feedstocks will help to protect second generation platforms from changing production conditions that may result due to inconsistency in plant biomass yield, climate variability, and market volatility. One way tomove toward this goal is by pretreating plant biomass, which has traditionally been used to overcome lignin inhibition. While pretreatment helps to mitigate these issues, new waste streams are introduced, toxic inhibitors are released that hinder the growth of the fermenting organism, andhigher enzyme loadings are required [123], [140]–[142]. A more promising strategy is, thus, to identify enzyme platforms that readily degrade diverse untreated lignocellulose and are robust to variations in biomass composition. However, for this to be industrially economical, high fermentable sugar conversions are a necessity. Key challenges include product inhibition of the cellulases and the release of lignin among other contaminants that can inactivate the secreted enzymes. Despite these barriers unengineered anaerobic gut fungi such as *Piromyces sp. UH3-1* show strong conversions on untreated plant biomass (Figure 3-5). When grown on milled (~0.5 mm), untreated corn stover, *Piromyces sp. UH3-1* converts at least 58 and 28 percent of the available glucan and xylan, respectively (Appendix A, Table A.6). These values are comparable to current commercial enzyme cocktails which release 48 and 30 percent of the available glucan and xylan, respectively, on ball-milled (~100 micron) corn stover [143]. While hydrolytic rates were not measured as the secretome of *Piromyces sp. UH3-1* is a crude preparation of lignocellulolytic enzymes and other unrelated proteins that are released over time, similar studiesof anaerobic fungi suggest that our observed glucan conversion are not limiting and would improvewith increased enzyme loading and or time [13]. In contrast, conversion with current cocktails saturate at these conversions unless supplemented with additional enzymes [144], [145]. Thus, the ability of *Piromyces sp. UH3-1* to degrade these untreated feedstocks without additional

supplementation at comparable efficiencies to commercial cocktails, supports further study of these CAZymes, which may provide next generation solutions to critical issues with lignocelluloserecalcitrance.

Engineering fungi for enhanced enzyme production has been a subject of considerable research [107], [146]. Both *Aspergillus* and *Trichoderma* are widely used to produce industrial enzyme cocktails, yet these organisms are not the strains that were originally isolated given their natural deficiencies. [147], [148]. For example, *Aspergillus* is known to express low amounts of endoglucanases, which is critical for efficiently degrading cellulose [149]. Additionally, the QM6Astrain of *Trichoderma reesei* (previously *T. viride*) that was initially isolated went through multiplerounds of mutagenesis to obtain the hyper-producing, catabolite repression resistant strain Rut- C30 that is the basis for commercial enzyme production [117]. Similar to the original *Trichoderma*QM6A, fungi of *Neocallimastigomycota* are known to have catabolite repression that directly represses CAZyme expression [18], [147]. Despite the presence of catabolite repression, gut fungi still robustly degrade untreated lignocellulose. Manipulating anaerobic gut fungi, through mutagenesis or genome engineering would likely lead to improved conversions and make anaerobic fungal enzymes more competitive with current commercial formulations. Similarly, further analysis of how these fungi alter CAZyme expression for diverse untreated lignocellulosemay identify new enzymes optimized for certain classes of feedstocks that could be exploited for efficient bioenergy production. However, full-scale industrial exploitation will also require the development of new technologies to cultivate anaerobic fungi at large scale, and may be energetically limited by the inherent anaerobic nature of such processes.

3.5 Conclusions

In this work, we present the isolation, taxonomic placement, and characterization of a novel speciesof anaerobic gut fungus. We tested fungal growth on diverse untreated feedstocks to estimate the full range of CAZyme activities, and their ability to degrade plant biomass at rates sustainable forfungal growth. *Piromyces sp. UH3-1* thrives on an array of untreated agricultural residues and bioenergy crops by hydrolyzing and fermenting the cellulosic and hemicellulosic fractions of these substrates. Importantly, we show for the first time that anaerobic fungi, such as this isolate, grow and release sugars to similar efficiencies regardless

of lignin composition. Thus, this study not only highlights the ability of unengineered gut fungi to degrade diverse untreated lignocellulose, but also suggests that novel adaptations to overcome compositional variability may exist. Characterizing these adaptations and isolating the responsible enzymes may lead to more efficientenzyme cocktails that can more fully use available renewable biomass for lignocellulosic biofuelproduction

3.6 Methods

3.6.1 Isolating a novel species of anaerobic gut fungi

We suspended fresh donkey feces in Hungate tubes containing sterile anaerobic medium Csupplemented with 15% clarified rumen fluid (150 ml: Bar Diamond Inc., Parma, ID, USA) under 100% CO_2 headspace [150]. Suspensions of donkey feces were serially diluted 1000-fold and used as a 10% inoculum in Hungate tubes containing 9 ml anaerobic medium C, supplemented with switchgrass as a carbon source (1% w/v) and chloramphenicol (25 µg/ml; Fisher Scientific, Waltham, MA, USA). After inoculation, the cultures were incubated at 39 ˚C for 72-96 hours.

To obtain axenic cultures, we inoculated roll tubes with liquid fungal culture and propagated individual colonies. Roll tubes were prepared by adding agar (2% w/v), glucose (0.45% w/v), and chloramphenicol (25 µg/ml) to anaerobic medium C under 100% CO_2 headspace[150]. We melted solid sterile media at 98 ˚C in a water bath and cooled the media to ~45-50 ˚C prior to the addition of chloramphenicol and 1 ml of inoculum from a liquid fungal culture in mid-exponential phase. Upon inoculation, the tubes were transferred to a benchtop and immediately rolled horizontally creating a uniform agar-inoculum completely coating the walls. The tubes were incubated at 39 ˚C until colonies were visible, typically between three and five days. Following incubation, we extracted individual colonies from the agar with a sterile needle while under CO_2 headspace and transferred them to new Hungate tubes containing 9 ml anaerobic Medium C, switchgrass, and antibiotics (chloramphenicol [25 µg/ml in 40% ethanol], streptomycin [40 µg/ml], penicillin [50 µg/ml], and kanamycin [25 µg/ml]). After 72-96 hours, we used these cultures to inoculate new roll tubes. Colonies were passaged three times to obtain axenic cultures.

3.6.2 Substrate preparation

Lignocellulosic substrates were dried by placing them in a Fischer Scientific Isotemp convection oven at 45 ˚C until they reached approximately 10% moisture. Similarly, we collected the food waste (i.e. orange peel), washed it with deionized water, and dried it to approximately 10% moisture. We milled the dry substrates to 20 mesh (~ 0.85 mm) in a rotary mill. Milled substrates were loaded at 1% w/v prior to the addition of medium C [150]. For all soluble carbon sources, substrates were dissolved in anaerobic medium C at 0.5% w/v prior to being aliquoted into individual Hungate tubes under 100% CO_2 and autoclaved. Non-lignocellulosic substrates included arabinoxylan from beechwood (Megazyme, Bray, Ireland), xylan from beechwood (Crescent Chemical, Islandia, NY, USA), glucose, arabinose, xylose, cellobiose, filter paper, carboxy-methyl cellulose (Fisher Scientific, Waltham, MA, USA), Sigmacell Type 50, and Avicel pH 101 (Sigma Aldrich St. Louis, MO, USA). Genetically modified lines of poplar containing varying molar ratios of syringyl and guiacyl lignin were used to assess the response of *Piromyces sp. UH3-1* to lignin composition [151], [152]. Poplar at approximately 10% moisture was milled to 40mesh (~ 0.5 mm), and tubes were loaded with 1% w/v substrate. We tested eight different lines ofdebarked poplar. Two different lines of wild type poplar were used in this experiment; NM6, whichis a global standard, and INRA 717 from which all of the modified lines were constructed [139], [151]. While autoclaved, all biomass in this study is effectively untreated; empirical calculations of the extent of pretreatment or severity factor are four orders of magnitude smaller than mild forms of pretreatment (Log R_0 – 2.10) [153]. Similarly, preliminary studies did not demonstrate significant increases in fungal growth rate or total fungal biomass accumulation when unautoclaved corn stover is used as the substrate (Appendix B: Figure B.11). The autoclaved biomass has not been washed to remove any potential fermentation inhibitors which hinder enzyme activity [140], [154].

3.6.3 Microscopy

All images of *Piromyces sp. UH3-1* were collected via confocal microscopy (Nikon Eclipse Ti Microscope and A1-multiphoton imaging system). Mature fungal cultures containing lignocellulosic material were immobilized in 10% polyacrylamide prior to imaging with 4',6-diamidino-2-phenylindole (DAPI) (Thermofisher, Waltham, MA, USA). Zoospore images were

collected using 3-day corn stover cultures, which were placed in Eppendorf tubes and fixed with formaldehyde to a 4% final concentration.

3.6.4 Species classification

An axenic stock culture (described in Isolating a novel species of anaerobic fungi) was used to inoculate 50 ml serum bottles containing medium C with glucose 0.45% w/v, and chloramphenicol (25 μg/ml in 40% ethanol) [150]. These serum bottles incubated at 39 ˚C for three to four days upon which the gDNA was harvested for species classification. Fungal genomic DNA was isolated with the MoBio PowerFecal kit (Carlsbad, CA, USA), yielding sufficient quality genomicDNA (260/280: 1.9 & 260/230: 1.5) at approximately two μg DNA per 50 ml culture. PCR (Phusion DNA polymerase, Thermoscientific, Waltham, MA, USA) was used to amplify the Internal Transcribed Spacer 1 (ITS1) and ITS2 regions of the isolated genomic DNA via JB206/205 primers (5' GGAAGTAAAAGTCGTAACAAGG 3' and 5' TCCTCCGCTTATTAATATGC 3') yielding an expected amplicon of approximately 700-750 base pairs [155]. We also amplified the D1/D2 portion of the 28S rRNA large subunit (LSU) gene with the NL1/NL4 primers (5' GCATATCAATAAGCGGAGGAAAAG 3' and 5' GGTCCGTGTTTCAAGACGG 3') [111]. DNA was amplified with the following PCR settings for 30 cycles: annealing at 56 ˚C, elongating for 60 seconds at 72 ˚C and melting at 98 ˚C. All of the same conditions were used for the LSU PCR reaction except the annealing temperature was changed to 67 ˚C [111]. DNA amplification was checked on an agarose gel and imaged with a c600 Azure Biosystems imager. We concentrated these PCR products with the Zymogen DNA Clean & Concentrator (Zymo Research, Irvine, CA,USA) kit prior to sequence submission at GENEWIZ (South Plainfield, NJ, USA). We assembledthe forward and reverse sequence reads of the ITS1 and ITS2 region into a single contig by trimming the ends of reads with poor base calls (>3 Ns in a 20 base window) and assembling readswith 85% overlap over at least 20 bps with the contig assembly feature in GeneStudio bioinformatics package (ver. 2.2.0.0, GeneStudio, Inc., Suwanee, GA, USA). ITS sequences werealso validated by cloning PCR products into the pGEM-T Easy cloning vector (Promega, Madison,WI, USA) following the manufacturer's instructions and sequencing three resulting colonies. Phylogenetic reconstruction was performed using MEGA7 (v 7.0.14) [156]. Due to the lack of homogeneity in coverage across the ITS1 and ITS2 sequences in gut fungi, only ITS1 and LSU sequences were used [111]. ITS1 and LSU sequences were analyzed

with the maximum likelihoodmethod using a Tamura Nei nucleotide substitution model with 1000 bootstrap replications to estimate the confidence in node clustering.

3.6.5 Growth curve analyses for characterizing the substrate range of *Piromyces* sp. UH3-1

Fungal growth was tracked according to the method introduced by Thedorou et al. [116]. Briefly, Hungate tubes containing anaerobic medium C and untreated substrate were autoclaved prior to assessing growth (Appendix A: Tables A.1-A.5, Appendix B: Figures B.4-B.6) [150]. Every substrate was tested at least in triplicate for growth. Additionally, duplicate uninoculated tubes were used as negative controls for each substrate. Specific growth rates were determined by performing a linear regression of a semi-log plot of accumulated pressure (in psig) versus time (in hours). The Microsoft Excel LINEST function was used for each plot to calculate the slope and exponential phase. The data points in the exponential phase that were linearly increasing and had an R^2 of approximately 0.90 or higher (typically between 48 and 120 hours) were used to calculate the specific growth rates on each substrate. We prepared fresh media as described above. Lignocellulosic and insoluble substrates were loaded at 1% w/v while soluble substrates were loaded at 0.5% w/v to keep the total mass of fermentable sugars relatively constant. Tubes were inoculated in a random order to prevent systematic bias in inoculum quality. Pressure accumulation was measured with a pressure transducer (APG, Logan, Utah, USA), every eight hours for seven days. The growth of *Piromyces sp. UH3-1* on wild type and genetically modified lines of poplar was tested to evaluate the effect of lignin composition on fungal growth (Appendix A: Tables A.2-A.5) [139], [151]. For all analyses, individual growth rates and total accumulated pressures were calculated. For data normalization to glucose (Figure 3-3D), the average accumulated pressure (in psig) across culture (biological) replicates at 168 hours for each substrate was divided by the average accumulated pressure of glucose at 168 hours for all of the inoculated tubes. The error for these measurements was propagated accordingly. For data normalization to wild type poplar (Figure 3-5 A-B), the same procedures were followed for both growth rate and accumulated pressure.

3.6.6 Isolation of the carbohydrate active enzymes (CAZymes)

We used a pull-down purification protocol similar to the one by Solomon et al. to isolate and concentrate fungal CAZymes [18]. This procedure exploits the cellulose-binding domains of CAZymes to isolate lignocellulose degrading enzymes [18]. Cultures were centrifuged at 12,800g and the supernatant was transferred to a tube containing approximately 0.4% (w/v) Sigmacell type 50. These tubes were incubated overnight at 4 °C with gentle agitation. Tubes were then centrifuged at 12,800g and the supernatant was discarded. 0.1M pH 7.0 Tris-NaCl buffer was added to the Sigmacell to elute the cellulose-binding enzymes. The elutions were then stored at 4 ˚C for further analysis. Protein concentrations were determined by the method introduced by Bradford (Fisher Scientific, Waltham, MA, USA) [157].

3.6.7 SDS PAGE and zymography analyses for detailed enzyme characterization

Cellulose-binding proteins were separated and visualized on 10% acrylamide gels run for 70 minutes at 110V. Gels were then stained with Sypro Ruby Protein Stain (Fisher Scientific, Waltham, MA, USA). These proteins were also tested for activity via zymography with 0.2% w/v carboxy methyl cellulose (CMC) or 0.4 % w/v pectin added to the resolving portion of a 10% acrylamide gel under non-denaturing conditions. The SDS was removed from the gel with slight modification to the procedure of Tseng et al 2002 [158]. The gels were rinsed with ddH_2O and placed in 0.1M pH 7.0 Tris-NaCl (TN) buffer containing 25% (w/v) isopropanol (TNI) buffer. Zymogram gels incubated for 30 minutes at 4º C in TNI buffer with gentle agitation. The TNI buffer was then removed, and the gel was rinsed two more times with fresh TNI buffer. The zymograms were then washed with 0.1M pH 7.0 TN-buffer prior to incubating at 39 ºC for substrate hydrolysis. CMC zymograms were incubated for one hour while pectin zymograms were incubated for 24 hours. Zymograms were then stained in 0.1% w/v Congo red stain (Fisher Scientific, Waltham, MA, USA), for 30 minutes, and de-stained with 1M NaCl (Fisher Scientific, Waltham, MA, USA) until the hydrolysis zones appeared relative to the red background. We fixed the zymograms with 0.1M acetic acid (Fisher Scientific, Waltham, MA, USA) prior to imaging.

3.6.8 Sugar reducing assay for xylanase activity

Piromyces sp. UH3-1 xylanase activity was measured after harvesting the cellulose binding proteins as discussed above. Briefly, we followed the 96 µl microplate procedure introduced by Xiao et al [159]. However, we used 0.05M potassium phosphate buffer (pH 7.0) in place of citrate, and a 2% solution of xylan from beechwood (Crescent Chemical, Islandia, NY, USA) as the substrate. Substrate hydrolysis proceeded for six hours at 50º C before the generated reducing sugars were measured at 540 nm on a Synergy Neo plate reader (Biotek, Winooski, VT, USA). All samples were measured in triplicate and normalized by total protein. To determine the extent of non-enzymatic xylan degradation, enzyme-free and protein (Bovine Serum Albumin [BSA], Fisher Scientific, Waltham, MA, USA) controls were tested.

3.6.9 Analyzing the composition of lignocellulosic material after fungal growth

To test the effect of syringyl lignin composition on sugar consumption by *Piromyces sp. UH3-1*, we grew the isolate in 100 ml serum bottles with 50 ml working volume and 1.4% (w/v) solids loading to generate sufficient spent biomass for analysis. Three different poplar constructs were used: 0998-45 (5% S-lignin), wild type INRA 717 (64% S-lignin), and F5H-64 (98% S-lignin) [139], [151]. After seven days, the spent lignocellulosic biomass and associated fungal residues were separated from the fermentation media by centrifuging at 5,000 RPM for five minutes. After centrifugation, the liquid phase was decanted and the solids were dried for five days at 45 ˚C. The sugar composition of the spent biomass was determined according to standard methods (Appendix A, Table A.6) [160]–[162]. Carbohydrates were determined using HPLC analysis (Waters 1525 Pump,Waters Corporation, Milford, MA, USA) equipped with an Aminex™ HPX-87H column (Bio- Rad, Hercules, CA, USA) maintained at 65 °C. The mobile phase was 5 mM aqueous H_2SO_4 at a flow rate of 0.6 mL/min. 50 µL of sample was injected, analyzed using a Waters 2414 Refractive Index detector (Waters Corporation, Milford, MA, USA) and quantified using Empower Pro Software (Waters Corporation, Milford, MA, USA). The differences in glucan and xylan composition between the raw and spent biomass were calculated, and one-way ANOVA analyseswere performed to evaluate the differences in composition.

4. COMPLETE GENOMES OF LIGNOCELLULOLYTIC ANAEROBIC FUNGI PAVE THE WAY FOR GENETIC TOOL DEVELOPMENT

This chapter is being prepared for submission as a research article. Manuscript preparation is ongoing and will be completed after graduation.

4.1 Abstract

Anaerobic fungi (phylum Neocallimastigomycota) are an emerging platform for the conversion of recalcitrant plant material into a variety of products including biofuels and commodity chemicals. The genomes of anaerobic fungi indicate they possess a wealth of enzymes for degrading plant biomass and potential for unique biosynthetic capabilities. However, creating genetic tools to harness their genetic potential is hindered by fragmented genome assemblies that lead to incomplete or inaccurate genomic context for prediction of function. Specifically, these assemblies lead to incomplete gene clusters that delineate biosynthetic pathways, truncated promoter or regulatory sequences, and low confidence in intergenic sequences needed for development of gene integrations. Here, we address the issues with fragmented genomes by first optimizing the extraction of high molecular weight genomic DNA and sequencing it with long read PacBio platforms that overcome some of the issues related to assembling the highly homopolymeric genomes of anaerobic fungi. Higher quality genomic DNA improved the most recent PacBio scaffolding by more than a factor of 2 compared to the most recent *Neocallimastix* assembly, increasing the scaffold sizes of our isolates up to 8.4 Mbp and N50 to 2.5 Mbp,. Additionally, by pairing this improved assembly with chromatin confirmation sequencing, we generated the first closed genomes of anaerobic fungi that increased the scaffolds up to 18.2 Mbp and N50 up to 10.4 Mbp. With genomic context, the organization of the Carbohydrate Active enZymes (CAZymes) and biosynthetic gene clusters (BGCs) was evaluated in order to understand their origins, potential functions, and impacts on regulation that were not possible with the previous incomplete genome drafts. Finally, using the high-fidelity assembly of intergenic regions in these closed genomes, we are now able to globally identify promoter sequences for gene expression, the centromeric interactions to identify components for plasmid segregation and replication, and the enhanced gene context to design homology arms for gene integration. While

creating a gene-editing system for anaerobic fungi is in progress, the parts leveraged form these genomes now move us closer to realizing the full potential of anaerobic fungi in the bioeconomy.

4.2 Background

Anaerobic fungi (phylum Neocallimastigomycota) are emerging biotechnology platforms that degrade recalcitrant plant material into fermentable sugars that can be transformed into a variety of products including biofuels and commodity chemicals [11]. As the bioeconomy continues to develop, anaerobic fungi may be instrumental in accessing the energy-rich carbon that is trapped in the abundant renewable plant biomass that is generated each year. At this time, the annual market for products made from these feedstocks is over $48 billion globally and growing [1]. Because the enzymes from the genomes of anaerobic fungi can degrade diverse feedstocks, anaerobic fungi enable these production platforms to incorporate additional feedstock and add to the growing bioeconomy.

The genomes of anaerobic fungi contain a wealth of enzymes that can be harvested for bioenergy and beyond. In terms of biomass-degrading enzymes, anaerobic fungi have the largest repertoire in the fungal kingdom [15] including *Aspergillus* and *Trichoderma,* which are commonly sourced for such enzymes. Some of these carbohydrate active enzymes (CAZymes) have been shown to be as active as commercially produced enzyme cocktails [18]. Additionally, like other fungi, anaerobic fungi possess a variety of biosynthetic genes clusters (BGCs) including polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS), and terpenes [23], [94]. Because products formed from these compounds make up the majority of medicines today [163], anaerobic fungi may also be an untapped source of pharmaceutical compounds. The potential of the anaerobic fungi genome, however, has yet to be fully leveraged due to the lack of genetic tools for these fungi.

Strategies to engineer anaerobic fungi are currently hindered by a lack of genetic information including needed gene regulatory elements. Anaerobic fungi genomes are historically difficult to sequence and the current genomes are highly fragmented due to their low GC-content and large intergenic regions rich in A/T homopolymers [164]. Consequently, the first assemblies of anaerobic fungal genomes resulted in over 30,000 fragments and N50 values below 0.01 Mbp – a metric that indicates at least half of the fragments are about 10 kb or smaller [165]. That is, anaerobic fungal genomes are highly fragmented potentially disrupting valuable parts and

inhibiting their discovery. While recent assemblies with PacBio long read technologies have resulted in genomes with N50 values of 0.75 -1.03 Mbp and 232 to 970 contigs, there is still a lack of gapless genomes for this family of organisms. Because the intergenic regions are homopolymeric, they are not only difficult to sequence but also to assemble; these features correspond to where many promoters reside and they are poorly annotated. As a result, only one reported promoter has been used to drive expression of heterologous genes in anaerobic fungi, and it was discovered from a genomic library pool rather than sourcing it from an assembled genome [166]. Without promoters for heterologous expression, the only molecular-level engineering strategy that has been reported is an RNAi approach that disrupts gene function [167]. Anaerobic fungi have also been incorporated in bottom-up [13] and top-down [168] engineering strategies to leverage their biomass-degrading abilities; however, these efforts still lack the precision to control and exploit the full potential of their genomes. Finally, to enable gene-editing or recombination of heterologous genes, high-quality sequences for intergenic regions that flank target loci are essential for efficient genome engineering.

By leveraging recent advances in genome sequencing and assembly, we resolved the genomes of three new anaerobic fungal isolates. Specifically, the gDNA extraction process was optimized and paired with the PacBio long read sequencing platform to overcome difficulties in assembling the homopolymer rich intergenic regions of the genome. Further, by pairing the large PacBio assemblies with chromosomal conformation capture sequencing, the spatial organization of the assembled contigs was obtained and used to reconstruct the genome with chromosomal resolution for the first time in anaerobic fungi. With genomic context, the organization of the CAZymes and BGCs was evaluated in order to identify horizontally transferred gene clusters and understand their potential functions in ways that were not possible with the previous incomplete genome drafts. Finally, the basic parts of a genetic toolbox were created using the high fidelity assembly of intergenic regions in these closed genomes. We are now able to globally identify promoter sequences for gene expression, the centromeric interactions to identify components for plasmid segregation and replication, and the enhanced gene context to design homology arms for gene integration. Though creating a gene-editing system remains a work in progress, these genomes now allow us to leverage their components and, ultimately, advance toward realizing the full potential of anaerobic fungi in the bioeconomy.

4.3 Results and Discussion

4.3.1 Isolated anaerobic fungi are lignocellulolytic

Studied anaerobic fungi belonging to the phylum Neocallimastigomycota were isolated from herbivores from a local zoo: *Neocallimastix sp. GfMa3-1* (Figure 4-1A, Appendix C: Figure C.1) was isolated from a giraffe (*Giraffa reticulata*) while *Neocallimastix sp. WI3-B* (Figure 4-1B, Appendix C: Figure C.2) was isolated from a wildebeest (*Connochaetes gnou*). These isolates have been serially passaged and cryopreserved alongside *Piromyces sp. UH3-1*, whose isolation was previously described in Chapter 3 (Figure 4-1C, adapted from [14]). Based on ITS and LSU phylogenetic analysis (Figure 4-1D), these isolates were assigned to the *Neocallimastix* genus; this designation was independently confirmed with morphological characterization. Specifically, these isolates are polyflagellated with rhizoidal, monocentric thalli, which are defining characteristics for *Neocallimastix* [30]. Like *Piromyces sp. UH3-1*, these isolates both grow on simple and cellulosic feedstocks as demonstrated by *Neocallimastix sp. GfMa3-1* in Figure 4-1E in both semi-defined Medium B and rumen-fluid-based Medium C. Adding these genomes to the Mycocosm portal [94] will increases the number of publicly available genomes by ~43% and will aid future comparative genomic studies as we try to unravel the physiology of this unique group of fungi. Based on the phylogenetic analysis and comparison to the seven other genomes publicly available through Mycocosm, these three isolates (Figure 4-1) represent novel isolates whose genomes hold untapped potential for biomass degradation and bioproduction.

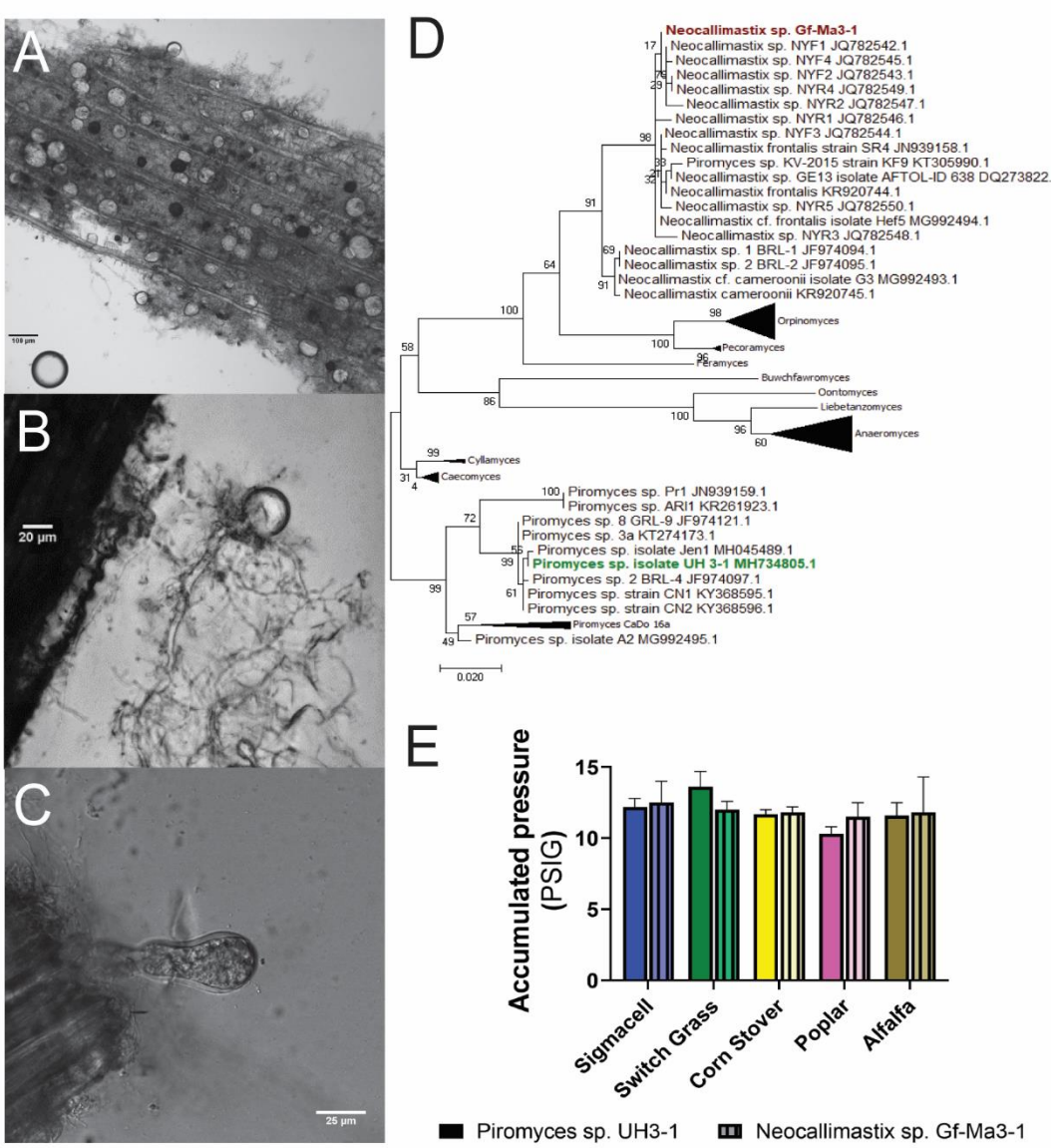


Figure 4-1: Isolated anaerobic fungi. Micrograph of A) *Neocallimastix sp. GfMa*3-1, B) *Neocallimastix sp. WI3-B*, and C) *Piromyces sp. UH3-1* growing on anaerobic crude biomass. D) Large subunit (28S rRNA) phylogenetic tree showing genus level designations of isolates GF and UH. E) Accumulated pressure (PSI) from GF and UH on lignocellulosic feedstocks in Medium B. Error bars = standard deviation.

4.3.2 Plant genome isolation methods reduce genome fragmentation

Because anaerobic fungi are strict anaerobes, they can be difficult to isolate, maintain, and cryo preserve; however, difficulty also surrounds the extraction and sequencing of their genomes [169]. While advances in sequencing technology and bioinformatics have improved the genomes

of anaerobic fungi, the best genomes are still left in hundreds of pieces due to the underlying AT-rich genes that contain multitudes of low-complexity repeat regions that complicate genome assembly [35]. Although technological advances have been able to tackle some of the issues related to sequencing and assembling the complex regions of the genome, they are still dependent on the extraction of high-quality, high molecular weight (HMW, >20 kb) genomic DNA. To overcome these challenges, the extraction of HMW genomic DNA was optimized and combined with the long-read sequencing advances of PacBio and the chromatin proximity capturing abilities of HiC to resolve genome assemblies into chromosomal scaffolds [170], [171]. To this point, however, the best methods to extract large molecular weight genomic DNA from anaerobic fungi [169] rely on kits with bead-based lysis that shears the genomic DNA (Figure 4-2A). So, in order to overcome the dense, chitin-rich membranes of anaerobic fungi [172], protocols were adapted from the genome extraction techniques used for other recalcitrant organisms and plant tissues [173], [174]. In collaboration with the Arizona Genome Institute, liquid N_2 chilled homogenization with mortar and pestle followed by a gentle CTAB and polyvinylpyrrolidone-based lysis yielded genomic DNA that was much less fragmented than the bead-beating kit extractions for all three isolates (Figure 4-2A, B). For example, DNA isolated from *Neocallimastix sp. WI3-B* ranged in size from 30-145 kb (Figure 4-2C) representing a 7-fold reduction in fragmentation compared to previous preps [169]. Although there was some degraded RNA evident in the gel (Figure 4-2B), this is removed during the library preparation for the PacBio platforms. Ultimately, the PacBio sequencing and assembly of these high molecular weight DNA preps resulted in 98% of the sequenced DNA residing in 169-317 scaffolds with L50s of 25-30 (Table 4- 1) indicating that the majority of these assemblies reside in larger contigs. Comparing these assemblies with *Neocallimastix lanati* v1.0, currently the most complete genome (N50 of 1.03 Mbp) with 970 scaffolds and largest contig of 4.01 Mbp, our assemblies yield up to 67% fewer scaffolds, more than 2.5 fold the N50, and scaffolds up to twice as large (8.4 Mbp). That is, our genomes are less fragmented than previous state of the art assemblies. Although these assemblies are improvements over most anaerobic fungal assemblies that range from 232- 32572 scaffolds [96], they still do not resolve the size or number of chromosomes that would enable the evaluation of genome organization, structure, and function.

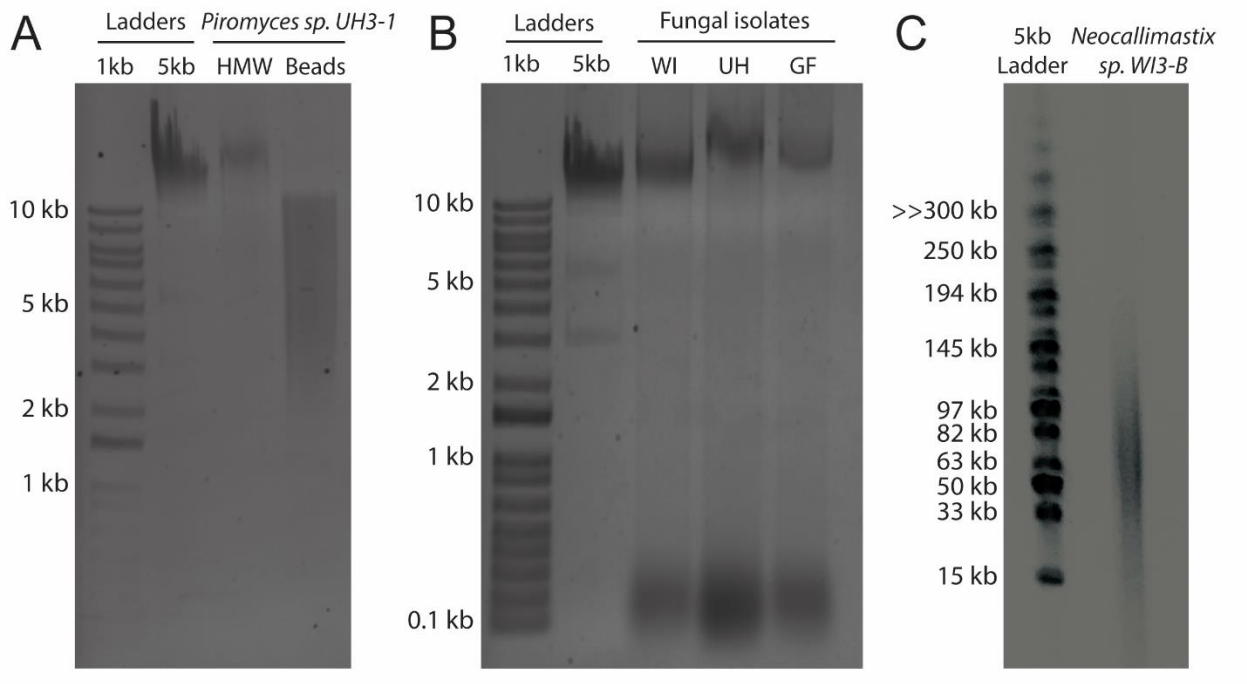


Figure 4-2: Optimization of high molecular weight genomic DNA extraction. A) 0.7% agarose gel showing high-molecular-weight- and silica bead-based genomic DNA from *Piromyces sp. UH3-1* showing the increase in size under the high molecular weight protocol. B) 0.7% agarose gel showing high-molecular-weight genomic DNA isolation for all three isolates; degraded RNA ~0.1 kb is common in anaerobic fugus genome preps [169]. C) CHEF gel showing the distribution of genomic DNA isolation from Neocallimastix WI3-B. Ladder sizes are shown to the left of each gel.

Table 4-1: Genome assembly comparison of PacBio and PacBio + HiC.

Genome Assembly	*Piromyces sp. UH3-1*	*Neocallimastix sp. GfMa3-1*	*Neocallimastix sp. WI3-B*
Genome Assembly size (Mbp)	84.1 Mbp	209.5 Mbp	206.8 Mbp
# of contigs (PacBio)*	317	236	169
Scaffold L50	30	27	25
Scaffold N50	0.87 Mbp	2.50 Mbp	2.68 Mbp
# of contigs (PacBio + HiC)*	12	24	25
Scaffold L50	5	7	8
Scaffold N50	7.30 Mbp	10.4 Mbp	10.1 Mbp[1]

[1] * = of contigs represents the number of contigs that represent >98% of the genome. Scaffold L50 = number of scaffolds that contains 50% of DNA, Scaffold N50 = Molecular weight of scaffold that contains 50% of DNA.

4.3.3 HiC sequencing enables complete assembly of genomes with chromosomal resolution

To resolve chromosomes, we employed HiC sequencing, a chromatin conformation capture technique that records the spatial organization of genomic DNA by crosslinking nearby DNA before fragmentation[171]. Crosslinked DNA tend to be captured in the same reads providing additional information for DNA assembly of standard sequenced libraries. Using the large PacBio assemblies generated as initial scaffolds, the chromosomal scaffolds (Figure 4-3) were computationally recreated using the interactions detected from the library of crosslinked DNA sequences. Because PacBio sequences generated from the optimized gDNA extraction method produced large scaffolds, it enabled HiC data to be incorporated where it could not previously perform well due to inferior assemblies. Once the HiC interaction data was incorporated, we were able to assemble our genomes onto 12-25 main scaffolds (Figure 4-3B, Table C.1) representing the number of chromosomes in each organism, where the chromosome sizes range from 4.2 to 11 Mbp for UH, 0.7 to 15.0 for WI, and 0.7 to 18.2 for GF (Table C.1). These assemblies represent the first complete assembly of gut fungal genomes resolved to the chromosomal level.

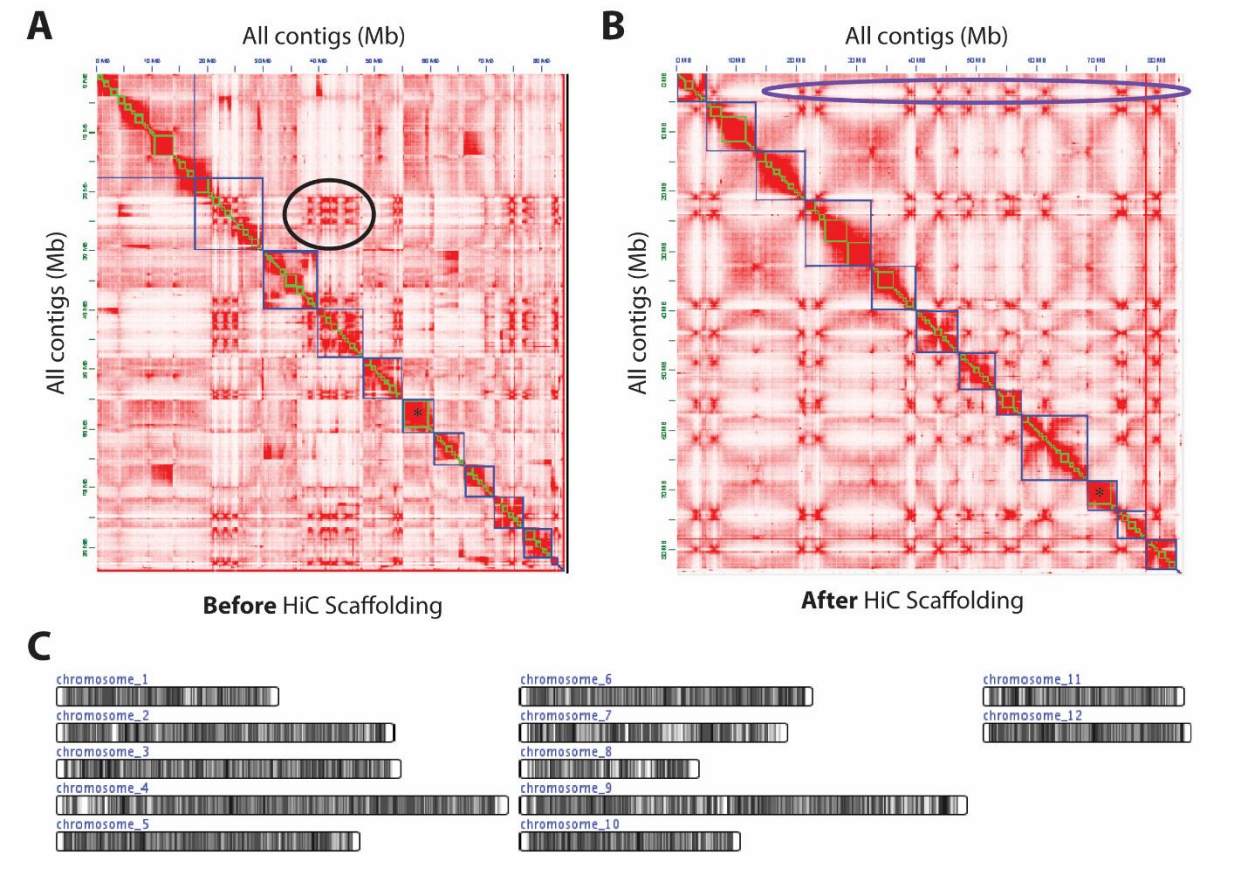


Figure 4-3: Interaction plot of *Piromyces sp. UH3-1* assemblies. A) Before and B) after HiC scaffolding. C) Representative chromosome map showing relative chromosome sizes. Tick marks indicate 10 Mbp. Black circle = example of off-diagonal interaction that suggests contigs need to be reorganized; purple ellipse represents off-diagonal that represents centromeric interactions of chromosomes; blue boxes represent scaffolds; green boxes represent contigs.

An added advantage of HiC assemblies is their power to identify and correct misjoins of PacBio and other Next Gen Sequencing assemblies based on the proximity data. Though they do not occur on every contig, we identified 153 to 264 misjoins (Table C.1) in the PacBio assemblies in low complexity regions (i.e. homopolymer-rich regions) that were only resolved via the HiC interaction data (Figure 4-3A). The pre-HiC scaffolding interaction plots show several areas (one indicated by black circle in Figure 4-3A) where the contigs, or fragments of contigs, have strong interactions with other contigs. This suggests that these fragments of DNA are in close proximity and should be joined onto a larger scaffold or chromosome; the interaction-corrected chromosomes are shown in Figure 4-3B. There are, however, some interactions that cannot be resolved and are shown inside the purple circle on Figure 4-3B; these represent the centromeres of the chromosomes,

which are frequently in close proximity and, therefore, are real interchromosomal interactions captured by HiC. These centromeric interactions also corroborate the number of chromosomes resolved via HiC assembly. Each chromosome has a single centromere that can interact with corresponding centromeres in adjoining chromosomes (off-diagonal interactions within the purple circle; Figure 4-3B). The number of predicted centromeres is consistent with the number of resolved chromosomes (large scaffolds) for each assembly (Appendix C: Table C.1, Figures C.3 & 4). In addition to the corrected misjoins, there are additional sequences that could not be assembled into these chromosomal scaffolds, however, they represent less than 2% of the genome. Ultimately, these assemblies provide a much more accurate picture of the anaerobic fungal genomes and will enable the assessment of how the genome content is organized.

4.3.4 Genome annotation reveals a wealth of fungal enzymes for biotech and gene synteny information to parse gene function and evolution

At the time of this writing, only the annotation for both the PacBio and PacBio + HiC assemblies of *Piromyces sp. UH3-1* have been completed and will be the focus here; the other two *Neocallimastix* genome annotations are currently underway with the Joint Genome Institute (JGI) and expected to be complete in the spring of 2021. Interestingly, we find approximately 200 more genes in the HiC + PacBio assembly (1.3% more) (Figure 4-4A) than the PacBio only assembly. The approximately 17,000 genes identified for *Piromyces sp. UH3-1* fits well within the range of genes identified for other anaerobic fungi (12,800 to 28,900). Similarly, analysis of the EuKaryotic Orthologous Groups (KOG) identified in each *Piromyces* assembly shows that the genome content is highly similar for the various KOG pathways and number of CAZymes regardless of assembly method (Figure 4-4A). Among the available anaerobic fungal genomes, the number of CAZymes found here (~1,600) in the *Piromyces sp. UH3-1* genome is second only to *Neocallimastix lanati* (~1,800) despite the *N. lanati* genome being >2x the size of *Piromyces sp. UH3-1*. Interestingly, the most abundant CAZymes in *Piromyces sp. UH3-1* are the GH11 xylanase and GH5 glucanase (Figure 4-4B), however, *Piromyces sp. UH3-1* does not grow well solely on xylose [14]. Recent studies of the synergy of ruminal microbiomes showed that anaerobic fungi are primarily responsible for cellulose decomposition using their abundance of GH5, 6, 8, and 48 genes while bacteria are primarily responsible for the deconstruction of hemicellulose [175]. Anaerobic fungi can use these hemicellulases to cleave xylan to access additional glucan trapped in the complex

lignocellulose structure particularly when isolated in the absence of synergistic bacteria. Given that horizontal gene transfer (HGT) has been reported in anaerobic fungi [35], the abundance of hemicellulases may be due in part to similar events in *Piromyces sp. UH3-1*.

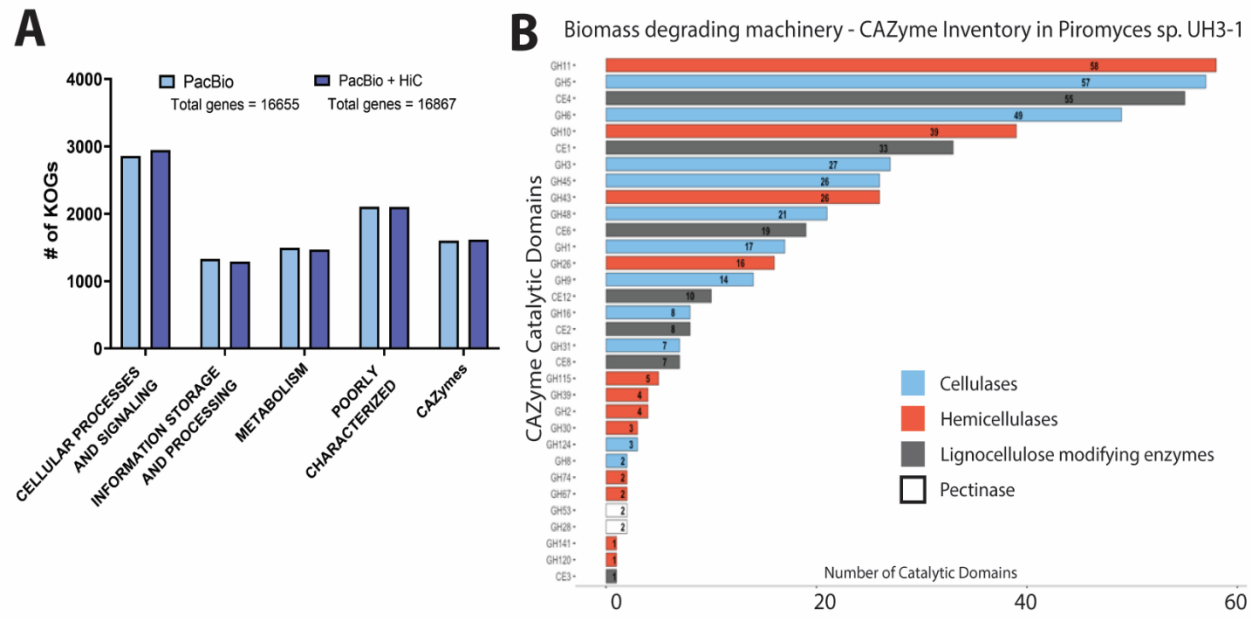


Figure 4-4: KOG comparison and CAZyme breakdown. A) Summary of KOGs by category for both PacBio and PacBio + HiC assemblies. B) Abundance of CAZyme domains in the PacBio assembly.

Additionally, the chromosomal resolution gained by pairing HiC and PacBio enables the evaluation of how the genomes are spatially organized and how HGT events are distributed throughout the genome. Because anaerobic fungi possess a great deal of carbohydrate active enzymes (CAZymes), we evaluated how they are distributed throughout the genome in order to look for patterns related to their function or origin. Anaerobic fungi have been shown to possess both bacterial and fungal CAZymes, but it is not clear whether these HGT events resulted in a cluster of several transferred CAZymes or whether they are more dispersed throughout the genome having occurred from separate events. Using a BLAST-based phylogenetic framework, we observed that roughly 11% of the *Piromyces sp. UH3*-1 genes have a bacterial origin (Figure 4-5A, Appendix C: Figures C.5 & 6). Out of the approximately 1,600 CAZymes annotated, we observed that roughly 31% of the CAZymes were of bacterial origin while about 51% can be attributed to fungi or anaerobic fungi (Figure 4-5B, Appendix C: Figures C.7 & 8). While we are

still cataloging all of the CAZyme transfer events, we see evidence on chromosome four of a cluster of 15-30 genes that appear to work together in the deconstruction of plant biomass (Figure 4-6A). This cluster includes dockerin proteins, carbohydrate-binding modules, xylan-retaining modules (GH10), and endo-β-1,4-glucanases (GH3 and GH5) that have primarily bacterial origins. This seemingly synergistic local organization suggests that these genes of non-*Neocallimastigomycota* origin, may have been obtained through a horizontal gene transfer event that could give them advantages in breaking down wider varieties of biomass. Similarly, the higher HGT frequency of bacterial CAZymes (31% of CAZymes compared to 11% of the all genes; See Appendix C: Figures C.5-8) suggests that this strategy of broadening the repertoire of enzymes to widen substrate versatility may be a particularly beneficial advantage in the competitive rumen environment. These clusters of HGT CAZymes appear widespread and occur throughout the entire genome, but HGT CAZymes are not exclusively found in clusters. Similarly, CAZymes native to *Neocallimastigomycota* are not organized into operons, but both clusters and individual anaerobic fungus CAZymes can also be found throughout the genome. At this time, the role that local and global organization of CAZyme plays in the expression and regulation of these genes is not yet clear but is currently being investigated with these high-resolution genomes. Understanding the organization and regulation of these enzymes will provide insight into how these genes and clusters of genes can be leveraged for enhanced bioproduction strategies.

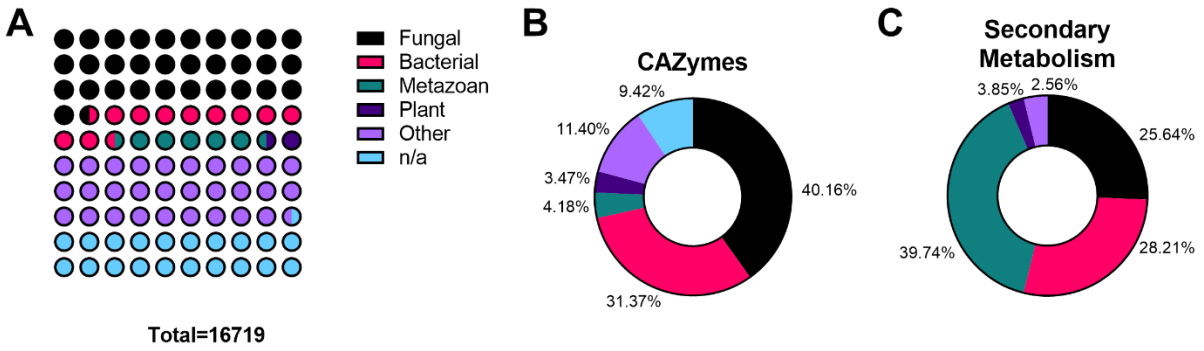
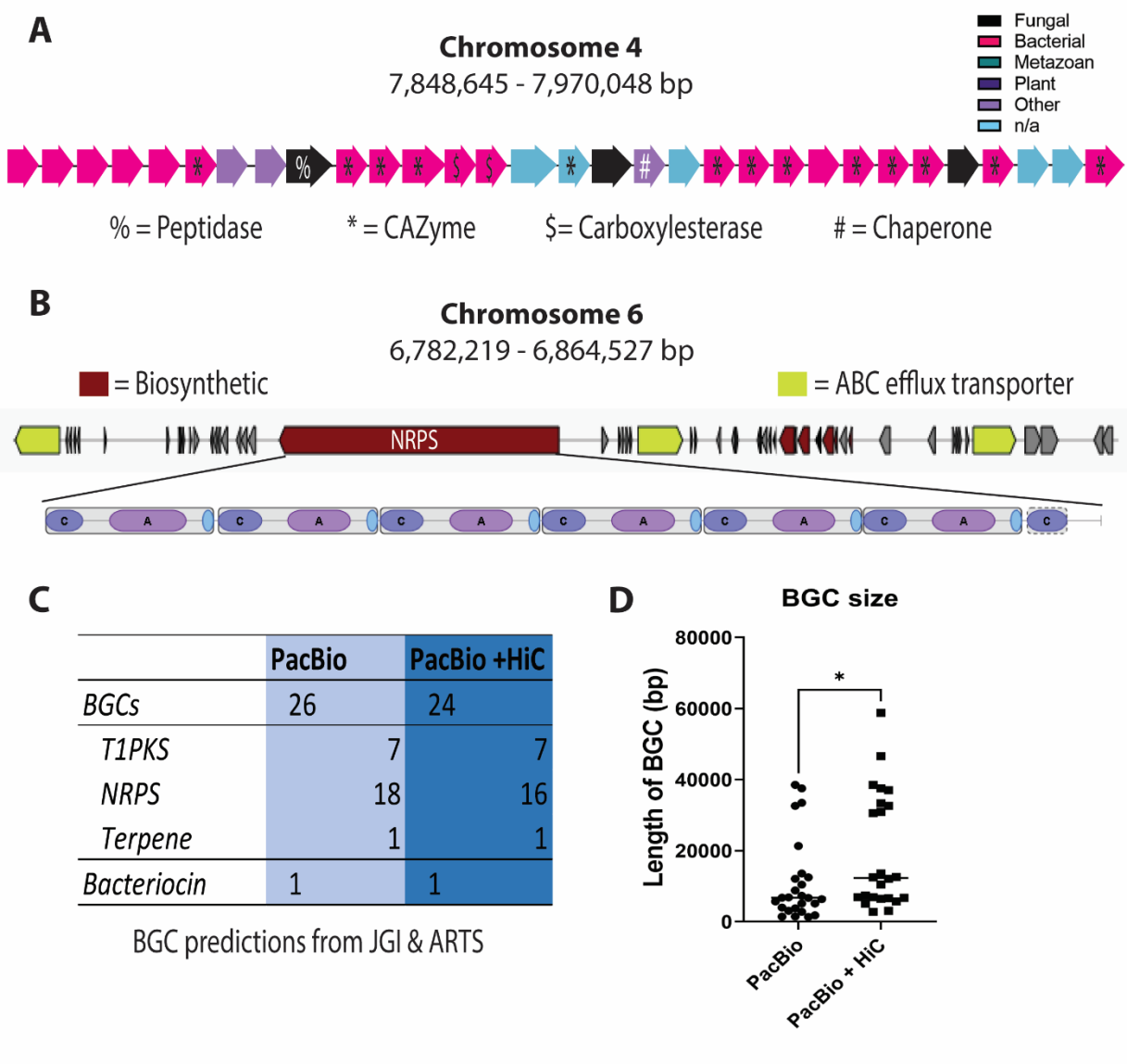


Figure 4-5: Horizontal gene transfer events. A) Waffle plot showing the fraction of the genome according to it best hit of non-Neocallimastigomycota origin. Fraction of B) CAZymes and C) BGCs by best hit of non-Neocallimastigomycota origin. Other = Virus, archeae, and miscellaneous Eukarya; Fungal = non-Neocallimastigomycota fungi; n/a = no non-Neocallimastigomycota BLAST result.

In addition to CAZyme organization and origin, we evaluated the biosynthetic genes found in the UH3-1 genome. In contrast to CAZymes that incur a selective advantage by equipping anaerobic fungi for biomass degradation, secondary metabolites often provide a selective advantage by mediating interactions with other microbes in the environment (e.g. antibiotics) as they compete for and exchange nutrients and resources. These compounds are typically formed from long biosynthetic pathways whose genes are frequently clustered. While the secondary metabolites from anaerobic fungi have not yet been evaluated for their functional activity, UH3-1 has an abundance (~25) of biosynthetic gene clusters (BGCs) that historically has been the source of antibiotics as well as bacteriocin polypeptides that target specific bacteria (Figure 4-6C). Non-biosynthetic genes within or in close proximity to BGCs, such as efflux transporters or duplicated gene variants of an essential enzyme, commonly confer resistance to the host and may be used to infer BGC function or molecular target [176] (Figure 4-6B). Using the Antibiotic Resistance Target Seeker (ARTS) framework [176], we found that the PacBio and HiC aided assemblies both yielded similar BGC predictions (Figure 4-6C). Additionally, the BGC predictions from both assemblies were able to identify nearby ABC transporters that may function as pumps associated with secondary metabolism. However, the PacBio resolved BGCs were often smaller and more fragmented (Figure 4-6D, $P = 0.044$) suggesting that the improved assembly can identify more complete BGCs for more complete pathway reconstruction. When we evaluated the origins of these genes, we found that the majority came from metazoan or bacterial origins (Figure 4-5C). The metazoan origins primarily pertained to transporters while the bacterial BGCs predictions were a mixture of Type 1 polyketide synthases (PKS) and non-ribosomal polypeptide synthetases (NRPS) that are associated with a variety of natural products. Because BGCs and CAZymes likely confer a selective advantage in the competitive rumen microbiome, they may be readily taken up by anaerobic fungi and enriched among the detected HGT events [177]. As highlighted by the CAZyme and BGC examples, the improved HiC assembly provides additional insight into gene function via genome organization and allows for more complete recapitulation of industrially relevant metabolic pathways. Overall, the added chromosomal resolution will allow the investigation of how gene expression changes on local and global regions in response to substrate composition and epigenetic inhibitor [35]. By pairing on-going epigenetic experiments and the related transcriptomic data, we will be able to better understand gene regulation and how it can be leveraged to enhance biomass deconstruction in anaerobic fungi.



	PacBio	PacBio +HiC
BGCs	26	24
T1PKS	7	7
NRPS	18	16
Terpene	1	1
Bacteriocin	1	1

BGC predictions from JGI & ARTS

Figure 4-6: Examples of CAZyme and Natural Product clusters. A) Expected bacterial HGT cluster of CAZymes and related enzymes from chromosome 4. B) Example of NRPS biosynthetic gene cluster in close proximity to efflux pumps that may confer resistance to toxic product synthesized. C) Natural products predicted in *Piromyces sp. UH3-1*. D) BGC size of different assemblies. * = P < 0.05, Unpaired T-test.

4.3.5 *Piromyces sp. UH3-1* is haploid

Given the rich biosynthetic potential of these organisms, we then investigated the ploidy or genetics of our isolate to determine how easily these genomes could be manipulated. Diploid or polyploid organisms have multiple alleles for genes making cellular phenotypes a complex function of dominant and recessive allele distributions. Based on gene annotation we found 232

potential loci with alleles, or variants of the same protein, over the entire UH3-1 genome. These "alleles" represent less than 2% of the gene in *Piromyces sp. UH3-1* and is far lower than expected for a true diploid or polyploid organism [178]. That is, it is likely that UH3-1 is a haploid organism with 12 chromosomes. It is unclear at this time if these genes are sequencing artifacts or if they are duplicated genes. In order to validate that this isolate is haploid, fluorescent flow cytometry was used to investigate the genomic content of extracted nuclei from *Piromyces sp. UH3-1*. We simultaneously evaluated juvenile and mature nuclei by extracting them from maturing biomass that contains a mixture of all lifecycle stages (see Appendix C: Figures C.1-C.2) [179]. Characteristic of a haploid organism, the *Piromyces* nuclei generated a single peak with propidium iodide (PI) staining of chromosomal DNA (Figure 4-7). In contrast to haploid (1n) organisms like the control *S. cerevisiae*, diploids produce two peaks where one shows the fluorescence intensity of the haploid (1n) stage and the other shows the diploid (2n) stage[180]. Both the *S. cerevisiae* and *Piromyces sp. UH3-1* nuclei produce a single peak suggesting they are haploids while the difference in peak height and area of fluorescence intensity is due to the difference in genome size between them. It is unclear if this is a feature of all anaerobic fungi, or just these isolates because recent genomic surveys have identified sexual genes in anaerobic fungi [19] and higher levels of allele heterozygosity in other isolates [35]. While anaerobic fungi are typically thought to be asexual with the mature fruiting body producing zoospore the mature into mature zoosporangia, the near-neighboring *Chytridomycetes* are able to reproduce both asexually and sexually [181]. However, given the relatively low frequency of heterozygosity of "alleles", we suspect that *Piromyces sp. UH3-1* reproduces primarily, if not solely, asexually. As we move toward genome engineering approaches for anaerobic fungi, the finding that this genome is haploid may simplify our approaches. For example, genome engineering in diploids requires that both alleles are modified; otherwise, the expression from an unedited allele may nullify a knockout or silence the expression of the modified allele. Additionally, the organisms may repair the modified allele during either the gene knockout process or even after the integration has been made. Ultimately, stable genome engineering in haploids has fewer barriers than diploids, and therefore, may simplify the genome editing process of anaerobic fungi.

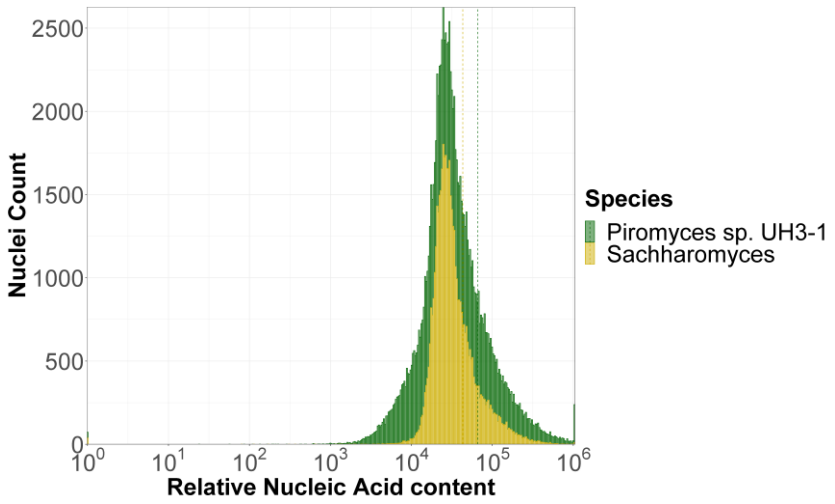


Figure 4-7: Nucleic acid content of anaerobic fungi suggests anaerobic fungi are haploid. Fluorescence intensity of extracted *Piromyces sp. UH3-1* (green) and *S. cerevisiae* (yellow) nuclei stained with propidium iodide.

4.3.6 Mining chromosomal assemblies for genetic tools

Despite the fact that haploid genomes can simplify downstream genome editing efforts, there are currently no basic tools for heterologous expression or genome editing in anaerobic fungi. The simplest genetic tools are promoters, reporters, and selection markers that control expression of proteins, validate expression, and select for populations displaying desired phenotypes, respectively. However, having high-quality genomes can help identify some of these basic parts. Promoters, specifically, can be identified from genomes by evaluating the regions directly upstream of the start site of genes [182]. In anaerobic fungi, however, the promoter regions are difficult to sequence as they often average 10-15% GC and contain stretches rich in poly A or T repeats that often hinder sequencing efforts [164]. By pairing high PacBio and HiC sequencing, however, we are able to obtain these sequences with high confidence. Without high-quality genomes, promoters have only been pulled from genomic libraries [166] that require large-scale screening efforts. Instead, we paired existing genomic and transcriptomic data to identify conserved genes that are constitutively expressed at high, medium, or low levels in multiple strains of anaerobic fungi and then used our high-quality genomes to obtain promoter sequences for the homologs in *Piromyces sp. UH3-1*. These sequences are currently being synthesized for testing. In addition to promoter regions, the PacBio + HiC assemblies allow accurate identification of sequences that can be used for homology arms to integrate sequences via homologous

recombination during genome editing. Specifically, these regions often contain up to 2 kb of intergenic sequence that is hard to obtain due to its low complexity and low GC%. Finally, because the PacBio + HiC assembly captures the frequent DNA proximity of the centromeric regions [171] (Figure 4-3), the centromere binding sequences, or CEN, can be extracted and used to improve the stability of extra-chromosomal plasmids by ensuring a copy of the CEN-containing plasmid is retained in each daughter cell [182]. Though we are still developing genetic tools to engineer the genomes of anaerobic fungi, these high-resolution genomes allow more accurately identification and design of parts by leveraging the high-confidence sequences for promoter and target loci, and ultimately, move us closer to being able to realize the potential of anaerobic fungi.

4.4 Conclusions

By improving the way genomic DNA is obtained and analyzed, we improved the genome sequencing and assembly of three isolates of anaerobic fungi (*Piromyces sp. UH3-1*, *Neocallimastix sp. GfMa3-1*, *Neocallimastix sp. WI3-B*). Higher quality genomic DNA improved the most recent PacBio scaffolding by more than a factor of 2, increasing the scaffolds up to 8.4 Mbp and N50 to 2.5 Mbp, compared to the most recent *Neocallimastix* isolate. These large scaffolds enabled the pairing of PacBio sequencing with chromosomal capture (HiC) sequencing to yield the first complete genome of any anaerobic fungi and increasing the scaffolds up to 18.2 Mbp and N50 up to 10.4 Mbp. Ultimately, the chromosomal resolution of *Piromyces sp. UH3-1* enabled the evaluation of CAZymes and BGCs organization in anaerobic fungi and sets the stage for future studies on how these clusters of genes are regulated. Additionally, the organizational information that these assemblies generated permitted the investigation of CAZyme and BGC function in the context of the neighboring genes where these clusters were previously truncated and fragmented. Future work can build on this dataset by leverage this chromosomal resolution to identify and design biological parts for a genetic toolbox that were previously hindered by low confidence sequences in AT-rich intergenic and promoter regions. These basic parts form the foundation of a genetic toolbox that will enable anaerobic fungi to be engineered for enhanced biofuel and bioproduct production.

4.5 Methods

4.5.1 Isolation and cultivation of anaerobic fungi

Fresh fecal material from zoo animals (giraffe and wildebeest, housed at the Indianapolis zoo) was suspended in Hungate tubes containing sterile anaerobic medium C supplemented with 15% clarified rumen fluid (150 ml: Bar Diamond Inc., Parma, ID, USA) under 100% CO2 headspace [150]. Fecal suspensions were serially diluted 1000-fold and used as inoculum in Hungate tubes containing 9 ml anaerobic Medium C, supplemented with switchgrass as a carbon source (1% w/v) and chloramphenicol (25 µg/ml; Fisher Scientific, Waltham, MA, USA). After inoculation, the cultures were incubated at 39 ˚C for 72-96 hours. As previously described, roll tubes were inoculated with liquid fungal culture and individual colonies were propagated and then isolated with two more successive rounds of roll tubes [14]. Roll tubes were prepared by adding agar (2% w/v), glucose (0.45% w/v), and chloramphenicol (25 µg/ml) to anaerobic medium C under 100% CO2 headspace[150]. Sterile media was melted at 98 ˚C in a water bath and cooled the media to ~45-50 ˚C prior to the addition of chloramphenicol and 1 ml of inoculum from a liquid fungal culture in mid- exponential phase. Upon inoculation, the tubes were transferred to a benchtop and immediately rolled horizontally creating a uniform agar-inoculum completely coating the walls. The tubes were incubated at 39 ˚C until colonies were visible, typically between three and five days. Following incubation, we extracted individual colonies from the agar with a sterile needle while under a $H_2/CO_2/N_2$ headspace in an anaerobic chamber (Plas-Labs, Inc, Lansing, MI, USA) and transferred them to new Hungate tubes containing 9 ml anaerobic Medium C, switchgrass, and chloramphenicol (25 µg/ml in 40% ethanol). After 72-96 hours, new cultures were inoculated and this process was repeated for a total of three rounds of isolation.

4.5.2 Phylogenetic and morphological characterization of anaerobic fungi

All images of *Neocallimastix sp. GfMa3-1 and WI3-B* were collected via confocal microscopy (Nikon EclipseTi Microscope and A1-multiphoton imaging system). Mature fungal cultures containing were grown three to four days as described above on lignocellulosic material and mixtures of both supernatant and biomass were stained with 4',6- diamidino-2-phenylindole (DAPI, ~300 nM) (Thermofisher, Waltham, MA, USA) prior to staining.

For phylogenetic analysis, an axenic stock culture was used to inoculate 50 ml serum bottles containing Medium C with glucose 0.45% w/v, and chloramphenicol (25 μg/ml in 40% ethanol) [150]. These serum bottles incubated at 39 ˚C for three to four days upon which the gDNA was harvested for species classification. Fungal genomic DNA was isolated with the MoBio PowerFecal kit (Carlsbad, CA, USA), yielding sufficient quality genomicDNA (260/280: 1.9 & 260/230: 1.5) at approximately two μg DNA per 50 ml culture. PCR (PhusionDNA polymerase, Thermoscientific, Waltham, MA, USA) was used to amplify the Internal Transcribed Spacer 1 (ITS1) and ITS2 regions of the isolated genomic DNA via JB206/205 primers (5' GGAAGTAAAAGTCGTAACAAGG 3' and 5' TCCTCCGCTTATTAATATGC 3') yielding an expected amplicon of approximately 700-750 base pairs [155]. We also amplified the D1/D2 portion of the 28S rRNA large subunit (LSU) gene with the NL1/NL4 primers (5' GCATATCAATAAGCGGAGGAAAAG 3' and 5' GGTCCGTGTTTCAAGACGG 3') [111]. DNA was amplified with the following PCR settings for 30 cycles: annealing at 56 ˚C, elongating for 60 seconds at 72 ˚C and melting at 98 ˚C. All of the same conditions were used for the LSU PCR reaction except the annealing temperature was changed to 67 ˚C [111]. DNA amplification was checked on an agarose gel and imaged with a c600 Azure Biosystems imager. PCR products were concentrated with the Zymogen DNA Clean & Concentrator (Zymo Research, Irvine, CA,USA) and submitted to GENEWIZ (South Plainfield, NJ, USA) for sanger sequencing. Forward and reverse sequence reads of the ITS1 and ITS2 region were joined into a single contig by trimming the ends of reads with poor base calls (>3 Ns in a 20 base window) and assembling readswith 85% overlap over at least 20 bps with the contig assembly feature in GeneStudio bioinformatics package (ver. 2.2.0.0, GeneStudio, Inc., Suwanee, GA, USA). Phylogenetic reconstruction was performed using MEGA7 (v 7.0.14). Due to the lack of homogeneity in coverage across the ITS1 and ITS2 sequences in gut fungi, only ITS1 and LSU sequences were used [111]. ITS1 and LSU sequences were analyzed with the maximum likelihoodmethod using a Tamura Nei nucleotide substitution model with 1000 bootstrap replications to estimate the confidence in node clustering.

4.5.3 DNA extraction and sequencing

While the Qiagen extraction kit generates sufficient quality and quantity DNA for phylogenetic analysis via PCR, genomic DNA sequencing from these cultures is limited due to the shearing associated with the lysis beads and silica clean-up columns. Therefore, we set out to

improve this method. Fungal biomass for genomic DNA isolation was obtained for each by pooling six 50 ml cultures grown in serum bottles of Medium B [183] with glucose (0.45%) as the carbon source. Cultures were grown three to four days at 39 °C to generate mats of fungal biomass, after which the media was separated from the biomass and excess media was removed by centrifugation (15,000g for five minutes) before the biomass was frozen at -80 °C until extraction. Before extracting the DNA, a mortar and pestle were chilled with liquid nitrogen and then the frozen biomass (~1.5g wet weight, approximately 150-200 mg) was transferred into the mortar and further chilled with liquid nitrogen; approximately 500 µl of zirconia beads were added to the biomass as well to help with the subsequent lysis step. Fungal tissue was then ground in liquid nitrogen with mortar and pestle, and the powder was transferred to a 50mL falcon tube. Immediately 0.1g PVP (polyvinylpyrrolidone, molecular weight 40,000) powder was added on top of the ground tissue followed immediately by 20mL 2X CTAB (0.2% (w/v) CTAB, 0.8%, (w/v), .20 mM EDTA, 100 mM Tris) buffer that was pre-warmed to 65 °C. Next, 200 µL β-mercaptoethanol was added from the stock solution followed by 20mg Proteinase K powder and the tube was mixed. To promote lysis the tube was immediately placed in a 50 °C water bath for 30min with intermittent gentle inversions every five to ten minutes to keep the tissue suspended in the buffer. After 30 minutes, the tube was cooled to room temperature for five minutes. Next, 10mL 5M Potassium acetate (not pH adjusted) was added and gently mixed by inverting the tube by hand about 40 times. After mixing, the tube was put on ice for 30-40 minutes with no additional mixing and the solution began to turn milky. After cooling, 20 ml of phenol:chloroform: isoamyl alcohol (P:C:I, 25:24:1) was added, gently inverted by hand 40 times, and placed on a slow shaker (~60 rpm) at room temperature for 10 minutes. After incubation at room temperature, the tubes were centrifuged at 4 °C and 5,000g for 20 minutes. The aqueous phase was then transferred into a fresh 50ml tube being careful not to disturb the interphase. To the aqueous phase, 200uL of 50mM PMSF was added and shaken slowly on a shaker (~60 rpm) for 10 minutes at room temperature. Next, 20 ml chloroform: isoamyl alcohol (C:I, 24:1) was added and gently mixed by hand 40 times and then place in a slow shaker at room temperature for 10 minutes. After incubation at room temperature, the tube was spun at 5000g for 20 minutes at 4 °C. The aqueous phase was transferred into a new 50 ml tube being careful not to disturb the interphase. Next, 2/3 volume isopropanol (~20 ml) was added and mixed very gently until the phases are no longer separate, and then spun at 5000g for 15 minutes. After spinning, the liquid was decanted and the pellet was washed with 70% ethanol

and spun at 5000g for five minutes; the wash step was repeated twice decanting the ethanol after each. The pellet was air-dried 15-30 minutes and then dissolved at room temperature overnight in 5 mL in TE. In the morning, 50uL of RNAse A (10mg/mL) was added and incubated at 37 °C for 1 hour. The RNAse was precipitated out by adding equal volumes chloroform: isoamyl alcohol (C:I, 24:1), mixing gently by hand 40 times and placing it in a slow shaker (~60 rpm) at room temperature for 10 minutes. After shaking, the tube was spin at 5000g for 15 minutes at 4 °C and the aqueous phase (top phase) was carefully transferred into a new 50 ml tube without disturbing the interphase. To the aqueous phase, 0.3 volumes (~1.5 ml) of 100% ethanol was added and gently mixed 40 times. Next, the tube was spun at 5000g for 20 minutes at 4 °C. Avoiding any pellet, the DNA solution was transferred to a new 50ml tube. About 10 ml (1.7X) volumes of 100% ethanol was then added to the DNA solution, gently mixed 40 times, and left on the bench 5-10 minutes to precipitates the final DNA. Afterward, the solution was spun at 5000g for 20 minutes at 4°C and the supernatant liquid was decanted. The DNA pellet was washed with 70% ethanol, spun for five minutes and decanted two times leaving the clean DNA pellet in the tube. The pellet was air-dried 15-30 minutes and dissolved in 100-300uL TE depending on the size of the DNA pellet; the pellet was allowed to hydrate in TE overnight at room temperature and then checked for quantity and quality by nanodrop, Qubit HS DNA fluorimetry, and by agarose gel and CHEF electrophoresis.

In collaboration with the Joint Genome Institute and Hudson Alpha, the genomes we sequence using the PacBio platform and assembled with either MECAT (*Piromyces sp. UH3-1*) or Flye (*Neocallimastix* isolates). Briefly, PacBio libraries were prepared by treating the gDNA with DNA Prep to remove single-stranded ends and then with DNA damage repair mix, followed by end repair, A tailing, and ligation of PacBio overhang adapters using SMRTbell template preparation kit 1.0 (Pacific Biosciences). The final library was size selected with BluePippin (Sage Science) at a 20-kb cutoff size and purified with AMPure PB beads. PacBio Sequencing primer v3 was then annealed to the SMRTbell template library, and sequencing polymerase was bound to them using a Sequel binding kit 3.0. The prepared SMRTbell template libraries were then sequenced on a Pacific Biosciences RSII sequencer using four hour sequencing movie run times [35].

Chromatin conformation capture data was generated using a Phase Genomics (Seattle, WA) Proximo Hi-C 2.0 Kit, which is a commercially available version of the Hi-C protocol [184].

Following the manufacturer's instructions for the kit, intact cells from two 50 ml cultures were crosslinked using a formaldehyde solution, digested using the DPNII restriction enzyme, and proximity ligated with biotinylated nucleotides to create chimeric molecules composed of fragments from different regions of the genome that were physically proximal in vivo, but not necessarily genomically proximal. Continuing with the manufacturer's protocol, molecules were pulled down with streptavidin beads and processed into an Illumina-compatible sequencing library. Sequencing was performed on an Illumina HiSeq 4000, generating a total of 175,956,490 PE150 read pairs.

The draft assembly (AT287932.20191203.main.contigs.fa) was used along with provided long-read data to run purge_haplotigs. The following cutoffs were used based on the read coverage histogram: -l 25 -m 120 -h 190. Reads were aligned to the primary contigs output by purge_haplotigs (curated.FALC.fasta) also following the manufacturer's recommendations (https://phasegenomics.github.io/2019/09/19/hic-alignment-and-qc.html). Briefly, reads were aligned using BWA-MEM [185] with the -5SP and -t 8 options specified, and all other options default. SAMBLASTER [186] was used to flag PCR duplicates, which were later excluded from analysis. Alignments were then filtered with samtools [187] using the -F 2304 filtering flag to remove non-primary and secondary alignments. FALCON-Phase [188] was used to correct likely phase switching errors in the primary contigs and alternate haplotigs from FALCON-Unzip and output its results in pseudohap format, creating one complete set of contigs for each phase.

Phase Genomics' Proximo Hi-C genome scaffolding platform was used to create chromosome-scale scaffolds from FALCON-Phase's phase 0 assembly, following the same single-phase scaffolding procedure described in Bickhart et al. [189]. As in the LACHESIS method [190], this process computes a contact frequency matrix from the aligned Hi-C read pairs, normalized by the number of DPNII restriction sites (GATC) on each contig, and constructs scaffolds in such a way as to optimize expected contact frequency and other statistical patterns in Hi-C data. Approximately 20000 separate Proximo runs were performed to optimize the number of scaffolds and scaffold construction in order to make the scaffolds as concordant with the observed Hi-C data as possible. This process resulted in a set of 10 scaffolds containing 81.9 Mbp of sequence (97.4% of the input assembly). Juicebox[191], [192] was then used to correct scaffolding errors, and FALCON-Phase was run a second time to detect and correct phase switching errors that were not detectable at the contig level, but which were detectable at the scaffold level. Juicebox,

additionally, was used to visualize the centromeric regions (Figure 4-3). Metadata generated by FALCON-Phase about scaffold phasing was used to generate matching .assembly files (a file format used by Juicebox) and subsequently used to produce a fully-phased, chromosome-scale set of scaffolds using a purpose-built script (https://github.com/phasegenomics/juicebox_scripts).

4.5.4 Annotation, identification of horizontal gene transfer, and natural product predictions

All genomes were annotated using the JGI Annotation Pipeline and are available via the JGI fungal portal MycoCosm (http://genome.jgi.doe.gov/neocallimastigomycota/). KOG, CAZyme, and BGC data were obtained from this portal. Once genome annotation is complete for the Neocallimastix isolates, genome assemblies and annotations will also be deposited at GenBank. HGT events were identified using a previously developed pipeline [35] that combines BLAST and phylogenetic tools to identify potential HGT events. Here it was modified to include non-Neocallimastigomycota fungi, plants, metazoans, other eukaryotes, viruses, and archaea. In addition to the biosynthetic gene cluster predictions of the JGI pipeline, ARTS [176] was used under default discovery parameters to identify potential biosynthetic gene clusters and any potentially associated genes in close proximity.

4.5.5 Flow Cytometry

To evaluate at the ploidy of anaerobic fungal nuclei, we first extracted nuclei using an in-house developed method and then evaluated the nuclear content with flow cytometry. To extract the nuclei, frozen fungal biomass (~1.5-2g wet weight) from six 50 ml cultures were ground under liquid nitrogen in a chilled mortar and pestle with zirconia beads in the same manner as was done for gDNA extraction. After grinding, about 20 ml of chilled Homogenization Buffer(10 mM Trizma base, 0.5M Sucrose, 60mM KCl, 14 mM NaCl, 5 mM MgCl2, 1 mM CaCl2, 1% W/V Triton X-100) was added to the ground tissue (100ml for every 10g of ground tissue). The homogenization slurry was then agitated on a stir plate at 4 °C for 20 minutes to promote swelling and lysis. Particulate matter and zirconia beads were removed by filtering through two layers of prewetted cheesecloth that was pre-wetteted with sterile cold Milli-Q water for prewetting. Keep everything cold on ice, the slurry was further lysed with a Dounce homogenizer using the loose pestle. The homogenate was centrifuged at 1800g for 15 minutes in a swinging bucket rotor at 4

°C. The pellets were very gently resuspended in 20-30 ml of Nuclei Wash Buffer (10 mM MES, 0.5M Sucrose, 60 mM KCl, 14 mM NaCl, 5 mM MgCl2, 1 mM CaCl2, 0.5% W/V Triton X-100) and filtered through one layer of prewetted Miracloth. The filtered homogenate was centrifuged again at 1800g for 10 minutes at 4 °C and the pellet resuspended in 10 ml of chilled Nuclei Wash Buffer. Next a Percoll gradient was constructed by overlaying 5 ml of 60% Percoll on ice-cold 2.5 M sucrose and then gently adding the resuspension without mixing the layers of the gradient. The gradient was then centrifuged at 1200 g for 30 minutes at 4 °C in a swinging bucket rotor. After centrifugation, the liquid above the gradient was discarded and the 60% Percoll layer (about 5 ml) was collected. The 60% Percoll gradient was then diluted with five volumes of Nuclei Wash Buffer and incubated on ice with gentle shaking (~60 rpm) for 10 minutes. The nuclei were then pelleted by centrifuging at 1800g for 10 minutes at 4 °C, and then resuspended in 5 ml of wash buffer. Nuclei were then carefully overlaid on 5 ml of 35% Percoll solution in Nuclei Wash Buffer and then centrifuged at 1200g for 10 minutes at 4 °C. Nuclei were washed by resuspending the pellet in 5 ml of Nuclei Wash Buffer and centrifuge at 1800g for 10 minutes at 4C. Nuclei finally resuspended in 2.5 ml of Nuclei Wash Buffer and mixed with 2.5 ml of propidium iodide (PI) staining solution (40mg/L PI, 500mg/L RNAse in PBS, 70% methanol) [193]. Nuclei were stained for approximately 30 minutes and were then sampled from by the Attune NxT (Applied Biosystems) equipped with a Blue Excitation Laser (488 nm), Red Excitation Laser (638 nm), Violet Excitation Laser (405 nm), and Yellow Laser (561 nm). Signals from approximately 60.000 events were captured in the PI channel (excitation: 488nm, emission filter: 610nm) at approximately 2,000 events/sec. An acquisition protocol was defined to measure forward scatter (FSC) and side scatter (SSC) on a log scale and mean fluorescence intensity on a log scale using the Attune™ NxT software (ThermoFisher). Data were analyzed with FlowJo (Tree Star, OR, USA) software version 10 and visualized with R scripts.

4.5.6 Promoter identification

Promoters were identified by pairing transcriptomics with genomics and leveraging our genomes. Transcriptomes were acquired from a previous dataset for *Piromyces finnis* and *Neocallimastix californiae* in a previous study of how anaerobic fungi degrade various feedstocks and thus provided expression data across a variety of conditions [18]. To identify constitutive promoters, genes that were not expressed across all conditions were removed. Similarly,

differentially expressed genes were identified with DESeq2 and excluded from potential promoter candidates. Using the expression level, genes were then sorted into high, medium, or low expression groups based on the distribution of expression and their corresponding expression level. Finally, to narrow in on genes that could be identified in all anaerobic fungi, we narrowed our list to only genes that were present in all genomes with at least KO level annotations [194]. To validate that the regulatory regions of these genes were conserved across anaerobic fungi, we compared 1000 bp of the region upstream of the translational start site of all the available genomes on Mycocosm. These upstream putative regulatory regions were then pulled from the genomes of our isolates to be synthesized by the JGI and will be screened for activity with the pipeline of basic genetic tools (Chapter 7).

5. HYDROLYSIS OF LIGNOCELLULOSE BY ANAEROBIC FUNGI PRODUCES FREE SUGARS AND ORGANIC ACIDS FOR TWO-STAGE FINE CHEMICAL PRODUCTION WITH *KLUYVEROMYCES MARXIANUS*

This chapter has been adapted from a manuscript recently submitted as a research article

5.1 Abstract

Development of the bioeconomy is driven by our ability to access the energy-rich carbon trapped in recalcitrant plant materials. Current strategies to release this carbon rely on expensive enzyme cocktails and physicochemical pretreatment, producing inhibitory compounds that hinder subsequent microbial bioproduction. Anaerobic fungi are an appealing solution as they hydrolyze crude, untreated biomass at ambient conditions into sugars that can be converted into value-added products by partner organisms. However, some carbon is lost to anaerobic fungal fermentation products. To improve efficiency and recapture this lost carbon, we built a two-stage bioprocessing system pairing the anaerobic fungus *Piromyces indianae* with the yeast *Kluyveromyces marxianus,* which grows on a wide range of sugars and fermentation products. In doing so we produce fine and commodity chemicals directly from untreated lignocellulose. *P. indianae* efficiently hydrolyzed substrates such as corn stover and poplar to generate sugars, short chain fatty acids, and ethanol, which *K. marxianus* consumed while producing 2.4 g/L ethyl acetate. An engineered strain of *K. marxianus* was also able to produce 550 mg/L 2-phenylethanol and 150 mg/L isoamyl alcohol from *P. indianae* hydrolyzed lignocellulosic biomass. Despite the use of crude untreated plant material, production yields were comparable to optimized rich yeast media due to the use of all available carbon including organic acids, which formed up to 97% of free carbon in the fungal hydrolysate. This work demonstrates that anaerobic fungal pretreatment of lignocellulose can sustain the production of fine chemicals at high efficiency by partnering organisms with broad substrate versatility.

5.2 Introduction

Renewable plant biomass represents a rich source of trapped carbon poised to accelerate the growth of the bioeconomy [1], [2]. These feedstocks are inexpensive, ubiquitous, and rich in carbohydrates (making up to 55-75% of their composition) that can be used in a broad range of bioprocesses [3]. Current and emerging technologies convert these feedstocks into a variety of products including biogas and biofuels, bioplastics, surfactants, oils and fatty acids, food additives, detergents, adhesives, lubricants, paints, and pigments that together make up a global market of over $48 billion each year [1], [5]–[7]. However, current strategies to degrade recalcitrant plant material rely on expensive enzyme cocktails and/or harsh pretreatment techniques that produce compounds inhibitory to subsequent microbial growth and bioproduction [7]–[10]. On the other hand, fungal pretreatment of biomass could reduce costs and increase efficiency of lignocellulose decomposition because it is low energy, low cost, and does not create inhibitory compounds [16].

Anaerobic fungi are an appealing solution to this problem as they are able to degrade crude, untreated biomass into fermentable sugars that can be converted into a wide variety of products [11]–[13]. One example, *Piromyces indianae (Piromyces sp. UH3-1),* has been shown to degrade a variety of feedstocks including food waste and lignin-rich poplar [14]. This substrate flexibility is due in part to the fact that anaerobic fungi have the largest repertoire of biomass degrading enzymes in the fungal kingdom[15] – even larger than the well-known biomass degraders *Trichoderma reesei*,[10] white rot fungi,[16] and wood decay fungi [17]. More importantly, enzymes produced during cultivation of anaerobic fungi were competitive with commercial *Aspergillus* and *Trichoderma* cocktails on untreated biomass [18]. As native symbionts of the rumen of large herbivores, these fungi have adapted strategies to efficiently degrade the material ingested by their hosts. One such strategy being the ability to tailor the Carbohydrate Active enZymes (CAZymes) that they express in response to the substrates that they encounter [13], [14], [18]. These strategies require that they establish synergy with archaea and other microbes present through cross-feeding and niche degradation [168], [175], [195]. Until their potential can be realized with genetic engineering tools,[11], [23], [196] the value of anaerobic fungi lies primarily in their ability to degrade a wide variety of recalcitrant renewable materials.

By pairing the extensive degradative abilities of anaerobic fungi with other microbes that specialize in production of chemicals and fuels, we can create cost-effective, sustainable bioproduction platforms. Previously, the range of organisms that can partner with anaerobic fungi

was explored *in silico* [197]. However, because of the mismatch in growth rates and metabolic needs of many potential partnering organism, most organisms cannot be cocultured with anaerobic fungi. Instead, a two-stage platform must be pursued where anaerobic fungi hydrolyze lignocellulosic biomass in the first stage into fermentable sugars that microbes in the second stage convert to product. Such a two-stage bioproduction system has been demonstrated for anaerobic fungi partnered with *E. coli* and *S. cerevisiae* [12], [13]. These studies, however, focused primarily on the production and utilization of sugars while regarding the anaerobic fungal produced organic acids as byproducts of fermentation. The organic acids, however, represent a significant carbon source, frequently larger than the free sugars, which can be efficiently captured and used by a variety of organisms.

Kluyveromyces marxianus is one such microorganism that can consume lactate, acetate, and ethanol in addition to the liberated sugars [198]–[201]. Often noted as the fastest growing eukaryote, *K. marxianus* is a facultative anaerobic yeast that produces large amounts of ethyl acetate (EtAc) when metabolizing glucose, whey, and other substrates [202], [203]. In addition to EtAc, which can be used as a solvent or scent in perfumes, *K. marxianus* produces a variety of esters and branched alcohols that may serve as next generation biofuels [204], [205]. Recent work has demonstrated the biosynthesis of 2-phenyl ethanol[204] (2-PE), a higher alcohol that can be used a food additive or potentially a biofuel, and isoamyl alcohol [206] (IsA), a solvent and flavoring chemical [207]. With a complete gene-editing toolbox [198], [204], [208], [209] and wide range of culturing conditions,[210] *K. marxianus* is an emerging organism for bioproduction.

In this study, we pair *P. indianae* with engineered *K. marxianus* and evaluate the ability of a two-stage system to produce fragrances and advanced biofuels. The substrate versatility of *K. marxianus* enables capture of more available carbon while broadening the range of products generated via anaerobic fungal pretreatment. We demonstrate comparable production yields in a two-stage system relative to *K. marxianus* bioproduction on rich media for EtAc, 2-PE, and IsA, and confirm the use of ‘waste’ fermentation acids such as acetate and lactate. This work provides further evidence for the feasibility of two-stage bioproduction systems with anaerobic fungal pretreatment and demonstrates the production of high-value products directly from inexpensive agricultural wastes.

5.3 Results and Discussion

5.3.1 Anaerobic fungi pretreatment of lignocellulose supports growth of *K. marxianus*

To evaluate the feasibility of a two-step bioproduction system pairing *K. marxianus* and *P. indianae*, we tested how well the yeast grows on anaerobic fungal media before and after anaerobic fungal hydrolysis of cellulose-rich substrates. Because *K. marxianus* is not a cellulolytic organism, it was unable to grow on fresh unhydrolyzed anaerobic fungal media containing Sigmacell, a form of cellulose, as a carbon source (Figure 5-1A). In contrast, *K. marxianus* grew on spent *P. indianae* media up to about 60% of that on YPD containing 2% glucose, a standard yeast growth medium. Because spent media sustains *K. marxianus* growth, it suggests the anaerobic fungus liberates enough free carbon to make a two-stage bioprocessing system feasible. To optimize growth of *K. marxianus*, we also investigated the effect of anaerobic fungal media type and lignocellulosic feedstock.

Anaerobic fungi are typically grown on one of two types of media: Medium B [183] or Medium C [150] (Appendix D: Figure D.1). Medium C is a richer medium that is rumen fluid based while Medium B is a minimal medium that is more defined and contains only 20% of the undefined yeast extract and 10% of the undefined peptone components compared to Medium C. We evaluated *K. marxianus* growth in Medium B and Medium C supplemented with 12 g/L Sigmacell after *P. indianae* hydrolysis. The richer Medium C hydrolysate allowed *K. marxianus* to grow to an OD_{600} about 3-fold higher than Medium B (Figure 5-1B). Growth in Medium B was similar to the standard defined (SD) yeast minimal medium containing 20 g/L glucose. However, as Medium C supports more yeast biomass, we selected it as the primary medium for our bioproduction studies.

While *P. indianae* can break down both simple and lignin-rich plant material without any genetic engineering[14], it is not clear if their hydrolysates from various feedstocks could sustain subsequent yeast cultures. Here, we compared *K. marxianus* growth on a variety of hydrolyzed feedstocks that ranged from simple to complex and lignin-rich. Specifically, we used Sigmacell as an example of crystalline cellulose, as well as three lignocellulosic feedstocks: corn stover, poplar and alfalfa. After four weeks hydrolysis of media supplemented with 12 g/L of Sigmacell, each of the resulting hydrolysates supported growth of *K. marxianus* to similar levels (OD_{600} between 25 and 30) in spent Medium C (Figure 5-1C). This suggests that *P. indianae* hydrolysis of both simple and complex cellulosic substrates provided similar levels of usable carbon for sustaining *K.*

marxianus growth, or at least enough carbon so that another nutrient became limiting. Similar systems have previously demonstrated that the model organisms *E. coli* and *S. cerevisiae* could be paired for ethanol production on other anaerobic fungus hydrolysates [12], [13]. Growth of *K. marxianus* on hydrolysates, however, shows that we can expand this type of bioproduction platform to include non-model organisms that produce various other industrially useful compounds [203], [204], [206], [210].

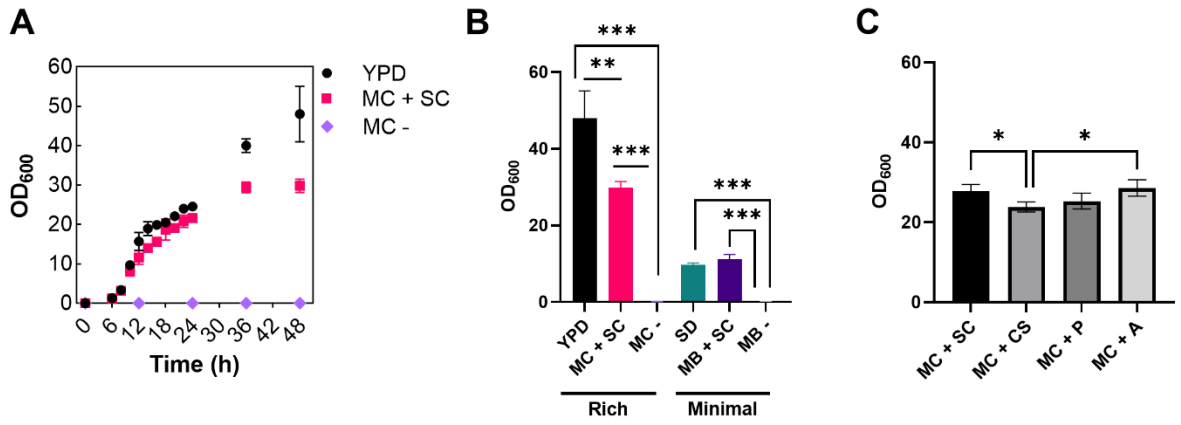


Figure 5-1: Aerobic growth of *K. marxianus* on anaerobic fungal-treated lignocellulosic feedstocks. (A) Growth curves of *K. marxianus* on standard yeast media (YPD) and anaerobic fungal media with Sigmacell before (MC -) and after *P. indianae* hydrolysis (MC+SC). (B) Final OD of *K. marxianus* after growth for 48 hours. (C) Final OD of *K. marxianus* after growth for 48 hours as a function of feedstock used in the spent fungal media. Error bars represent standard deviation of triplicate cultures. MC = Medium C, MB = Medium B, SC = Sigmacell, CS = corn stover, P = Poplar, A = Alfalfa. Results from Unpaired T-tests are indicated by: * = $P < 0.05$, ** = $P < 0.005$, *** = $P < 0.0005$.

5.3.2 *K. marxianus* growth is primarily supported by anaerobic fungal generated fermentation products

Anaerobic fungus release carbon in the form of free glucose, xylose, lactate, formate, acetate, and ethanol from lignocellulosic feedstocks [12]. However, it is unclear how these products vary with substrates or are used by *K. marxianus.* We tracked sugar and organic acid levels as a function of time in both stages to determine what carbon was being released by *P. indianae* and which was used by *K. marxianus.* With anaerobic fungi, growth phase is tracked by monitoring culture pressure (Figure 5-2A), as particulate lignocellulosic biomass interferes with conventional OD measurements [116]. Accumulation of free carbon in the medium does not occur

until approximately the mid-late log stage of *P. indianae*'s growth (Figure 5-2B and C). In stationary phase (≥ 96 h), the fungus stops producing organic acids although its CAZymes continue to liberate free sugars from the feedstocks. Enzymatic hydrolysis continues long after *P. indianae* growth, up to at least three weeks after *P. indianae* entered stationary phase, consistent with observations in other anaerobic fungi [12], [13]. After four weeks of *P. indianae* hydrolysis, there was up to 8.7 g/L glucose or 96 mM acetyl-CoA equivalent metabolites from glucose and up to 200 mM of acetyl-CoA equivalent metabolites in the form of xylose, arabinose, lactate, formate, acetate, and ethanol (Figure 5-2D).

The form of the available carbon varied dramatically with feedstock composition. Free sugars were only released from Sigmacell, a more amorphous synthetic cellulose that is >90% glucose, and corn stover (45.0 ± 1.1 % glucan, 26.4 ± .04 % xylan, 4.1 ± 0.9% arabinan, 17.7 ± 0.4 % lignin) (Figure 5-2E) [14]. Sigmacell was hydrolyzed better than the other feedstocks with the hydrolysate reaching about 8.7 g/L (or 48 mM) of free glucose. Hydrolysis of lignocellulose, however, seemed to be hindered by the complex structure (Table D.1). For example, only about 1 g/L (or 7 mM) glucose, xylose, and arabinose are made available from corn stover. In poplar (48.4 ± 0.3 % glucan, 23.6 ± 0.6 % xylan, 0% arabinan, 23.9 ± 0.5 % lignin) and more lignin-rich alfalfa (28.5 ± 4.4 % glucan, 9.5 ± 1.8 % xylan, 3.9 ± 0.1% arabinan, 77.2 ± 1.6 % lignin), released sugars decreased an order of magnitude to < 0.5 mM despite comparable glucan fractions to corn stover. This may potentially be attributed to some CAZymes becoming deactivated by lignin[211] through irreversible binding of the CAZymes onto the feedstocks,[10] or reduced hydrolysis due to lignin and other structural polymers blocking access to hydrolysable carbon. Nonetheless, the dominant carbon source for *K. marxianus* growth was not free sugars but the organic acids generated by the anaerobic fungus. Up to 60 mM of lactate, formate and acetate were produced from anaerobic fungal hydrolysis (Figure 5-2E). Yields for more complex poplar and alfalfa were slightly reduced at 50 mM and 40 mM, respectively, perhaps reflecting reduced substrate hydrolysis. Interestingly, hydrolysis of lignocellulosic feedstocks produced mainly formate and acetate, whereas lactate was the dominant product on Sigmacell. Finally, ethanol was produced at around 0.3-0.4 g/L (15-22 mM) from Sigmacell, corn stover, and poplar, but only about 0.1 g/L (or 6 mM) from alfalfa (Figure 5-2F). It should also be noted that there is a sizeable amount of ethanol present (~125 mM) in the media - some comes from the crude rumen fluid but the majority is from the antibiotic

solution that is supplemented to keep the cultures sterile (Figure 5-2F). Despite the variations in the form of available carbon, however, the total carbon available for *K. marxianus* is similar.

We next evaluated the ability of *K. marxianus* to use the different forms of available carbon. >95% of the free acetyl-CoA equivalents generated by *P. indianae* were consumed by *K. marxianus* (Figure 5-2G). All metabolites except formate were completely consumed from the Sigmacell, corn stover, and poplar hydrolysates (Figure 5-2H, I). Formate is a *K. marxianus* end-product[199] and a small amount is produced (2-9 mM) [212]. Since *K. marxianus* can utilize nearly all of the carbon forms released into the *P. indianae* hydrolysate, we pose that pairing these two organisms will allow us to increase the amount of carbon captured in the final products and reduce the amounts that go to unwanted byproducts.

5.3.3 *K marxianus* converts spent fungal media to high-value commodity chemicals with high efficiency

Since all feedstock hydrolysates supported growth of *K. marxianus*, we assessed bioproduction from each spent media and compared the titers to those on rich YPD media. Here, we evaluated two engineered *K marxianus* strains: WT-u-h and ENGR. The WT-u-h parent strain is a derivative of a high-producing EtAc CBS 6556 strain [203] with *URA3* and *HIS3* knockouts. The high-producing 2-PE strain (ENGR) is a derivative of the parent strain that contains an EAT1 knockout that suppresses *K. marxianus*' ability to produce ethyl acetate and overexpresses $ARO4^{K221L}$, $ARO7^{G141S}$, *PHA2*, and *ARO10* to increase flux to the 2-PE biosynthesis pathway (Figure 5-3A)[204]. The EAT1 knockout also prohibits *K. marxianus* from forming 2-phenylacetate (2-PEAc) or isoamyl acetate (IsAc), acetylated derivatives of 2-PE and IsA, respectively. Depending on whether we want to produce EtAc, 2-PE, or IsA we can select or create new strains that enhance production of a particular product from our spent media.

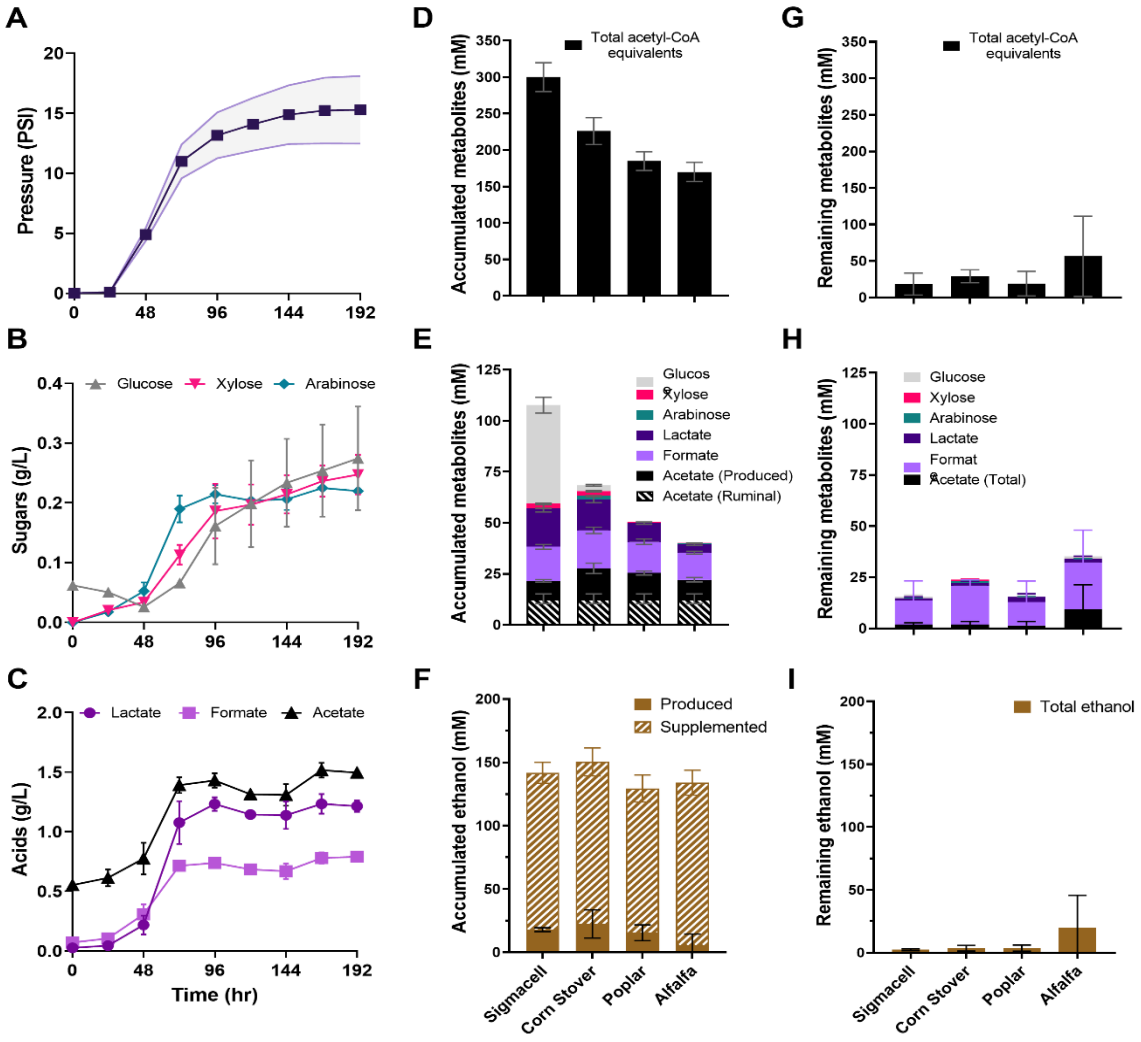


Figure 5-2: Composition of *P. indianae*-treated corn stover Medium C before and after growth by *K. marxianus*. (A) Representative 8-day *Piromyces indianae* growth curve where pressure is used as a proxy for fungal growth[116]; the shaded confidence interval represents the standard deviation. (B) Free sugars released by *P. indianae* hydrolysis. (C) Fermentation acids produced by *P. indianae* growth. (D) Acetyl-CoA equivalents liberated from diverse feedstocks after 28 days, where the total carbon is a sum of the organic acids, sugars, and ethanol. 1 mM glucose = 2 mM acetyl-CoA equivalents (E) Free sugars and organic acids released from diverse feedstocks after 28 days; ruminal acetate = media baseline. (F) Ethanol accumulated in *P. indianae* cultures on diverse feedstocks after 28 days; supplemented ethanol = media background + antibiotic supplemented ethanol. (G) Acetyl-CoA equivalents remaining after *K. marxianus* CBS6556 WT-u-h growth on feedstock hydrolysates (H) Remaining acids and sugars, and (I) ethanol after 48 hours of *K. marxianus* growth. Error bars = standard deviation.

Ethyl acetate is commodity chemical used in several industrial processes as a solvent and fragrance, and it is a product made by wildtype *K. marxianus* strains from ethanol and acetate (Figure 5-3A). We found that the parent WT-u-h strain of *K. marxianus* made 2.4 g/L EtAc on the Sigmacell hydrolysate in Medium C (see Appendix D: Figure D.2 for data on Medium B). This represented a 33% increase from the 1.8 g/L made on YPD, likely due to the high levels of ethanol and free sugars and acids present in spent media. WT-u-h cultures from the lignocellulosic feedstock hydrolysates that contained lower levels of free sugars, however, produced less EtAc reaching about 0.75 g/L for corn stover and around 0.2 g/L for alfalfa and poplar (Figure 5-3). This reduction cannot be explained by the relative abundance of acetyl-CoA equivalents. Lignocellulosic hydrolysates with low sugar levels (corn stover, poplar, alfalfa) exhibited reduced EtAc yields (50%-90% reduction) suggesting that EtAc production requires high sugar (Figure 5-3B). Due to *EAT1* knockout in the ENGR strain, this strain did not produce any EtAc and likely redirected the flux of acetyl-CoA equivalents to other pathways.

In contrast to the EtAc that is formed directly from central metabolism, 2-PE is a product of the Shikimate and Ehrlich pathways. As a fine chemical, 2-PE is a valuable food additive and next generation biofuel due to its increased energy density[205]. Recently, the *K. marxianus* ENGR strain was created to pull flux through the Shikimate pathway and increase titers produced in rich media to the g/L scale [204]. We found that the both the WT-u-h and ENGR strain were able to produce 2-PE from spent media; however, the *EAT1 k*nockout boosted the yield 4-5 fold (Figure 5-3A). For the WT-u-h strain, 2-PE titers scaled with available acetyl-CoA equivalents (Figure 5-3B, C). That is, spent fungal media resulted in comparable production yields of 2-PE in WT-u-h strains. For the ENGR strain, however, rich YPD media produced more 2-PE (908 mg/L) than the richest hydrolysate (551 mg/L), which did not scale with available acetyl-CoA equivalents, potentially due to the formation of alternate products.

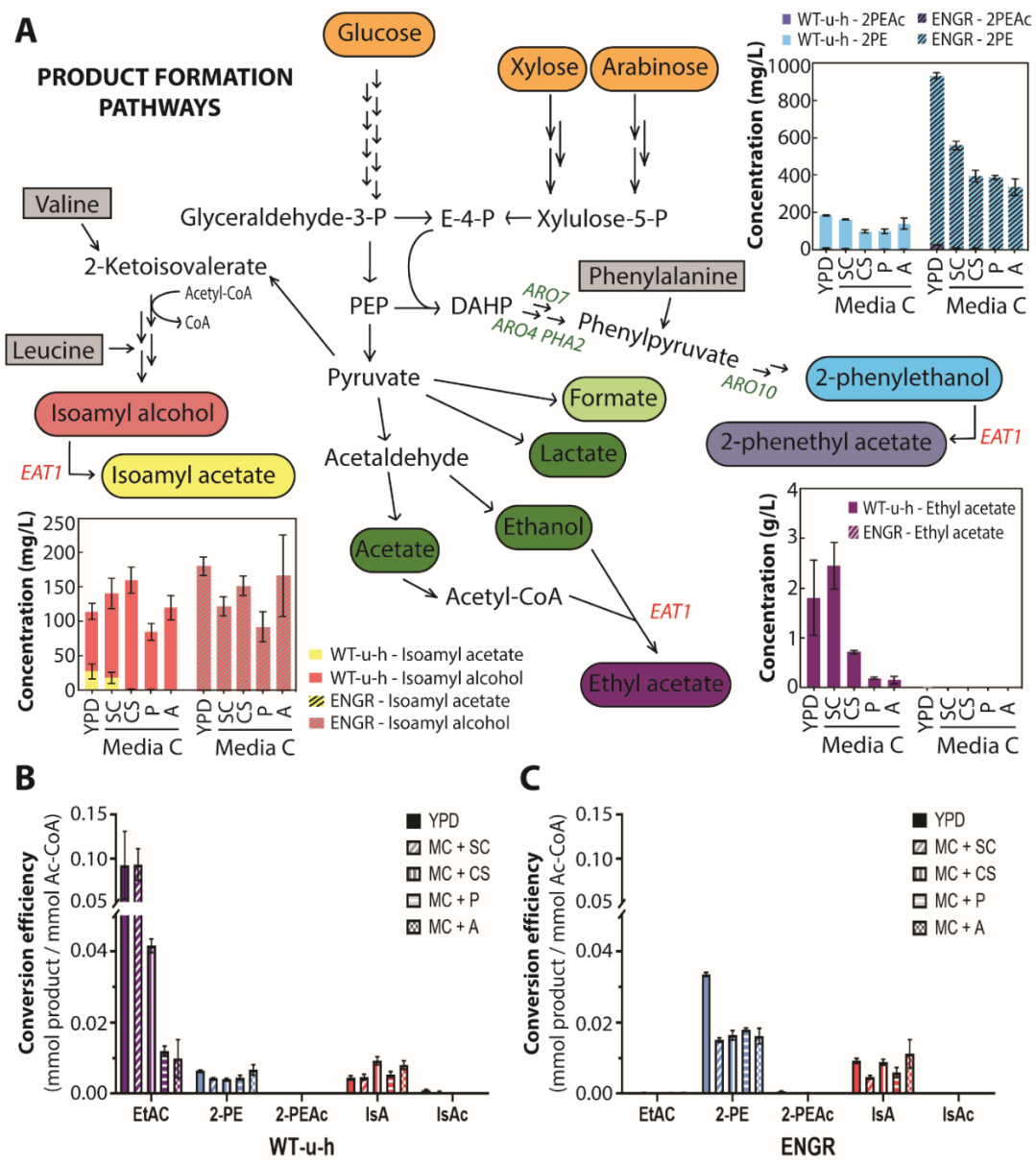


Figure 5-3: *K. marxianus* synthesis of acetate esters and alcohols from *P. indianae* hydrolyzed biomass. A) Metabolic pathways of *K. marxianus* where orange indicates sugars in *P. indianae* hydrolysates, while dark green indicates organic acids and alcohols in the hydrolysates that are consumed by *K. marxianus*. Formate (light green) is also present in the hydrolysate. Product titers form *K. marxianus* cultures are shown next to their biosynthetic pathways for EtAc (dark purple) 2-PE (blue), 2-PEAc (light purple), IsA (red), and IsAc (yellow). WT-u-h (solid) and ENGR (striped) values are shown. Genes knocked out in the engineered strain (ENGR) are shown in red and overexpressed genes are shown in green. Alternate substrates are shown in grey. (B) Conversion efficiencies of *K. marxianus* WT-u-h and (C) *K. marxianus* ENGR to produce bioproducts from the various medias. Conversion efficiency is calculated using the product yields and total free acetyl-CoA in the hydrolysates (i.e. generated acid, ethanol, and sugar + native acids and supplemented ethanol). Error bars = standard deviation.

IsA is valuable solvent, scent, and flavoring produced from the Ehrlich pathway reactions (Figure 5-3A). While the strains here were not specifically engineered for enhanced IsA production, it is made as a byproduct of both WT-u-h and ENGR metabolism (Figure 5-3A, Appendix D: Figure D.2 for Medium B). In YPD medium, the ENGR strain makes about twice as much IsA as the WT-u-h strain, which is consistent with the fact that the *EAT1* knock-out suppresses acetylation from IsA to IsAc, and the overexpression of *ARO10* enhances the downstream pathway biosynthesis from α-ketoisocaproate to IsA. Interestingly, we observed that the IsA levels do not vary between the WT-u-h and ENGR strains when they are grown on the lignocellulosic hydrolysates (Figure 5-3B, C) suggesting there is no change in the flux to IsA when grown in spent media. Overall, our results demonstrate that not only can *K. marxianus* capture and use the released sugars and fungal degradation products, but that they may be converted to product with high efficiency.

5.3.4 *K. marxianus* captures 'lost' carbon from anaerobic fungal pretreatment increasing efficiency

To determine the overall efficiency of a two-stage process, we then evaluated how much of the feedstock hydrolyzed by *P. indianae* was converted to product by *K. marxianus*. Compositional analysis of the substrates before and after anaerobic fungal growth revealed that nearly all (92%) of the loaded Sigmacell biomass was hydrolyzed (Table D.1). For lignocellulosic feedstocks, between 20-50% of the feedstocks were hydrolyzed (Figure 5-4A) although in all cases this encompassed more than 55% of all the available sugars in the substrate (Figure 5-4B). Despite the high utilization of feedstocks, 47-98% of the cleaved sugars were metabolized by *P. indianae* (Figure 5-4C) resulting in the low observed sugar yields (Figure 5-2E). As little as 2% of this sugar was made available for direct conversion to product (Figure 5-4D). That is, current 2-stage bioproduction systems with anaerobic fungi that rely on sugar transfer lose a significant amount of the carbon released from plant biomass. Nonetheless, *K. marxianus* is able to recover the metabolized carbon by using lactate, acetate, and ethanol to grow and generate product. Despite low sugar yields, our two-stage system makes 37-72% of the hydrolyzed carbon available depending on substrate (Table 5-1). The usable acid and ethanol from lignocellulosic hydrolysates account for 78-97% of the available carbon or 3.7- 54 times the amount of carbon from sugars

alone demonstrating the potential that these substrates possess if they are presented to the right organism.

Table 5-1: Carbon released by anaerobic fungi and made available to *K. marxianus*¥

		Sigmacell	Corn Stover	Poplar	Alfalfa
Stage 1	*% hydrolyzed*	99.5%	69.7%	59.0%	70.0%
	Released carbon (mmol)	6.68	3.50	3.04	1.10
Stage 2	*Available carbon (mmol)*	4.84	1.87	1.12	0.67
	Sugars (mmol)	3.40	0.40	0.02	0.04
	Lactate and acetate (mmol)	0.98	0.83	0.64	0.41
	Ethanol (mmol)	0.46	0.63	0.45	0.22
	% hydrolyzed carbon available	72%	53%	37%	61%
	Ratio of avail. acids + ethanol to sugars	0.42	3.65	54.50	15.75

[2]

Ultimately, the exogenous and liberated substrates manifest in product titers of EtAc, 2-PE, and IsA up to 0.20, 0.046, and 0.014 g product/g feedstock hydrolyzed in our two-stage system (Table D.2). These titers compare reasonably well with the yield of ethanol per gram biomass of other two-stage platforms (0.22)[213] including those with anaerobic fungi (0.04 - 0.14)[12], [13]. However, to our knowledge, this is the first report of a two-stage system for producing esters and higher alcohols. A similar approach was recently used to create 82 µg of poly-α-olefins per g of substrate from the hydrolysate of *Clostridium cellulolyticum* and subsequent culture of *Acinetobacter baylyi* ADP1[214]. The olefin production phase was sustained from a hydrolysate that reached a maximum concentration of ~5 mM glucose, 5 mM acetate, and 7 mM lactate produced from crystalline cellulose (Avicel) via *C. cellulolyticum*. However, our anaerobic fungal hydrolysates are able to make available at least twice the carbon (Figure 5-2D) on untreated corn stover demonstrating that anaerobic fungi can hydrolyze even crude untreated agricultural residues to release more carbon than more established strains on purified substrates. Based on the glucan, xylan, and arabinan content, we calculated that our system reached up to 41%, 12%, and 17% of the theoretical yield for various feedstocks (Table D.2). While the theoretical yield is not achievable due to carbon needed for cell biomass and energy, these hydrolysates produced more EtAc, 2-PE, and IsA than would theoretically be yielded by sugars alone for the more complex

[2] ¥amounts of hydrolyzed carbon reported in Acetyl-CoA equivalents

lignocellulosic feedstocks (Table D.2). Taken together, our system increases efficiency by capturing carbon that would otherwise be discarded or viewed as unusable.

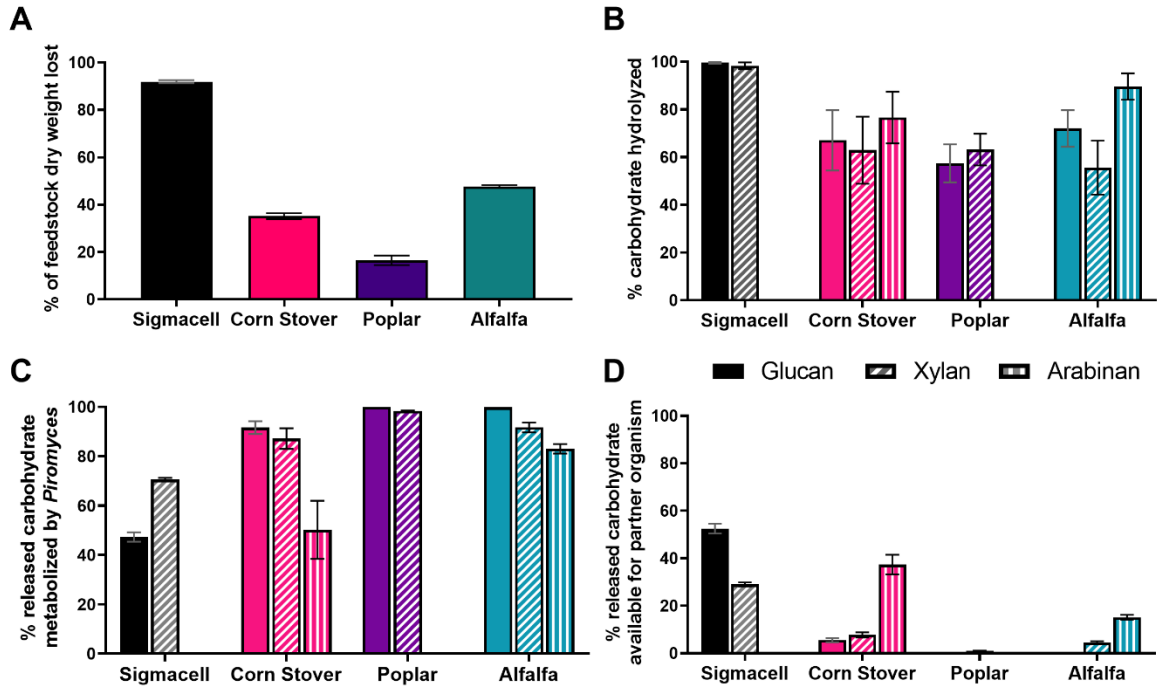


Figure 5-4: Hydrolysis of untreated feedstock by *P. indianae.* (A) Feedstock biomass lost due to hydrolysis. (B) Percentage of glucan, xylan, and arabinan hydrolyzed. (C) Percentage of sugar that was metabolized by *P. indianae* (*e.g.*, was hydrolyzed but not released into the supernatant). (D) Percentage of sugar released into the supernatant. Error bars = standard deviation.

5.3.5 Increasing substrate loading enhances hydrolysis and results in higher yields of liberated sugars.

As formation of EtAc and other products appeared sugar-limited, we assessed whether increasing substrate loading can increase the sugar and acid yields. We doubled the feedstock loading (from 12 g/L to 24 g/L) for all four feedstocks and tracked the metabolites from *P. indianae*. When we doubled the loading of Sigmacell, corn stover, and poplar, the growth of *P. indianae* was unaffected while alfalfa growth increased about 20% (Figure 5-5A). We suspect that the alfalfa cultures were carbon limited under 12 g/L loading conditions because the glucan and xylan content of alfalfa is so low. However, the amount of free glucose from double loaded alfalfa increased 2.53-fold suggesting that it is no longer carbon limited and can generate free sugar. Glucose levels

in the double loaded Sigmacell and corn stover hydrolysates also increased from about 8 g/L to 10.7 g/L (43 mM to 58 mM; 1.28-fold) and from about 0.5 g/L to 1.7 g/L (2.9 mM to 9.1 mM; 3.26-fold), respectively (Figure 5-5B). Similar trends were obtained for hydrolysates in Medium B, albeit at lower yields (Appendix D: Figure D.3). All double-loaded feedstock hydrolysates, however, had increased amounts of free xylose and arabinose, for those with measurable arabinan content (Figure 5-5B). Interestingly, these did not appear to be maximum yields as sugar release did not taper off after 21 days (Appendix D: Figure D.4). In comparison, at 50 g/L loading of reed canary grass *A. robustus* and *N. californiae* also yielded up to 2 g/L glucose as well as up to 1 g/L xylose and 1 g/L arabinose [13]. Here, *P. indianae* hydrolyzed similar levels of glucose and xylose in half the loading of corn stover (24 g/L; Figure 5-5B). The similar sugar yields for our fungi under conditions with half the substrate suggest that yields are a complex function of feedstock composition, substrate loading, and fungal strain. Interestingly, acid production did not change with loading suggesting that given more feedstock, the excess is being hydrolyzed to sugars and not metabolized by the fungus (Figure 5-5C). Thus, increasing substrate loading can produce excess free sugars and should be optimized to improve product formation efficiency in the second stage.

In addition to optimizing system parameters such as substrate loading, efficiency could perhaps be tuned further with a suite of genetic tools for anaerobic fungi to control CAZyme expression and organic acid production. Specifically, the development of gene-editing tools, like CRISPR-Cas9 that allow knock-ins and knock-outs would allow us to exchange weak promoters for stronger ones that overexpress hydrolytic enzymes to enhance feedstock conversion. In addition, these tools could be used to knockout xylose metabolisms genes that may hinder yields in anaerobic fungi. Advanced synthetic biology tools could create kill-switch constructs that stop *P. indianae* metabolism of the feedstock while allowing the hydrolysis to reach higher yields. As we work to advance these organisms for biotechnology and the bio-based economy, these tools will be extremely useful to optimize feedstock conversion and potentially bioproduction by anaerobic fungi.

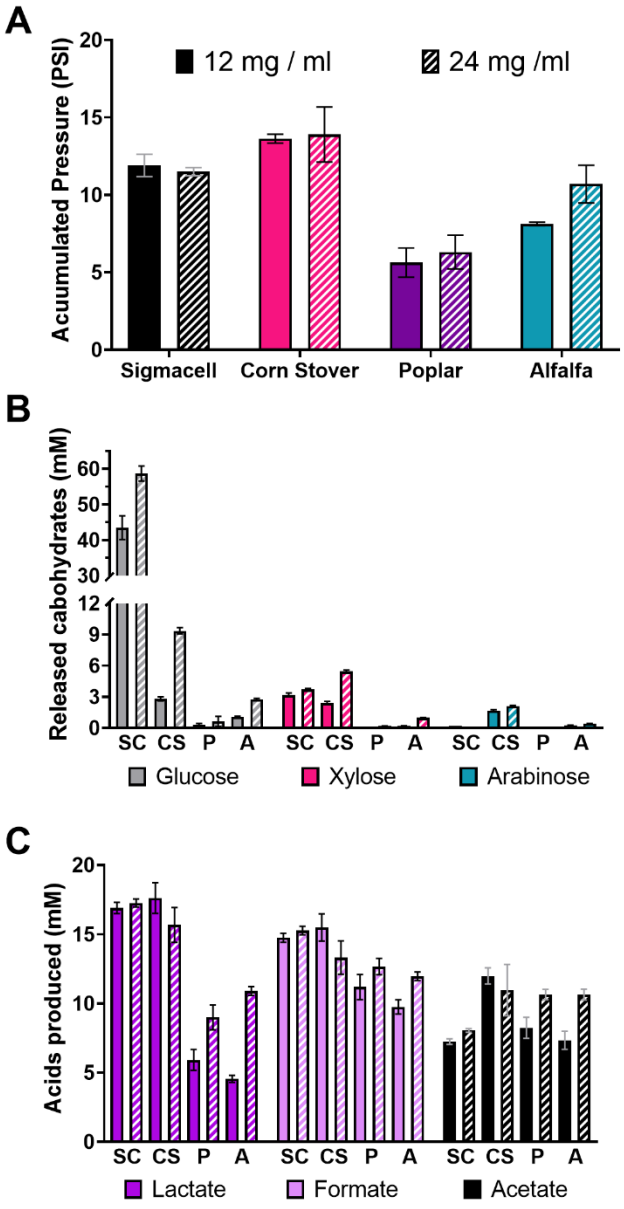


Figure 5-5: Effects of substrate loading on *P. indianae* hydrolysis. (A) Growth of *P. indianae* on Media C with 12 or 24 g/L substrate loading. (B) Released sugars from *P. indianae* hydrolysis with 12 or 24 g/L loading. (C) Organic acids produced from *P. indianae* hydrolysis with 12 or 24 g/L loading. Error bars represent the standard deviation or triplicate experiments.

5.4 Conclusions

We built a two-stage bioprocessing system pairing *P. indianae* and *K. marxianus* to produce fine and commodity chemicals from lignocellulosic feedstocks. We first assessed the ability of *P. indianae* to metabolize a variety of lignocellulosic biomass to generate free sugars, organic acids,

and ethanol. While the simpler feedstocks generated more free carbon, we found that even the most complex substrates could produce enough carbon to support *K. marxianus* growth. Recently engineered strains of *K. marxianus* were shown not only to grow on the hydrolyzed media from *P. indianae* but also to produce ethyl acetate, 2-phenyl ethanol, and isoamyl alcohol. In comparison to commonly used yeast media (YPD), the hydrolysates performed at comparable efficiencies for ethyl acetate and isoamyl alcohol and at half the efficiency for phenylethanol. Despite the low sugar content of some hydrolysates, the organic acids and ethanol was converted to useful product by *K. marxianus*. Finally, we identified substrate loading as a limiting factor for efficient sugars yields that could improve our system. While this two-stage bioproduction system is unoptimized, this work demonstrates the adaptability of this platform to incorporate effective degrader and producer specialists to produce a wide array of chemicals from lignocellulosic waste.

5.5 Materials and Methods

5.5.1 Strains, media, and culture conditions for feedstock hydrolysis by anaerobic fungi

The anaerobic fungus *Piromyces indianae* (*Piromyces sp. UH3-1*) was previously isolated and characterized by Hooker et al [14]. *Piromyces* cultures were inoculated from 50 mL starter cultures containing corn stover biomass in anaerobic Medium C [150] supplemented with 15% v/v rumen fluid. Medium C is a phosphate and carbonate-buffered media containing some yeast extract and casitone as a nitrogen source, and rumen fluid for additional nutrients. Serum bottles with 12 g/L of substrate in 50 mL of either Medium C or semi-defined minimal Medium B (a semi-defined media with more defined nitrogen sources such NH_4Cl as replacing some complex nitrogen and a complex mix of fatty acids, trace metals, hemin, and vitamins in place of rumen fluid)[183] were inoculated with 3 mL of starter culture. Substrates used were Sigmacell (Type 20, Sigma Aldrich), corn stover, alfalfa, or poplar. Plant biomass was milled to 20 mesh. Media and substrate were autoclaved at 121 ˚C for 30 minutes and then supplemented with sterile filtered chloramphenicol dissolved in 60% v/v ethanol (final concentration of 3.5 μg/mL) prior to inoculation with *Piromyces indianae*. After *P. indianae* growth, spent media were centrifuged for 10 minutes at ≥ 12,000 g to separate the biomass for analysis from the supernatant for *K. marxianus* cultures. The hydrolysates were stored at -20 ˚C until they were to be used. To assess the effects of substrate loading, the aforementioned substrates were loaded at either 12 or 24 g/L in 10 mL of either

Medium B or Medium C. These cultures were treated with antibiotic as above, inoculated with 1 mL of a *Piromyces* starter culture, and monitored by HPLC every seven days (See *HPLC analysis of supernatants* below).

5.5.2 Two-step *K marxianus* cultivation on spent media

Kluyveromyces marxianus CBS 6556 *ura3Δ his3Δ* was used as the parent strain for the experiments described here and is referred to here as WT-u-h [204]. The strain engineered for high 2-PE production was derived from the parent as described previously [204]. Briefly, WT-u-h was modified by disrupting ARO8, EAT1, and ABZ1, overexpressing feedback insensitive variants of ARO4 and ARO7, and overexpressing wild type alleles of PHA2 and ARO10. Synthetic defined (SD) medium is defined as 6.7 g/L BD Difco™ Yeast Nitrogen Base without amino acids, 0.79 g/L CSM powder (Sunrise Science Products), and 20 g/L D-glucose. *K. marxianus* strains were also cultivated in rich YPD medium (YPD: 10 g/L Gibco™ Bacto™ Yeast Extract, 20 g/L Gibco™ Bacto™ Peptone, 20 g/L D-glucose). All *K. marxianus* strains were precultured in 1 mL YPD medium for 12 h from single colonies, and then inoculated to an initial OD600 of 0.05 in 10 mL of indicated media in 50 mL baffled shake flasks. Spent *P. indianae* hydrolysates were sparged with compressed air for 10 minutes and sterile filtered before being aliquoted into 10 ml cultures. Culturing was conducted at 250 rpm in an INFORS HT Multitron incubation shaker with temperature control set to 30 °C.

5.5.3 HPLC analysis of supernatants

Supernatants were analyzed for sugars and acids using 1 mL of culture supernatant collected from cultures and controls at indicated time points. The supernatants were stored at -20 ˚C until they were to be analyzed and were then cleared of particulate debris by three cycles of centrifugation at 21,000g for 10 minutes each cycle. Clarified HPLC samples were kept at 4° C prior to analysis in an Agilent 1260 HPLC with an Aminex HPX 87H anion exchange column using a 50 °C, 5 mM H_2SO_4 mobile phase at 0.6 mL/min. An Agilent 1260 Infinity II Refractive Index Detector (RID) (Agilent Technologies, Santa Clara, California, USA) was used to detect analytes over a 30 minute run from a 20 μL injection. Concentrations were determined from standard curves of each analyte prepared from commercial standards.

5.5.4 GC-FID analysis of *K marxianus* products

Analysis of 2-phenylethanol (2-PE), 2-phenylethyl acetate (2-PEAc), isoamyl alcohol (IsA), isoamyl acetate (IsAc), and ethyl acetate was carried out on a Shimadzu GC-2010 Plus equipped with a Shimadzu AOC-20s autosampler and a Shimadzu AOC-20i auto-injector. The GC suite was coupled to a flame ionization detector (FID). Compounds were separated on an Agilent J&W DB-WAX Ultra Inert column (length: 30 m; inner diameter: 0.32 mm; film thickness: 0.5 µm). 2-PE and 2-PEAc were detected and separated using a 21 minute temperature program as follows: start temperature of 100 °C, 20 °C/min to 140 °C, 10 °C/min to 150 °C, 5 °C/min to 160 °C, hold for two minutes then increase by 1 °C/min to 170 °C, hold for two minutes, and finally 25 °C/min to 220 °C. For IsA, IsAc, and ethyl acetate, the temperature was held at 40 °C for two minutes, and then increased at 20 °C/min to 70 °C, finally increased to 220 °C from 70 °C by 50 °C/min, and held at 220 °C for two minutes [215]. Helium was used as the carrier gas at a flow rate of 1.9 mL/min. The sample injection volume was one microliter where split mode was used for injection and the ratio was 20:1.

For sample preparation, a 330 µL sample of cell culture was centrifuged for 1 min at 5000 g. 300 µL of supernatant was collected and transferred to a clean 1.5 mL tube with 300 µL of organic extractant. 300 µL of cyclohexane was used as an extractant for 2-PE and 2-PEAc, and 300 µL of hexane was used for IsA, IsAc, and ethyl acetate. Samples were vortexed thoroughly for 30 min, centrifuged at 10,000g for 1 min, and 100 µL of the organic layer was transferred into a 2 mL clear Agilent GC vial with glass insert. Standard curves depicting the linear correlation between the concentration of five compounds (2-PE, 2-PEAc, IsA, IsAc, and ethyl acetate) with the area of peaks from FID were obtained to quantify extracellular metabolite accumulation under different media conditions. A series of YPD solutions with known concentrations were made and extracted by cyclohexane or hexane accordingly following the same procedure which was used to extract these five compounds from the supernatants of cell cultures.

5.5.5 Compositional analysis of feedstocks

The sugar composition of raw plant biomass was determined according to NREL methods [160]–[162]. Carbohydrates were determined using the same HPLC analysis configuration as

above, (see *HPLC analysis of supernatants* above), however a temperature of 65 ˚C and run time of 45 minutes was used for separation of analytes.

The sugar composition of the plant biomass after inoculation with anaerobic fungi was determined by first separating the fermentation liquor from the solids via centrifugation. Samples were centrifuged for five minutes at 4 ˚C at 8000 g. The wet pellets containing spent plant biomass and anaerobic fungal biomass were dried to constant weight at 45 ˚C. The resulting dry biomass was used for NREL gravimetric analysis and carbohydrate quantification following standard methods. To correct for the amount of anaerobic fungal biomass present in the pellet at the end of the first stage of fermentation, the amount of formic acid measured in the supernatant (See *HPLC analysis of supernatant* above) was used to calculate the fungal biomass yield and was then subtracted from the dry biomass [216].

6. CODON-OPTIMIZATION ENABLED HETEROLOGOUS EXPRESSION OF ANAEROBIC FUNGAL MEVALONATE PATHWAY IN *E. COLI*

This chapter is in preparation for submission as a research article; manuscript preparation is ongoing and will be completed after graduation.

6.1 Abstract

Anaerobic fungi are an emerging biotechnology platform with genomes rich in biosynthetic potential. Heterologous expression of their biosynthetic pathways, however, has not yet been demonstrated in model hosts like *E. coli*. Although *E. coli* are a powerful tool for heterologous expression, previous expression of anaerobic fungal proteins has yielded limited success. One reason for this limited success maybe that gene composition of anaerobic fungi is extremely AT-biased making the genes poor candidates for heterologous expression in *E coli*. Specifically, the codon usage of *P. indianae* is extremely biased toward codons that are not the preferred codons of *E. coli*. When *P. indianae* genes that bear these extreme biases for low abundance codons are highly expressed in *E. coli*, they create growth deficiencies (up to 69% reduction in growth) in *the* model host. Even *E. coli* strains engineered to express AT-biased genes, like BL21-CodonPlus, are not able to fully rescue cell growth (growth reduction up to 57%). Instead, by optimizing the codons distribution to mimic the codon usage of *E. coli,* these genes are expressed without a growth defect due to codon deficiencies. By unveiling the mechanism to express anaerobic fungal genes in *E. coli*, we demonstrated the ability of anaerobic fungal genes to synthesize mevalonate at levels comparable to or greater than the known yeast homologs (up to 2.5 g/L). Ultimately, this work lays the foundation for how additional biosynthetic pathways from anaerobic fungi can be expressed in model hosts like *E. coli*.

6.2 Introduction

Over the last century, fungi have been a source of valuable products in biotechnology from medicines and insecticides to food additives and enzymes [23]. Within the past decade, anaerobic fungi have emerged as an untapped source of enzymes for biotechnology [11], [15], [196]. Genomic studies of anaerobic fungi show they have the largest array of biomass-degrading

enzymes among fungi which can be used for platforms that convert lignocellulosic waste to biofuels and chemicals [13], [15]. In addition to plant biomass-degrading enzymes, anaerobic fungi have unique biosynthetic pathways[217] and may be a valuable source of natural products like other fungi [23], [35]. However, the full potential of these fungi remains unrealized because there are few tools for their engineering.

Heterologous expression partially addresses this gap by evaluating and using promising genes from hard to engineer organisms in more amenable model organisms such as *E. coli* and *S. cerevisiae* [218]. These strategies have been used to make products in the food, pharmaceutical, and agricultural industries [219]. While heterologous expression is a very powerful tool, many potential pitfalls such as codon bias [220], post-translation modification [20], and dissimilarity of host environment [221], [222] can cause expression to fail [54]. Previous attempts of heterologous expression of anaerobic fungal genes have often encountered similar issues in both *E. coli* and *S. cerevisiae* [20], [21]. However, *E. coli* has successfully been used as a host for several proteins including hydrogenosomal and scaffoldin proteins from anaerobic fungi [11], [35]. Despite these successes, issues related to the expression of anaerobic fungal biosynthetic pathways in *E. coli*. have not been resolved.

One of the factors that limit the heterologous expression of anaerobic fungal genes in these hosts is that the genome composition of anaerobic fungi is divergent from most non-model hosts. The genomes of anaerobic fungi have the lowest %GC in the fungal kingdom (~20% GC) and are largely AT-rich in both the intergenic and coding regions [19]. For example, in the recently isolated anaerobic fungus *Piromyces indianae* (*Piromyces sp*. UH3-1) [14], while AT-rich regulatory regions of DNA regions skew the overall %GC lower, the coding regions are still relatively AT-biased around 27.9% GC. On the other hand, the genomes of *E. coli* and *S. cerevisiae* are more balanced around 50% and 38% GC, respectively with little differences between coding and intergenic regions [223], [224]. The disparities in gene composition between anaerobic fungi and *E. coli* and the underlying codon utilization has hindered the heterologous expression of anaerobic fungal genes. Overuse of low abundance codons in model organisms with dissimilar codon usage is anticipated to impact expression strain cell health by depleting tRNA pools vital to the expression of essential genes like those involved in cell proliferation [225].

In this study, we evaluated the effects of expressing native and codon-optimized genes from anaerobic fungi in *E. coli*. We focused on the mevalonate pathway from recently isolated

Piromyces indianae; a pathway composed of three genes (*atoB*, *HMGS*, and *HMGR*) that convert acetyl-CoA into mevalonate – a valuable terpenoid precursor. We demonstrate that without codon optimization, some genes from anaerobic fungi cause moderate to severe growth defects in *E. coli* even in strains designed to express difficult genes. Further, we provide evidence that codon-optimization enables *E. coli* to produce compounds from the biosynthesis pathways of anaerobic fungi, and that specific *P. indianae* homologs increased production approximately 2.5-fold compared to the canonical yeast pathway.

6.3 Results and discussion

6.3.1 Compared to *E. coli* the overall codon usage of *P. indianae* varies greatly throughout the whole genome.

The codon adaptation index (CAI) metric evaluates the potential expression level of a target gene based on its codon usage compared to that of highly expressed genes in the host of interest [226], [227]. The CAI of every *P. indianae* gene with *E. coli* as the host was evaluated with a publicly available expression dataset [227], and the majority of *P. indianae* genes (>98%) were found to have *E. coli* CAI scores under 0.50 (Figure 6-1). The mean and median CAI of *P. indianae* genes are close to 0.35 and 0.34 (Figure 6-1), respectively, which is about 30% lower than the mean and median CAI of the *E. coli* genome (~0.50). Highly expressed genes have a CAI of ~0.7 reflecting an optimal distribution of rare and abundant codons that does not overly deplete any individual tRNA pool. Given the low *E. coil* CAI of *P. indianae* genes, these genes are not likely to be highly expressed in *E. coli*. It is not surprising to see that the mean and median of the *E. coli* genome are close to 0.50 given that its genome and codon usage are largely balanced (~50 %GC) [223] and the distribution is a normal Gaussian distribution like we also across most genomes [228]. The *P. indianae* distribution is also relatively normal while the mean is more tightly centered at a lower CAI suggesting it is not as unbiased as *E. coli*. Furthermore, there are only 10 *P. indianae* genes with a CAI of 0.60 or higher with the maximum CAI around 0.77. However, with the exception of two genes (ribosomal protein S29 and thioredoxin 2), these genes remain unannotated, hypothetical proteins, with no evidence of their origins as determined by BLAST. Given that these genes are more GC rich (34.5 – 49.7% GC), we suspect that they are highly conserved proteins or the result of horizontally transferred genes from prokaryotes in the native rumen [35]. While we

can see most genes use different codons (Figure 6-1), the specific codons that are over- or under-utilized cannot be determined by CAI.

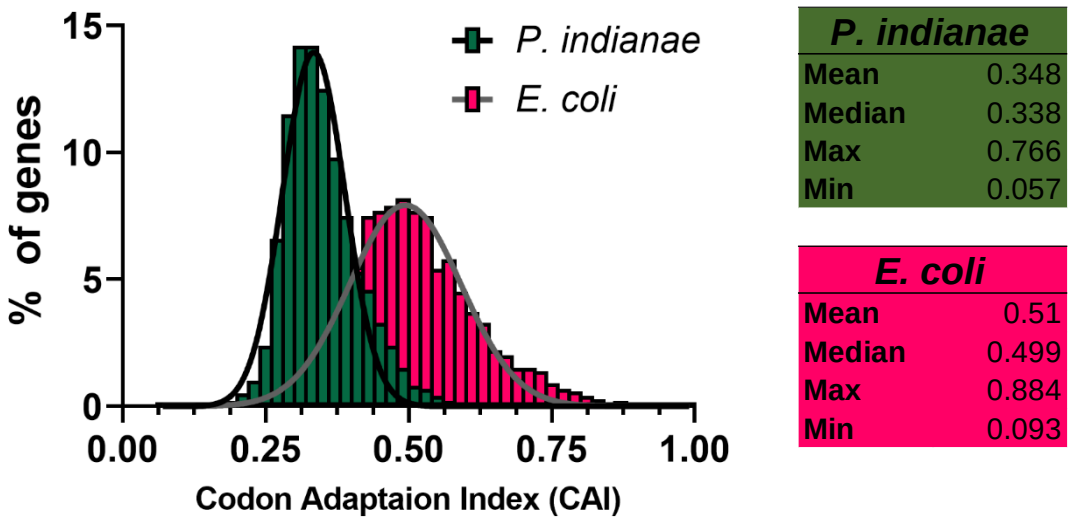


Figure 6-1: Distribution of Codon adaptation index (CAI) scores for *P. indianae* and *E.coli* genes with respect to their fitness in E. coli. Mean, median, maximum, and minimum scores are reported for the distributions. Normal Gaussian distributions are shown for each.

To further explore what makes the *P. indianae* genes poor fits for *E. coli*, we evaluated the relative codon usage for each amino and overall codon usage across the genome of *P. indianae*. The relative codon usage measures the preferred codon for an amino acid in an organism and is the frequency at which a particular codon is used relative to synonymous codons [229]. The overall usage, on the other hand, is the frequency at which a codon is used with respect to the codons for all amino acids. While *E. coli* uses a variety of AT-rich and GC-rich codons [223], anaerobic gut fungi are heavily biased towards AT-rich codons [230]. Our comparison of the relative usage of *E. coli* and *P. indianae* found that the preferred codon was different for 13 out of the 18 amino acids with more than one codon (Figure 6-2); *S. cerevisiae* whose genome is somewhat AT-rich (~39% GC) and *E. coli* only share 8 of the same 18 preferred codons. The largest difference in preferred codons were for asparagine (Q), aspartate (N), arginine (R), proline (P), cysteine (C) and leucine (L) where codon usage of the *P. indianae* codon increased > 40% compared to the frequency that *E. coli* uses this codon. Additionally, for the five codons where the preferred codon was the same, the relative usage rates of *P. indianae* were skewed toward a heavy bias of one codon (using these 83 +/- 8% of the time) whereas *E.coli* balances its usage better (57 +/- 9% of

the time); *S. cerevisiae* is also more balanced in its usage of preferred codons utilizing them 58 +/- 8% of the time (Appendix E: Tables E.1-3). Unlike *E. coli* and other prokaryotes that compartmentalize codon usage for highly expressed genes [226], [228], anaerobic fungal codon usage is uniform across genes regardless of expression level like many eukaryotes (data not shown). These biases suggest that anaerobic fungi heavily prefer specific codons while *E. coli* has a more balanced codon-usage.

In addition to differences in synonymous codon preferences, *P. indianae* genes overused 13 of the 61 amino-acid-calling codons compared to their overall usage in *E. coli* (>2-fold increase in the overall codon usage; Figure 6-2) which is similar to *S. cerevisiae* (14 overused codons compared to *E. coli*). Not surprisingly, these codons were all AT-rich. Out of these 13 codons, two are rare *E. coli* codons (AGA_R and AUA_I; used in less than 0.5% of all codons) and two are semi-rare (UCA_S and AGU_S; used in less than 1% of all codons). Frequent use of rare codons is problematic for expression in *E. coli* because it frequently stalls translation and can create misfolded or truncated proteins [229]. Because these rare codons also have a special function in regulating the rate of protein synthesis and nascent chain folding [225], [231], [232], we suspect that over-expressed genes that consume the already small pools of rare codons could potentially tax the cells and slow down translation and, ultimately, growth [233]. The most overused codons, AAU_N, UUA_L, and AAA_K, however are not rare codons, but are overall used 18.9% of the time collectively in *P. indianae* compared to 6.4% of the time in *E. coli* (Appendix E: Tables E.1 & E.2), and also demonstrates there are also large differences in non-rare codons. Much less work has been done to investigate how the overuse of these codons affects cell physiology [234]. However, the genes of *P. indianae* are intriguing case studies because over a dozen of these codons are heavily used here. Overall, *P. indianae* genes use codons that are AT-rich and less frequent in *E. coli* and could potentially create translational burdens in heterologous hosts when overexpressed.

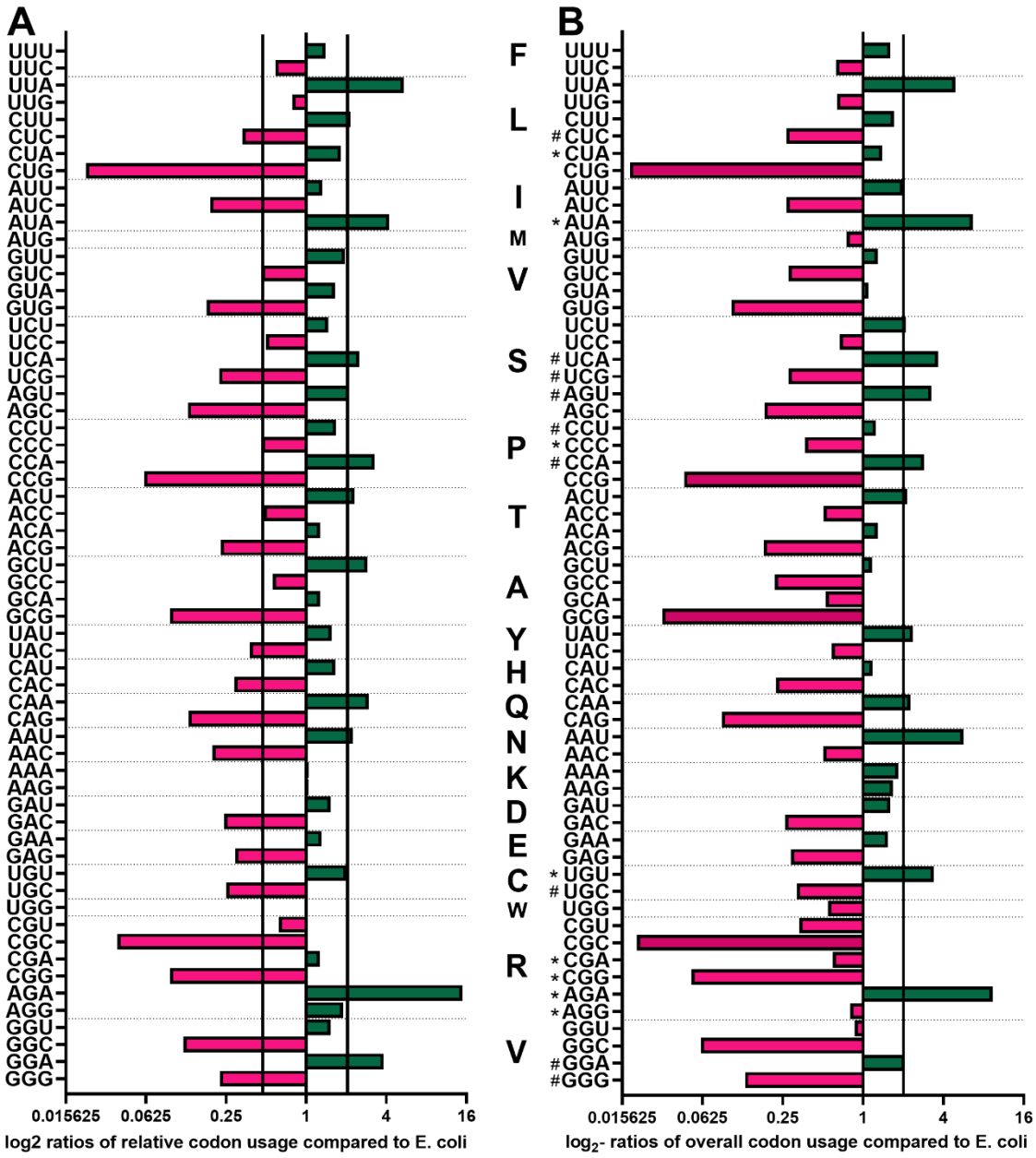


Figure 6-2: Relative and overall codon usage of *P. indianae* compared to *E. coli*. A) Log_2 ratios of *P. indianae*'s codon usage relative to codons for the same amino acids compared to *E. coli*. B) Log_2 ratios of *P. indianae*'s codon usage relative to all codons compared to *E. coli*. Ratios below 1 are shown in pink to indicate codons used more frequently in *E. coli* than *P. indianae*; ratios above 1 are shown in green to indicate codons used more frequently by *P. indianae* than *E. coli*. The solid black line marks the 2-fold increase or decrease in codon use compared to *E. coli*. * and # indicate *E. coli* codons that are rare (≤ 0.5% overall usage) and semi-rare (≤ 1.0% overall usage), respectively.

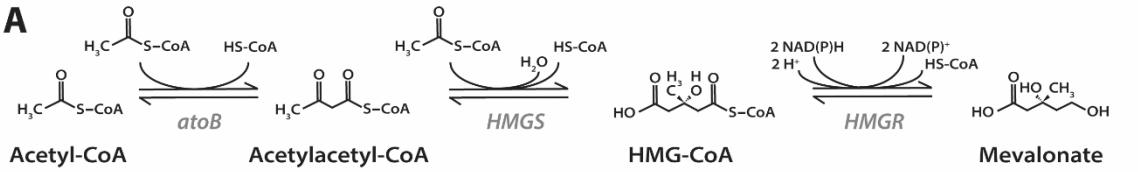
6.3.2 *E.coli* are not well equipped to express the AT-rich genes of *P. indianae*

To test the effects of expressing *P. indianae* genes in *E. coli*, we separately expressed three genes from a biosynthetic pathway from *P. indianae* (Figure 6-3A). We selected the mevalonate pathway (*atoB*, *HMGS*, and *HMGR*) as this is both a conserved pathway in fungi and because it can be used to support a broad range of isoprenoid bioproduction platforms [235]. Because the mevalonate pathway is not native to most bacteria including *E. coli*, characterization of these enzymes in this heterologous platform avoids substrate competition between the native and the heterologous pathways. Expression of these genes was compared to homologs from *S. cerevisiae* and *E. coli*, which have previously been expressed in *E. coli* [236]. Interestingly, the expression of *PI.atoB* and *PI.HMGS* reduced the growth of *E. coli* by about 69% and 53%, respectively (Figure 6-4), while expression of *PI.HMGR* had no effect on growth. These growth defects are unlikely to be the result of enzymatic activity because overexpressing the corresponding homologs from *E. coli* and *S. cerevisiae* (*Ec.atoB, SC.HMGS, SC.HMGR*) does not result in similar growth reduction compared to the uninduced controls. CAI also does not explain this discrepancy as the CAI of the yeast homologs do not reduce growth although they have similar CAI to those from *P. indianae* (Figure 6-3B).

Interestingly, we see that the *HMGR* homologs are nearly identical in both number of overused rare codons and their overall usage, and they both result in similar growth effects (Figure 6-4A). However, *PI.atoB* and *PI.HMGS* each only use two rare codons between 3-5% of the time, yet they have the largest effects on growth (Figure 6-4A) compared to *SC.HMGS*, *SC.HMGR*, and *PI.HMGR* that have three or more rare codons. Based solely on rare codons, it appears that fewer rare-codons results in larger growth deficiencies for these genes. We suspect that additional rare-codons may play a role in slowing down or stalling the translational machinery [229] that in turn slows the drain of the various overused tRNA pools, and thus, limit growth defects. Additionally, codon content was recently shown to correlate with the mRNA levels where rare-codons could reduce protein expression by decreasing the stability of mRNA [234]. The AUA_I codon specifically was shown to have the largest effect at attenuating protein expression and, thus, may play a role here in decreasing the lifespan of mRNA that overuse it, namely *SC.HMGS*, *SC.HMGR*, and *PI.HMGR*. In this case, even though these genes have the potential to overuse rare and semi-rare codons, the reduced mRNA levels keep the tRNA pools from being drained as quickly. In

either case, it would seem that the overexpression of these proteins is affecting the physiology of the cell, which agrees with recent global proteomic analysis when proteins are overexpressed [237].

A



B

Acronym	Origin	Gene	E.C. #	Size	CAI	Overused codons: Rare	Overused codons: Semi-rare
EC.atoB	*E. coli*	*atoB*	*2.3.1.9*	*1.2 kb*	*0.482*	CCC_P (1.02%)	AGU_S (1.78%)
SC.HMGS	*S. cerevisiae*	*HMGS*	*2.3.3.10*	*1.4 kb*	*0.452*	AUA_R (4.42%) CCC_P (1.99%)	AGG_R (5.96%) CGG_R (2.87%) UUA_L (3.53%)
SC.tHMGR	*S. cerevisiae*	*tHMGR*	*1.1.1.34*	*1.5 kb*	*0.360*	AGA_R (2.15%) UGU_C (0.98%) AUA_I (1.59%)	CCU_P (1.99%) CCA_P (1.59%) AGU_S (1.79%) UUA_L (4.38%) UCA_S (1.59%) UCC_S (1.99%)
PI.atoB	*P. indianae*	*atoB*	*2.3.1.9*	*1.2 kb*	*0.469*	AGA_R (1.67%) UGU_C (1.67%)	UUA_L (4.06%) CCA_P (4.06%) CUU_L (2.86%)
PI.HMGS	*P. indianae*	*HMGS*	*2.3.3.10*	*1.3 kb*	*0.465*	AGA_R (2.80%) UGU_C (1.94%)	UUA_L (5.38%) CCA_P (3.44%)
PI.HMGR	*P. indianae*	*HMGR*	*1.1.1.88*	*3.1 kb*	*0.370*	AGA_R (2.15%) UGU_C (0.98%) AUA_I (1.17%)	UUA_L (5.47%) AGU_S (1.98%) CCA_P (4.39%) CUU_L (2.15%) UCA_S (3.61%)

Figure 6-3: Mevalonate pathway and gene information. (A) Mevalonate biosynthesis pathway showing genes, cofactors, and substrate. (B) Homolog information for the genes evaluated in this study. Overused codons are used 2-fold or more in these genes compared to the host utilization. Rare and semi-rare codons are used 0.5% and 1.0% of the time or less, respectively. Codon-usage of each gene is shown next to the codon. Gene names: acetyl-CoA acetyltransferase (*atoB*), HMG-CoA synthase (*HMGS*), and HMG-CoA reductase (*HMGR*)

This translational burden can be alleviated somewhat by reducing promoter strength from a stronger T7 promoter to a weaker H9 variant [238] (Figure 6-4B). However, it should be noted that regardless of the promoter, expression was not detectable via SDS-PAGE suggesting that both strong and weak promoters resulted in weak or no expression (data not shown). Additionally, growth defects have been observed when cellulases from anaerobic fungi were expressed in *S. cerevisiae* [20] which may be the result of similar codon deficiencies. Regardless, CAI alone is insufficient to predict heterologous growth defects of anaerobic fungal enzymes due to their significant biases for rare AT-rich codons that *E. coli* cannot accommodate.

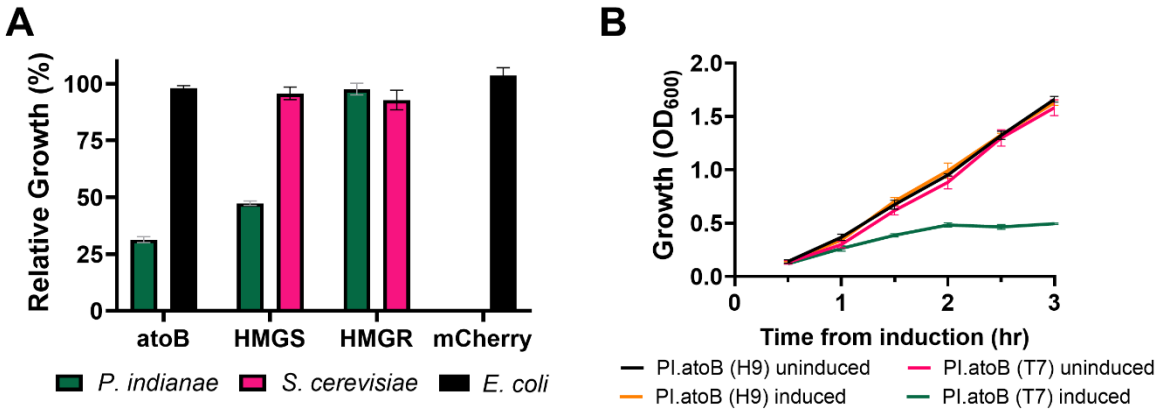


Figure 6-4: Relative growth of *E. coli* expressing the *E. coli, S. cerevisiae,* or *P. indianae* homologs. (A) Relative growth of *E. coli* BL21 when expressing individual genes of the mevalonate pathway under an inducible T7 promoter. (B) Growth curve of *E.coli* BL21 expressing the *PI.atoB* under the T7 promoter and a weaker H9 promoter variant. Growth is relativized to the growth of an uninduced control for each homolog

6.3.3 Strains with addition tRNAs for rare codons do not effectively relieve the burden of expressing *P. indianae* genes

Strains of *E. coli* have been created to allow expression of proteins with codon mismatches by adding additional copies of rare tRNA genes. We evaluated if one of these strains, BL21-CodonPlus (DE3) RIPL (Agilent Technologies), that provides an additional tRNA copy for AGA_R, AUA_I, CCC_P, and CUA_L, could alleviate the growth defects of *P. indianae* homolog expression. Like other prokaryotes., the tRNA gene copy number and codon usage are highly correlated in the *E. coli* genome, ultimately allowing them to use codons with more abundant tRNAs at higher rates [229], [239]. Therefore, increasing the available tRNA pools with additional copies of rare codons tRNA genes may allow increased expression of these genes where it was previously hindered. In this strain, the effect on *PI.HMGS* growth improved markedly where growth was only reduced by about 16%; however, the *PI.atoB* homolog still reduced the growth about by 57% (Figure 6-5A). Looking at the specific codons supplemented by this strain, only AGA_R is highly used in these *P. indianae* homologs and makes up 1.7% and 2.8% of the overall codon usage for *atoB* and *HMGS*, respectively (Figure 6-3B). Because the other RIPL codons that are supplemented (AUA_I, CCC_P, and CUA_L) do not alleviate the drain on the other overused codons such as AAU_N, AAA_K, and

UUA$_L$, this strain is ill-equipped to efficiently express the *P. indianae* homologs and growth is still hindered.

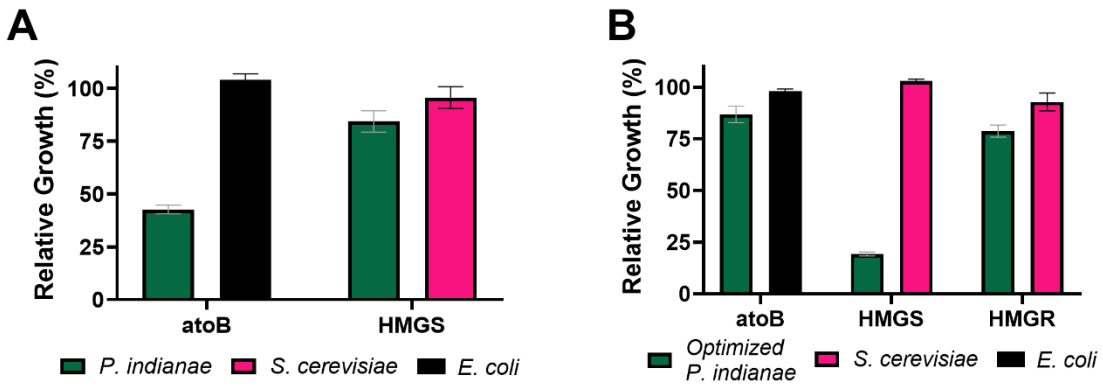


Figure 6-5: Relative growth in additional strains or with optimized genes. (A) Relative growth of E.coli BL21$^+$ RIPL expressing the *E. coli*, *S. cerevisiae*, or *P. indianae* homologs of the mevalonate pathway in. (B) Relative growth of *E. coli* BL21 expressing the *E. coli*-codon-optimized versions of the *P. indianae* genes compared to the respective *atoB*, *HMGS*, or *HMGR* homolog.

6.3.4 Codon optimization alleviates the growth deficiencies seen when the *P. indianae atoB* is expressed in *E. coli*

Because accommodating all of the codon deficiencies would require a large plasmid or an extensive strain engineering undertaking, we selected a codon-harmonization approach to resolve the majority of the underlying codon issues. Codon harmonization matches the codon usage of each gene of interest to that of the *E. coli* genome to avoid potential stresses related to overusing any one codon [240]. When the codon-optimized *P. indianae* homologs were expressed, we saw that the *PI.atoB* and *PI.HMGR* homologs had minor growth defects, growing to roughly 80% of the control level (Figure 6-5B). This suggests that by reducing the usage of suboptimal codons, *P. indianae* genes can be expressed without the same detrimental growth defects seen from the native unoptimized genes. The growth defect of *PI.HMGS*, however, was nearly twice as severe compared to the unoptimized gene. We suspected that at least part of the growth defect seen here is a result of protein activity. *PI.HMGS* in combination with the native *E. coli atoB* produce a toxic intermediate, hydroxy β-methylglutaryl-CoA (HMG-CoA), from acetyl-CoA pools [235], [241]. The unoptimized *PI.HMGS*, however, does not produce this toxic effect possibly due to lower expression. As seen by another group expressing optimized genes from anaerobic fungi in *E.coli*

[21], the growth deficiency could also potentially result from improper folding or membrane incorporation. However, HMM predictors of transmembrane domains [242] did not reveal the presence of any membrane-associated domains within *PI.HMGS*. When this gene was expressed as part of the complete pathway, we did not observe this same growth defect (see growth data, Appendix E: Figure E.2) suggesting that the toxic HMG-CoA was being converted into mevalonate (see section 6.3.6).

6.3.5 Expression of unoptimized genes hinders biosynthesis from *P. indianae* genes

Despite the low CAI scores of the yeast homologs, expressing them in *E. coli* can produce mevalonate and subsequently terpenoids [236]. Initially, we evaluated the full mevalonate pathways (Figure 6-3A) of yeast genes, native, unoptimized *P. indianae* genes, and a hybrid of the two (Figure 6-6A) in order to understand if the native genes could produce mevalonate despite also being poor matches for the *E. coli* chassis. To create the pathways, we took the individually cloned genes and combined them into one plasmid using the ePathBrick system [238]. In contrast to previous approaches that use an operonic approach to express all genes under one T7 promoter [236], [243], we expressed each gene at independent levels via separate T7-inducible promoters of varying strength [244]. This also enabled the investigation of how changing the expression level of each enzyme with different promoter strengths affected the mevalonate output (see Figure 6-6A). Using T7 promoters to drive expression of the yeast pathway, we found that *E. coli* produced similar titers of mevalonate (~0.3 g/L) after 20 hours as previously described despite not using the same operonic approach [236]. However, this titer was improved to ~1.0 g/L (approximately 3-fold) by increasing expression of the *HMGS* with a stronger promoter (Figure 6-6B). This high-producing variant was constructed as part of a combinatorial library of T7-inducible promoters of varying strengths that independently drive expression of each gene [245]. After screening >40 colonies for mevalonate production, we repeatedly found this $atoB_{H9}$-$HMGS_{C4}$-$HMGR_{H9}$ construct as one of the highest producers, showing that optimizing the expression of individual genes in this pathway can improve the mevalonate production (Appendix E: Figure E.1). In contrast to the yeast pathway, the *P. indianae* pathway did not produce significant amounts of mevalonate (< 0.10 g/L) under either the T7 or varied promoter strength configuration tested for the yeast constructs (Figure 6-6A). While the low yield of *P. indianae* homologs is not surprising given the associated growth deficiencies of individual genes, the growth when expressed together in the full pathway was

similar to the fully optimized pathway (see Appendix E: Figure E.2). As *PI.HMGR* expression was not associated with any growth defect in *E. coli* we swapped this homolog for the *Sc.HMGR* variant in the all yeast pathway and evaluated whether a hybrid of the yeast and *P. indianae* pathway (*EC.atoB-SC.HMGS-PI.HMGR*) could produce any mevalonate. Interestingly, we found that this hybrid pathway produced mevalonate titers similar to the all yeast pathway under the T7 promoters (Figure 6-6B). Though the titers are not as high as the H9 yeast pathway that it was derived from, the production of even 0.25 g/L mevalonate was encouraging given that this gene was not optimized and came directly from *P. indianae*. Therefore, we elected to proceed with the optimized pathways as a case study for how biosynthetic pathways from anaerobic fungi can be evaluated in *E. coli*.

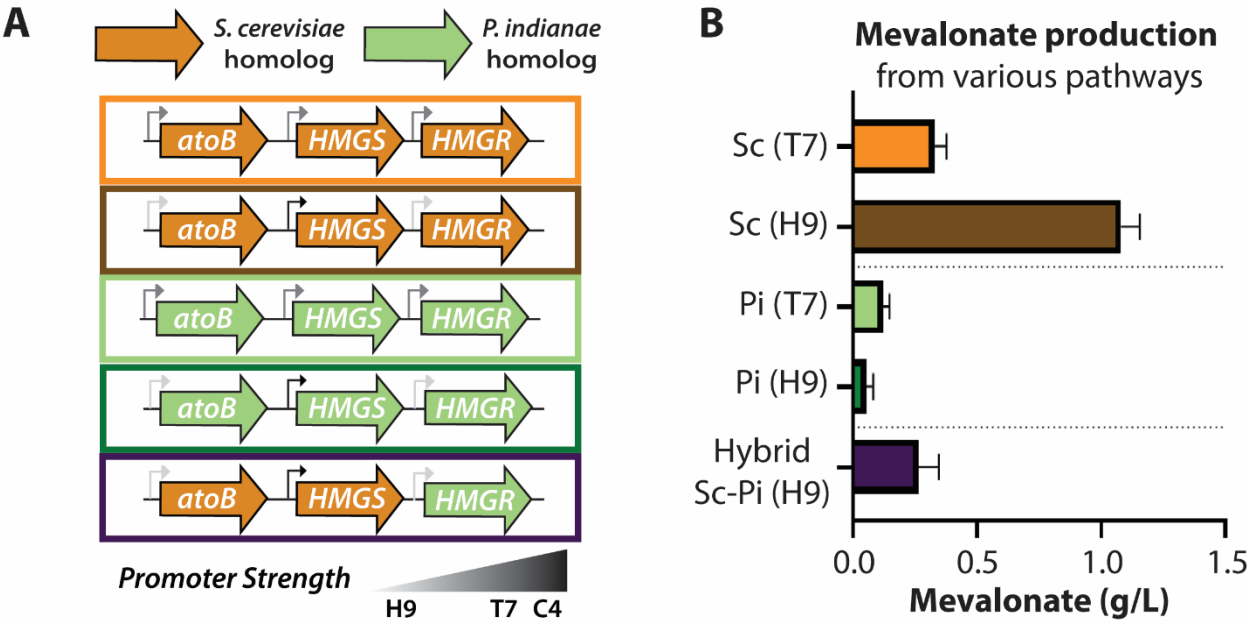


Figure 6-6: Mevalonate pathway variants and titers with unoptimized genes. (A) Different promoter and gene configurations of the Martin et al. (orange) and *P. indianae* (green) mevalonate pathways and a hybrid of them. (B) Mevalonate titers for the pathway variants; Sc = yeast, Pi = *P. indianae*; T7 = T7 promoters for each, H9 = H9-C4-H9 promoter configuration as shown in panel A. Error bars represent standard deviation.

6.3.6 Codon optimization allows heterologous production of mevalonate from *P. indianae* enzymes

Because optimization of the *P. indianae* genes relieved the growth deficiencies of the *atoB* gene, we evaluated if it also enabled increased production of mevalonate. Specifically, we investigated the production of the *P. indianae* pathway with both native and optimized genes, the yeast pathway, and hybrids thereof. In addition to mevalonate production, we also tracked the produced acetate to compare how these constructs affected the flux of the acetyl-CoA precursor. Hybrid optimized pathways were able to increase titers almost 2-fold suggesting that codon-optimized *P. indianae* homologs of the mevalonate pathway may be more catalytically active than yeast variants (Figure 6-7B). In particular, *PI.atoB* improved mevalonate production by shunting acetyl-CoA flux to mevalonate production, reducing acetate titers 88.4%. It also suggests that this first step may be the most important for producing high titers of mevalonate – at least under this promoter organization. Both *PI.HMGR* and *PI.HMGS* in hybrid pathways resulted in similar titers of mevalonate. However, both exhibited reduced acetate levels suggesting an accumulation of redirected carbon in the pathway. Unoptimized hybrid pathways, on the other hand, produced reduced levels of mevalonate similar to what we observed previously (Figure 6-6). The unoptimized native *PI.atoB* and *PI.HMGS* produced no mevalonate (data not shown) while the unoptimized *PI.HMGR* significantly reduced mevalonate production and led to higher accumulation of acetate (Figure 6-7). These results further demonstrate that genes from anaerobic fungi need to be optimized for both expression and mevalonate production in *E. coli*.

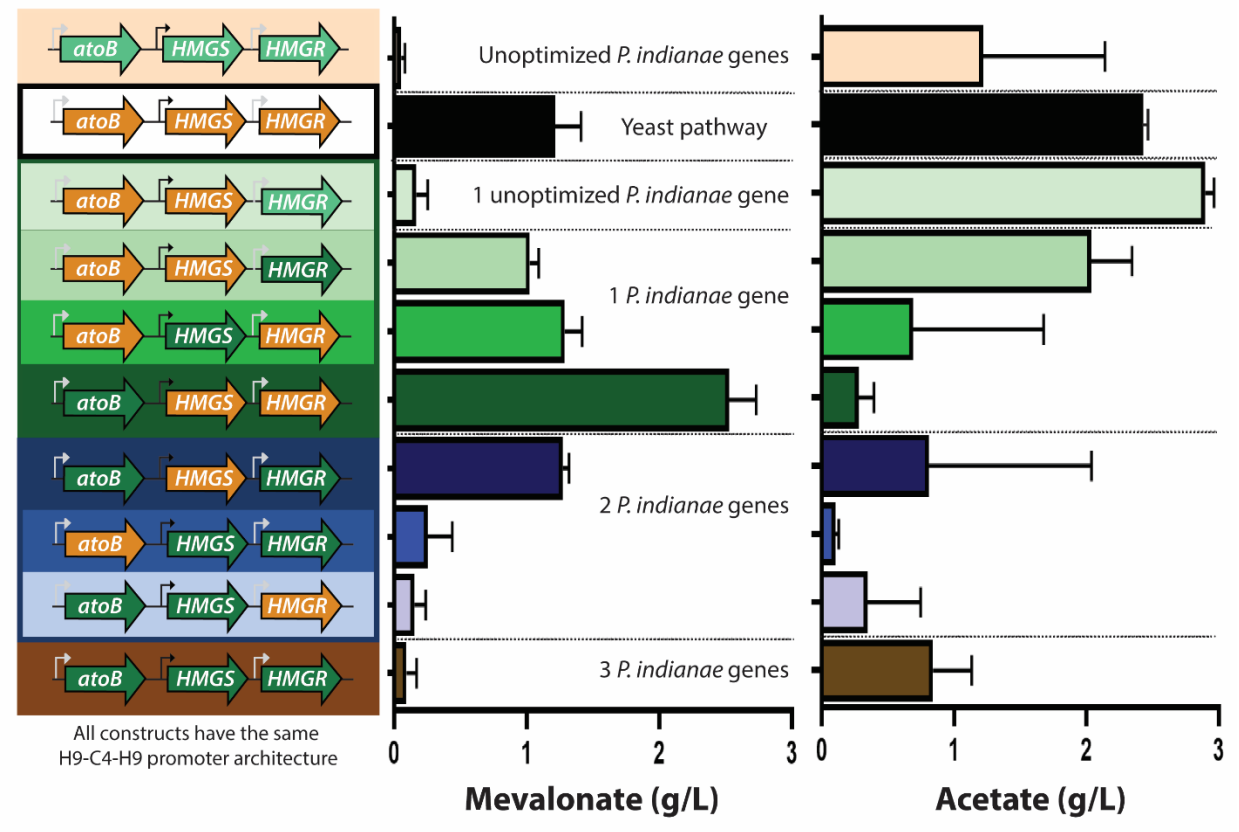


Figure 6-7: Mevalonate and acetate titers of original and hybrid mevalonate pathways. Pathway configuration (left), mevalonate production (middle), and acetate accumulation (right) from various mevalonate pathway hybrids containing individual genes from Martin et al (orange), native *P. indianae* genes (light green), or *E. coli*-codon-optimized *P. indianae* genes (dark green) after 20 hrs of culture. Green bars represent pathways with one PI homolog, blue bars and brown bars represent pathways with two and three PI homologs, respectively. All pathways are configured in the high-producing h9-promoter configuration. Errors bars represent standard deviation.

Most constructs with two *P. indianae* genes saw a decrease in the amount of both acetate and mevalonate when compared to the parent one-*P. indianae*-gene hybrid. The mevalonate produced by combining the first and last gene homologs (*PI.atoB* and *PI.HMGR*) with the *SC.HMGS* reached a level between the levels produced by either one-gene hybrid. Based on the lower acetate level of this two-gene hybrid, we suspect that *SC.HMGS* was hindering the flux of acetyl-CoA compared to other two two-gene hybrids containing *PI.HMGS*. The *PI.atoB-PI.HMGS-SC.HMGR* pathway made very little mevalonate (less than 0.25 g/L) considering that these one-gene hybrids individually both produced more than 1 g/L. However, when we expressed this pathway on a smaller scale (5 ml, Appendix E: Figure E.3), they produced about 1.6 g/L,

which is between the yields of the single-gene hybrids. While the reason for lower mevalonate titers at this larger scale (50ml) is unclear, it suggests there may be specific culturing conditions that might improve the performance of these genes. In addition to exploring other culturing conditions, we also expect different promoter organizations may improve the production of these genes given that this organization was optimal for the yeast genes [244]. This very well may be the case, because we observed that when all three anaerobic fungal genes are combined into one plasmid under the previous promoter organization, very little mevalonate was actually produced (Figure 6-7).

Although many parameters remain to be optimized, we demonstrate here the functionality and utility of these *P. indianae* genes. Compared to the previous batch titers of Ma et al., 2003 (~0.4 g/L) and Yang et al., 2012 (3.1 g/L), our maximum titer of 2.5 g/L titers is a promising starting point for future studies that focus on optimizing promoter organization and culture condition. Moreover, *P. indianae* sourced genes may proceed via distinct anaerobic catalytic mechanisms. In contrast to *SC.HMGR*, the annotation of *PI.HMGR* predicts that it uses NADH rather than NADPH as a cofactor in the synthesis of mevalonate (Figure 6-3A). While unvalidated, this putative cofactor preference may be beneficial for production in certain expression hosts and conditions.

6.4 Conclusions

Anaerobic fungi are an emerging platform for biotechnology with novel biosynthetic pathways. Yet, heterologous expression of their biosynthetic pathways has had limited success in model hosts like *E. coli*. We find that one reason may be the genome composition of anaerobic fungi, like *P. indianae,* are extremely AT-biased with a preference for particularly rare AT-rich tRNAs in *E coli*, which are not explicitly predicted by standard CAI metrics. Native *P. indianae* genes with these extreme biases create significant growth defects in the heterologous host. However, codon optimization rescues growth allowing for gene evaluation. In this manner, we demonstrate that anaerobic fungal homologs such as *PI.atoB* are more active than *S. cerevisiae* homologs increasing titers 2-fold and reducing waste carbon to acetate by almost 90% under the conditions tested. Our work highlights the need for in-depth codon utilization analysis in determining heterologous hosts for expression and demonstrates the potential for anaerobic fungal enzyme homologs in metabolic engineering.

6.5 Materials and Methods

6.5.1 Homolog identification, Primer design, PCR, RT-PCR, and Cloning

Using the Search terms *"acetoacetyl-CoA transferase", "hydroxymethylglutryl-CoA synthetase",* and *"hydroxymethylglutryl-CoA reductase,"*the genomic databse of *Neocallimastgomycota* on Mycocosm [247] was searched. The coding regions of the resulting protein sequences were analyzed with BLAST [248] to confirm homologs. The coding sequence of the homologs were obtained from Mycocoms and aligned using MEGA7 [156]. Degenerate primers were designed to amplify the gene sequences from *Piromyces sp. UH3-1* [14]. PCR amplicons were sequenced to confirm target identity of the *atoB*, *HMGS*, or *HMGR* homolog. To isolate genes without introns for cloning, we then amplified the gene from a cDNA library. Total RNA was extracted from *Piromyces sp. UH3-1* using the QIAGEN (Germantown, MD) AllPrep Fungal DNA/RNA/Protein kit. Reverse transcription PCR was performed using the QIAGEN QuantiTect Reverse Transcription Kit along with random octamers according to the manufacturer's instructions. From the resulting cDNA pool, the target genes were PCR amplified with primers flanked by a 5' BglII and a 3' XhoI site. Finally, these amplicons were digested with the flanking enzymes and ligated with Rapid T4 ligase (ThermoFisher; Waltham, MA) into the pETM6 backbone [238] digested with the same enzymes and dephosphorylated with calf intestinal alkaline phosphatase (CIAP). Following the ePathBrick [238] design, the full pathway was constructed in on backbone by first joining the *atoB* and *HMGS* homolog in one round of cloning, and joining the *HMGR* homolog with the first two genes in a final step. The analogous yeast pathway [236] was obtained from Addgene as a benchmark. In a manner similar to the original construction of the pathway, combinations of the yeast and *Piromyces* homologs were made to assess the productivity of the various homologs. Similarly, vectors with weaker or stronger T7 promoter derivates were generated and combined in attempt to enhance mevalonate production. Because a yeast homolog vector organized with the $atoB_{H9}$_$HMGS_{C4}$_$HMGR_{H9}$ pathway was found to be a high-producing stain, the *Piromyces* homolog pathways were constructed to match this promoter structure.

6.5.2 Growth Analysis

Vectors containing the homologs for the *PI.atoB*, *PI.HMGS*, or *PI.HMGR* gene were each transformed into an electrocompetent BL21 (DE3) *E. coli* strain for expression from the T7 promoter. Full and intermediate pathway vectors were transformed similarly. Overnight starter cultures were inoculated from single colonies and used to inoculate fresh LB/ampicillin 5 ml cultures to OD 0.05 the following day. The initial O.D. was adjusted to approximately 0.05 and IPTG was added to 100 µM at the time of inoculation. Cultures were grown at 37 ˚C while shaking at 250 rpm in parallel with uninduced cultures for approximately three hours with the O.D. being measured every 30 minutes. *E.coli* BL21 RIPL CodonPlus (DE3) strains were also transformed and grown in the same manner as described above with the addition of chloramphenicol in accordance with the CmR marker on the RIPL vector.

6.5.3 Codon Adaptation Index, Codon Usage, and Codon optimization

The protein encoding sequences of the *E. coli* and *P. indianae* genomes were downloaded from GenBank (Accession #: U00096.3) and Mycocosm (piromy1), respectively. Using a python package developed by Lee et al (2018), the CAI of the *E. coli* genome and *P. indianae* genomes were calculated from the coding sequences; the CAI histograms and the Gaussian distribution fit were calculated with PRISM v9.0. To calculate the codon usages, a BioPython [249] script was used to split the coding sequences every three base pairs to create trinucleotide codons for every gene. The codon occurrences were tallied and then frequencies were calculated with respect to all trinucleotides ("overall usage") and those that code for the same amino acid ("relative usage"). Trinucleotide frequencies, or codon usages, were compared for relative and overall usage between *E. coli* and *P. indianae* (Figure 6-2). Additionally, the codon usage of *E. coli* was used to determine the codon frequencies that should be used for optimizing *P. indianae* genes. These gene sequences were manually optimized by removing any rare codons, and replacing both rare and infrequently used codons with frequently used codons until the overall codon frequency of each gene roughly matched the *E. coli* relative codon usage. Genes were then synthesized by Twist Biosciences and cloned into the pETM6 vector as previously described (see Section 6.5.1).

6.5.4 Mevalonate production cultures and HPLC analysis

Overnight cultures of BL21 (DE3) strains containing variants of the yeast, *Piromyces,* or hybrid mevalonate pathways were grown in LB/amp at 37˚C. For mevalonate production, cultures were inoculated to an O.D. of 0.05 in 2 YT media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 1% glycerol; pH ~7.5). 5 ml 2YT cultures were grown at 37˚C while shaking at 250 rpm. When cultures reached an O.D. of 1, they were induced to a final concentration of 500 µM IPTG and left overnight (Total ~18 hrs). Final O.D. was measured, and 450 µl of cleared supernatant was mixed with 50 µl of 14% H_2SO_4 in order to convert produced mevalonate to mevalolactone for HPLC analysis [250]. The acidified supernatant was immediately chilled at 4˚C for a minimum of 1 hour and then analyzed by HPLC. Briefly, analysis of a 20 ul sample of the resulting supernatant was injected on an Agilent 1260 instrument with separation via a BioRad Aminex HPX-87H at 50 ˚C in 5 mM H_2SO_4 mobile phase (rate 0.600 ml/min). Detection was performed with a Refractive Index Detector (RID) in (+) signal polarity mode at 45 ˚C. To evaluate mevalonate production at larger scales and better aeration, 50 ml 2YT cultures were also grown under the same parameters as above but in baffled flasks.

7. DEVELOPING GENETIC TOOLS TO UNLOCK THE POTENTIAL OF ANAEROBIC FUNGI: PROGRESS AND TOOLS IN DEVELOPMENT

7.1 Abstract

Anaerobic fungi are potentially powerful platforms for biotechnology that remain unexploited due to a lack of genetic tools. These fungi have the largest repertoire of CAZymes for degradation of lignocellulose, making them attractive for biofuel production, and are predicted to possess novel biosynthetic capabilities. However, a genetic toolbox that further enables study and exploitation of these capabilities for synthetic biology remains underdeveloped. No plasmids, promoters, terminators, selection markers, reporter genes, or gene-editing systems have been characterized. Leveraging a naturally competent phase of the lifecycle, DNA and RNA can be introduced and used for transient heterologous gene expression. Existing transcriptomic resources were leveraged to identify putative promoters and terminators that regulate gene expression at different that appear conserved across 6 anaerobic fungal isolates. Using our transformation protocols, at least one promoter candidate was confirmed to promote gene expression of a flavin-based fluorescent reporter and developed a pipeline to characterize candidate gene regulatory elements. Additional experiments also successfully introduced antibiotic selection systems that will be extended to express alternative selection markers, fluorescent reporter genes, and CRISPR-based endonucleases. Important elements such as origins of replication and centromeric binding sequences are being developed in parallel to create a stable, plasmid-based expression system in anaerobic fungi. Ultimately, this toolbox enables the expression of genes from non-native pathways, of selection markers, and of engineering systems that will unlock a variety of genome engineering possibilities for strain development and various synthetic biology applications.

7.2 Introduction

Anaerobic fungi possess the highest diversity of biomass-degrading enzymes (CAZymes) across the fungal kingdom (>300 CAZymes), exceeding that of industrially used *Trichoderma* strains and lignin-degrading white rot fungi [15]. In the native large herbivore gut these lignocellulose-degrading enzymes include cellulases, hemicellulases, and pectin-degrading or lignin-modifying enzymes that are responsible for roughly 60% of the plant biomass degradation

and, therefore, may help overcome a significant bottleneck of bioenergy production from agricultural crop waste [15], [93]. Interestingly, their CAZyme activity is resistant to lignin-composition [14], and surveys of their genomes of anaerobic fungi show that anaerobic fungi possess significant untapped biosynthetic capabilities [18], [251]. Because of this significant genomic potential, anaerobic fungi may be able to help meet the growing need for bioenergy and bioproduction strategies using sustainable materials. Thus, these fungi are an emerging platform for robust conversion of woody biomass for downstream processes and biofuel production.

The potential of anaerobic fungi, however, remains unrealized because we lack the tools necessary to manipulate their gene expression or edit their genomes [196], [252]. Because regulation of CAZyme expression is controlled through catabolite repression in anaerobic fungi [18], [251], engineering robust biocatalysis strategies will require reprogramming the regulation of these CAZymes through promoter swaps and heterologous expression. Additionally, metabolic engineering strategies to enhance production of the natural products will require the ability to knock-out peripheral pathways and overexpress those of interest though CRISPR-Cas9 mediated approaches. To identify genes that will advance biofuel and bioproducts production, we need to create methods that manipulate enzyme expression and rapidly interrogate gene function in the natural context of the anaerobic fungi.

Genetic tools for anaerobic fungi, however, are vastly underdeveloped; only one promoter has been evaluated for heterologous expression [166]. While the lack of genomes previously hindered our ability to identify and utilize their genetic parts, our recently acquired genomes (Chapter 4) enable the development of parts for expression vectors, selection markers, and gene editing tools. Specifically, by taking genomic sequences immediately preceding highly expressed genes and fusing them to quantitative reporters, a range of promoters can be identified in anaerobic fungi as has been demonstrated in other fungi [182], [209]. Similarly, using genomic libraries and some highly conserved genetic loci, autonomously replicating sequences (ARS) and centromere binding sequences (CEN) can be identified in order to help stabilize expression vectors [253], [254]. Ultimately, these fundamental expression vector components enable the development of selection markers and gene-editing systems that will enable evaluation of target gene function [255] and create more efficient strains for lignocellulose degradation and biofuel production.

To unlock the potential of anaerobic fungi, this work focuses on bottom-up engineering of basic stable expression plasmids, characterizing genetic reporters, and developing a gene editing

system in anaerobic fungi. In addition to developing electroporation methods for anaerobic fungi, their naturally competent zoospores [256] were leveraged to heterologously express reporter genes and selection markers. Using genomic information, plasmids containing the enolase promoter (P_{ENOL}) from *Neocallimastix sp. GfMa3-1* were constructed to drive expression of two fluorescent reports and subsequently characterize additional parts for our genetic toolbox. Similarly, an antibiotic selection system for episomal plasmids was constructed and used to identify components for stable plasmid replication. Finally, a selection system that can identify genetic mutants was evaluated and constructs were built for genome modification through homologous recombination and CRISPR-Cas9 gene disruptions. Ultimately, the basic parts developed here will be the foundation of a genetic toolbox for aerobic fungi that will be expanded upon and used to create more efficient strains for lignocellulose degradation and biofuel production.

7.3 Results and Discussion

7.3.1 Transformation of anaerobic fungi

Evaluating the natural competence and electro-competence of anaerobic fungi.

Previously, two transformation methods have been evaluated for anaerobic fungi. The first, biolistic transformation was evaluated in a *Neocallimastix* isolate and using a putative enolase promoter, they found that this method could incur transient transformation up to seven days [166]. Biolistic transformations, however, require specialized equipment and gold particles that hinder the accessibility of this technology. The second, natural competence, leverages the juvenile zoospore's ability to naturally take up nucleic acids, namely RNA [256]. This method, however, was not used for heterologous expression but for RNAi to disrupt ethanol formation in *Pecoramyces* [167]. Here, natural competence was evaluated in *Piromyces* and *Neocallimastix* isolates for heterologous gene expression as well as electroporation as an alternative.

Similar to Calkins et al, small labeled oligonucleotides were used to visualize the uptake of nucleic acids. Random nucleic acid sequences were designed to be ~50% GC and cross referenced to not have homology to the publicly available genomes. For natural competency, cultures were grown in SWIM media [256] that allows liberated zoospores to freely move while the mature zoosporangia remain localized. The zoospores can be collected and separated from the mature fungi with a wash with sterile anaerobic water that enters the porous substrate collecting

them. Nucleic acids can then be added and allowed to incubate with the zoospores to be taken up naturally. When this was done this with *Piromyces sp. UH3-1*, RNA was taken up (Figure 7-1A) similar to what was previously shown with *Pecoramyces* [256].

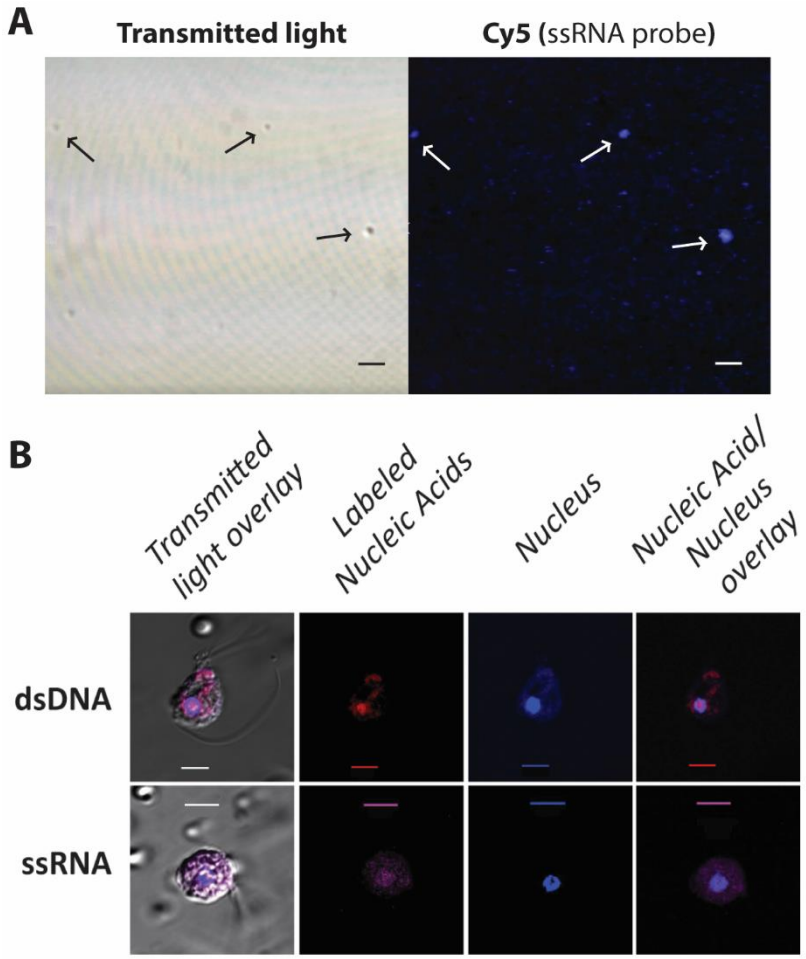


Figure 7-1: Transformation of anaerobic fungi. A) Natural competency transformation of *Piromyces sp. UH3-1* with labeled ssRNA probe; scale bar = 20 µm, arrows indicate competent zoospores. B) Electrical transformation of *Neocallimastix sp. GfMa3-1* showing colocalization of DNA (red)and RNA (purple)probes with nuclear content (blue); scale bar = 5 µm. dsDNA = double stranded DNA, ssRNA = single stranded RNA, nucleus stained with DAPI)

Electroporation is another commonly used transformation method for model microbes such as *E. coli*. Here, it was evaluated as an alternative, and potentially more efficient, method than natural competence. Electroporations, however, have several parameters that may need to be tuned for various organisms to avoid excessive currents that can damage or kill the cells. Like our natural competence experiments, zoospores were harvested from SWIM media in sterile anaerobic water;

however, they were transferred to a cuvette and mixed with DNA before being shocked for uptake. Both DNA and RNA were shown to transform into *Neocallimastix* sp. *GfMa3-1* using electroporation (Figure 7-1B) though more work is required to optimize this method of transformation. Because electroporation was more variable and has the potential to lead to cell death from high electrical currents, the heterologous expression experiments were carried out with the natural competency transformation method.

7.3.2 Identification and validation of the putative enolase promoter in anaerobic fungi

Bioinformatic identification of the enolase promoter

To date, the only identified promoter that has been used for heterologous expression is the enolase promoter from *Neocallimastix frontalis* [166]. Our analysis of this sequence with homologs from other sequenced fungi revealed that there was approximately 500bp of homology (sequence ID >70%, Figure 7-2A) conserved between them sequenced anaerobic fungi. Using these sequences, degenerate primers were developed to amplify about 1000 bp of the intergenic region upstream of the enolase genes of our isolates. Using PCR with low annealing and extension temperatures (~45 ˚C) that accommodate the high AT-content of the anaerobic fungal genomes, the putative enolase promoter region was amplified from the genome of *Neocallimastix sp. GfMa3-1* (Figure 7-2B). This sequence was subsequently cloned into the pETM6 backbone using the XbaI and BglII restriction sites. After sequence verification of the promoter, this backbone was used as a backbone for creating initial reporter and selection plasmids (Figure 7-2C).

Identifying low-, medium-, and high-expressing constitutive promoters

Recently, the genomes of *Piromyces sp. UH3-1*, *Neocallimastix sp. GfMa3-1*, and *Neocallimastix sp. WI3-B*, were sequenced with the Joint Genome Institute (JGI), which now enables the extraction of promoters directly from these isolates. By pairing transcriptomic and genomic data for *Piromyces finnis* and *Neocallimastix californiae*, promoters were identified that were constitutively expressed across a variety of feedstock growth conditions for two phylogenetically diverse species, suggesting they are likely constitutively expressed in other species as well. Similar to previous work in non-model fungi [182], these candidates were categorized into high-, medium-, and low-expression promoters to evaluate a range of promoters.

Interestingly, the enolase promoter was not found to be in the list of constitutively expressed genes and suggests that it may not be the best choice for driving expression of heterologous proteins. However, by selecting core genes that are present across all the available genomes, the focus was narrowed in on highly conserved genes and their promoters. Using these criteria, 10 promoters were selected from each expression level and are being synthesized for characterization with the JGI. Upon synthesis, these will be evaluated for gene expression using one of the reporters developed below (See section 7.3.3). By expanding our toolbox to include a range of promoter strengths, the expression of selection markers, reporters, biosynthetic pathways, and Cas9 editing systems can be finely tuned for genome engineering.

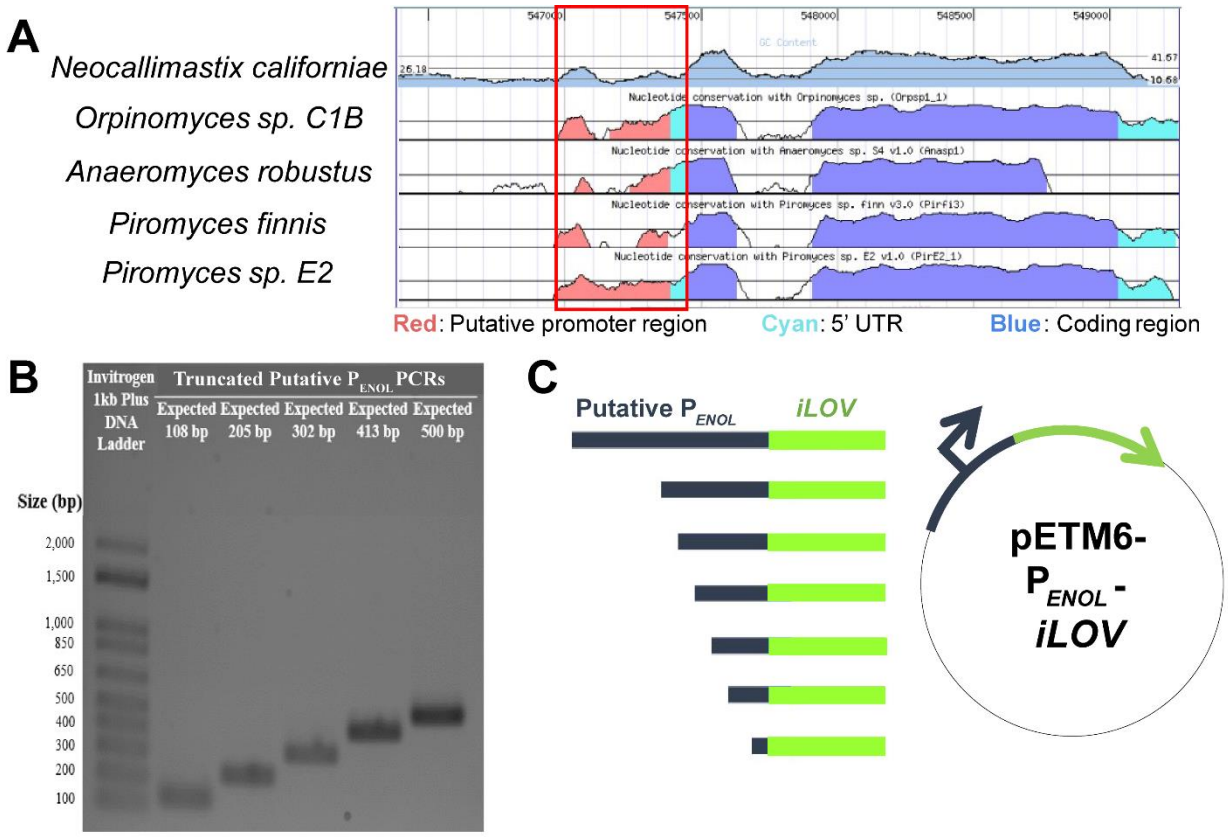


Figure 7-2: Identifying and cloning the putative enolase promoter. A) Alignment of the enolase regulatory (red/cyan) and coding (blue) regions of anaerobic fungal genes available through Mycocosm. B) PCR amplification of 100-500 bp truncations of the AT-rich enolase promoter region. C) Schematic of enolase (P_{ENOL}) plasmid and demonstration of the promoter characterization strategy that will be used to truncate functional promoters down to the essential sequence.

7.3.3 Developing oxygen-independent reporters in anaerobic fungi

Flavin and heme-binding fluorescent proteins

Because anaerobic fungi are strict anaerobes, they cannot use the majority of fluorescent proteins that require oxygen for proper chromophore folding [257], [258]. Therefore, in order to validate our enolase promoter, the iLOV flavin-binding protein gene, which fluoresces a cyan color [258], was used as a reporter in the enolase promoter vector to validate promoter function (Figure 7-2C). Here, naturally competent zoospores were subjected to five separate doses of the reporter vector (2 µg each) over a period of 72 hours and the cells were then imaged with fluorescent microscopy. The results of this dosing experiment show that the zoospores took up the plasmid, demonstrated functionality of the enolase promoter, and suggest that iLOV can be used as a reporter for anaerobic fungi (Figure 7-3A). Additionally, a nuclear localization sequence (NLS) necessary to direct gene-editing enzymes to the nucleus was identified from the N-terminus of histone 2B by comparing these sequences for all available anaerobic fungi. The 90bp NLS-tag was validated by fusing it to the N-terminus of the iLOV protein under the control of the enolase promoter (Figure 7-3A) and suggests that this tag can be fused to endonucleases like Cas9 for genome editing. While iLOV can be used as a reporter, the current transient expression system and the weak fluorescence generate do not generate enough signal above the media background and thus limits our ability to quantify expression with a fluorescent spectrometer. In addition to iLOV, whose excitation wavelength overlaps with the commonly used nuclear stain DAPI, the red-fluorescing heme-binding protein iRFP702 [259] was also evaluated as an alternative reporter. The iRFP702 gene was codon-optimized using the codon table for anaerobic fungi (Appendix Table E.1) and cloned into a cassette with NLS tags flanking it on both the N- and C-terminal ends and under the control of the enolase promoter. Using the same plasmid-dosing and natural competency transformation as for iLOV, iRFP is also able to visualize expression from the enolase promoter in *Neocallimastix sp. GfMa3-1* (Figure 7-3B). Further optimization of heme concentrations in the media may, however, be necessary to improve fluorescence of this reporter. With multiple reporters, more complex genetic circuits can be built and tested for promoter and terminator characterization [182].

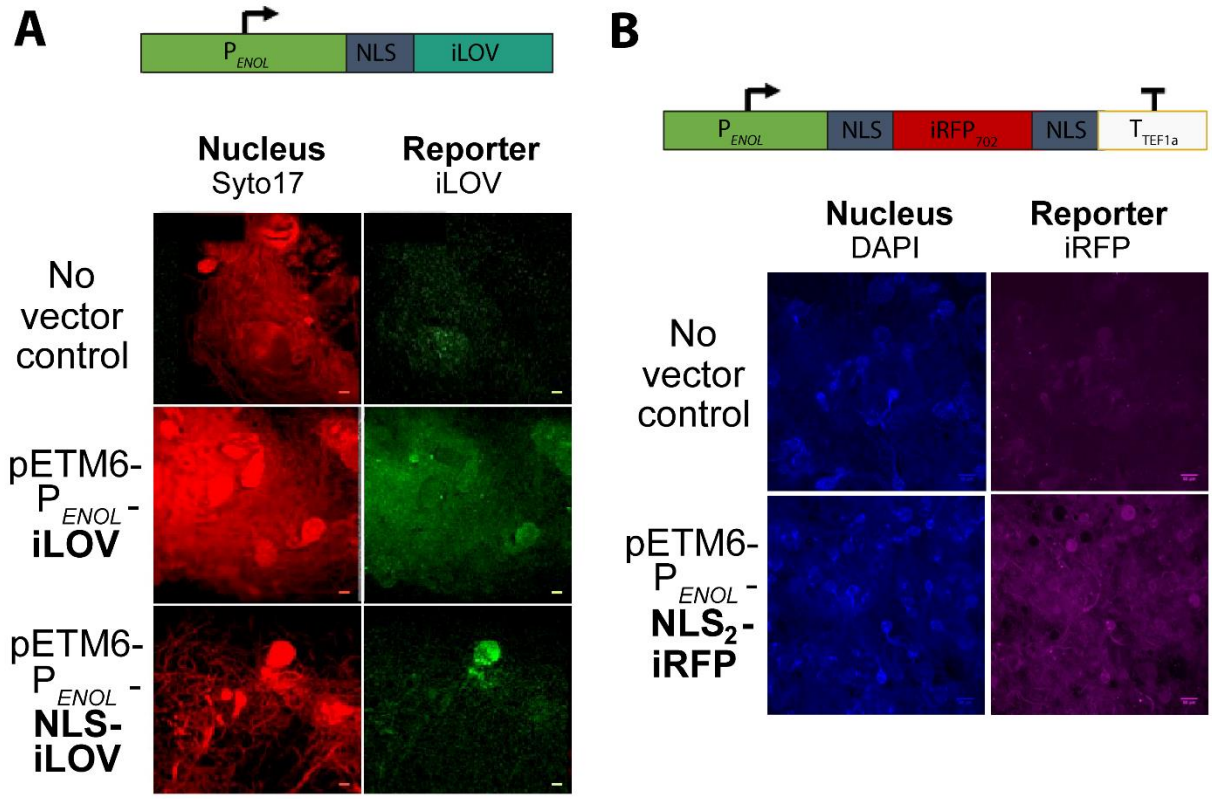


Figure 7-3: Heterologous expression of fluorescents reporters in *Neocallimastix sp. GfMa3-1*. A) Schematic example of an iLOV gene construct and demonstration of both its fluorescence and localization when combined with the nuclear localization sequence (NLS); scale bar -= 20 µm. B) Schematic example of the NLS_2-iRFPgene construct and demonstration of its fluorescence; scale bar -= 50 µm. Nuclear stains = Syto17 (red) and DAPI (blue); P_{ENOL} = enolase promoter, tTEF1a = translation elongation factor terminator.

Plate reader assay for quantifying expression using the heterologous GusA reporter

In addition to fluorescent proteins, tabulating gene activity through quantitative assays can validate heterologous expression. Quantitative assays for heterologous expression rely on orthogonal gene activity or low background activity in untreated conditions. With respect to anaerobic fungi, this means that the media itself should not have high background fluorescence that would otherwise interfere with measurement of gene activity. For example, the iLOV fluorescent excitation and emission wavelengths overlap with some fluorescent components of the media that strongly emit signal, and therefore, hinder quantitative measurement of iLOV expression in batch culture (data not shown). Here, a 96-well plate and fluorescent spectrometer were used to evaluate the use of a codon-optimized *GusA* (Figure 7-4A), a beta-glucuronidase

(GH2 family), reporter to cleave a fluorophore from a carbohydrate-linked substrate (MUG, 4-Methylumbelliferyl-beta-D-galactopyranoside). While most anaerobic fungi do not have more than one GH2 family protein, some background MUG-cleaving activity was detected in *Neocallimastix sp. GfMa3-1* (Figure 7-4B, C). However, when the exogenous *GusA* was expressed with natural competency transformation and plasmid dosing, the signal relative to growth is significantly higher than the background signal from endogenous MUG-cleavage (Figure 7-4B). In addition to the fluorescent reporter expression shown previously (Figure 7-3), the gene activity data here further supports the functionality of the enolase promoter. Because the mycelia interfere with accurate fluorescence measurements and this fluorescence is contained within the cells (Figure 7-4C), this plate assay is performed on lysates of the fungal biomass to accurately measure

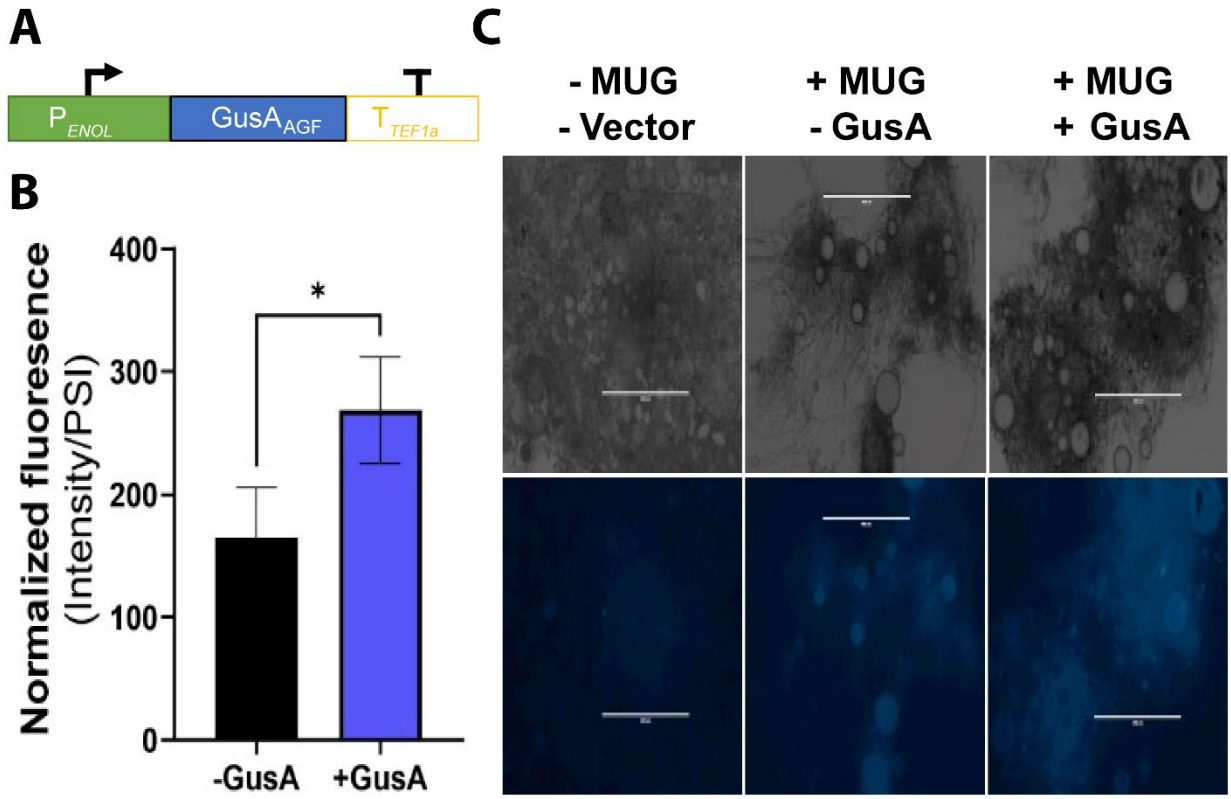


Figure 7-4: Using *GusA* as a semi-quantitative reporter. A) Gene schematic of *GusA* in under the control of the P_{ENOL} and T_{TEF}. B) Normalized fluorescent intensity of *Neocallimastix sp. GfMa3-1* lysates after three days with and without the heterologous *GusA* reporter; * = unpaired t-test $P < 0.05$. C) Micrographs showing the background fluorescence without (left) and without (right) MUG and after the addition of the *GusA* reporter (left). MUG = 4-Methylumbelliferyl-beta-D-galactopyranoside, scale bar = 200 µm.

the cleaved fluorescent compounds. As variability in the washing and lysis of the cells can introduce additional variability, future efforts should evaluate the addition of a secretion tag to allow the activity to be measured in the supernatant to increase throughput and avoid the variable steps of the assay. However, microscopy of the treated and untreated cultures also demonstrates the background activity of the MUG-only treatment was primarily in the sporangia while the fluorescent activity increased throughout the mycelia of the P_{ENOL}-*GusA* transformed culture (Figure 7-4C) suggests that this untagged protein was expressed throughout the entire cell. Recent genomic annotation of the *Neocallimastix sp. GfMa3-1* genome indicated that this isolate contains seven GH2-family CAZymes that may be responsible for the background MUG-cleaving activity (Figure 7-4B, C). While this assay further validated the activity of the enolase promoter here, additional optimization will be needed to standardize this assay for comparing promoter strength. Despite the fact that there is not a one-size-fits all reporter for our genetic toolbox, three separate tools are presented here that can be adapted to answer specific questions in a variety of contexts. These reporters form the foundation of our genetic toolbox that will be built upon and improved as we move toward engineering these fungi.

7.3.4 Creating plasmid-based selection systems for anaerobic fungi

Building an antibiotic selection system with hygromycin

Having validated the enolase promoter (P_{ENOL}), a selection system was constructed where P_{ENOL} drives expression of an antibiotic resistance gene (Figure 7-5A). After demonstrating that anaerobic fungi are susceptible to the antibiotic hygromycin (Appendix F: Figure F.1), the hygromycin resistance gene (*HygR*) was cloned into the pETM6-P_{ENOL} backbone. However, because these plasmid do not have a mechanism to replicate and are diluted out due to growth, additional plasmid was dosed in at 12 to 24 hour increments so that plasmid would be available for recently liberated zoospore to uptake (Figure 7-5B). In order to allow the fungi to mature and express *HygR*, antibiotic selection was not applied until 52 hours (Figure 7-5B). When *HygR* is expressed, it recovered approximately 70% of the growth compared to the positive control while the antibiotic- treated cultures without plasmid failed to generate substantial pressure (Figure 7-5C). When this experiment was repeated, it was found that the timing of the antibiotic treatment can greatly affect the results. Less pressure is generated if the antibiotic is added too early,

suggesting that the selective pressure is too restrictive for juvenile cultures; while more pressure is generated if the antibiotic is added later, suggesting there is not enough selective pressure (data not shown). It is unclear at this time, however, whether hygromycin is the best choice of antibiotic or if there are better options for a selection system. Similarly, other promoters could provide better expression levels that may be better suited for a selection system. While several parameters remain to be optimized, this selection system is one of the most fundamental components of our genetic toolbox and will continue to be improved as more parts are generated.

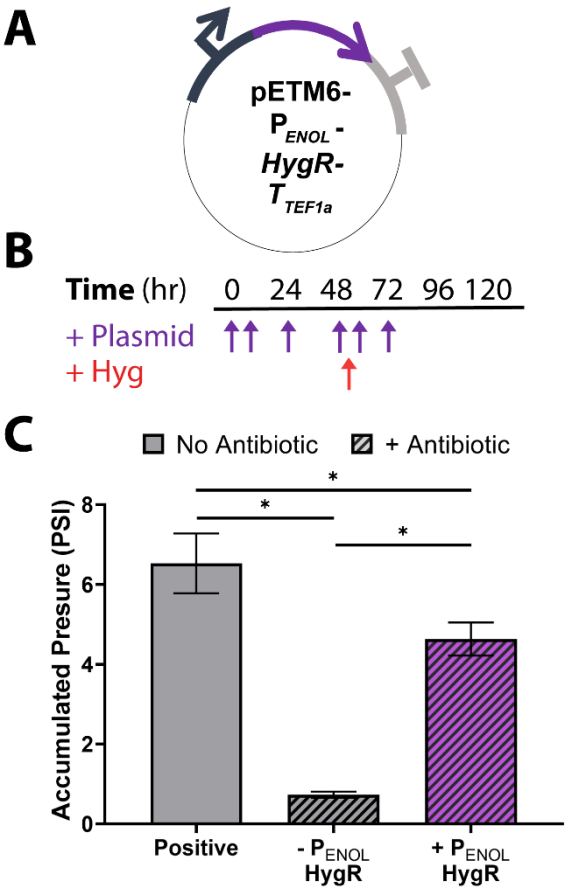


Figure 7-5: Developing an antibiotic selection system for anaerobic fungi with hygromycin. A) schematic of resistance vector. B) Plasmid dosing and antibiotic selection timeline. C) Accumulated pressure of untreated (positive), hygromycin-treated, and vector-containing, hygromycin-treated *Neocallimastix sp. GfMa3-1* cultures * = unpaired t-test $P < 0.05$.

Screening genomic libraries and conserved loci for autonomously replicating sequences

In order to create a plasmid that is stably replicated in the host organism, it needs to contain an autonomously replicating sequence (ARS). However, at this time no such sequences have been

identified in anaerobic fungi, nor are there are bioinformatic tools for identifying or predicting their sequences. Characterization of ARS in other fungi and yeasts suggest that they typically contain AT-rich motifs; though, looking for such motifs in anaerobic fungal genomes would likely yield a lot of false positives as the genome is typically 80% AT. Instead, ARSs have been identified in other non-model yeasts by creating libraries of genome fragments and then screening these random fragments for the ability to resist antibiotic selection through a functioning ARS [253]. Based on *in-silico* restriction enzyme digests of the known genomes, NdeI, EcoRI, and BglII were identified as enzymes that can digest the genome into the highest numbers of genome fragments between 1 and 10 kb for cloning into the resistance vector (Appendix F: Figure F.2). After digesting genomic DNA with these enzymes, the resulting fragments were cloned into the *HygR* resistance backbone (Figure 7-6a, top). Simultaneously, fragments of the rRNA operon intergenic spacer (Figure 7-6A, bottom), a location that has been shown in several fungi to have an functioning ARS [260], were also cloned in to evaluate if there is a conserved ARS here for anaerobic fungi. Because this region is large, up to four kilobases, it was split into three fragments around 1.3 kb so that they could be evaluated individually and in combinations to narrow in on a potential ARS [253]. Because ARS containing plasmids in this library should be stably replicating, they are only dosed into the plasmid once after the start of the culture, and once four hours before applying the selective hygromycin pressure. To evaluate the effectiveness of this resistance, the libraries were compared to a vector without any additional ARS (backbone control) as well as to both antibiotic treated and untreated cultures that show the effectiveness of the hygromycin. Interestingly, the IGS library was able to confer more resistance (~61% growth recovery) than the backbone control (35% growth recovery) and that the genomic library cultures may have conferred some resistance although with much higher variation (Figure 7-6B). When these seemingly resistant cultures were subsequently cultured into media with hygromycin, however, they did not propagate further suggesting the conferred resistance was not stable. This additional resistance conferred here without additional dosing is encouraging, but it further highlights the need for optimizing our selection system and heterologous expression from P_{ENOL} or other promoters.

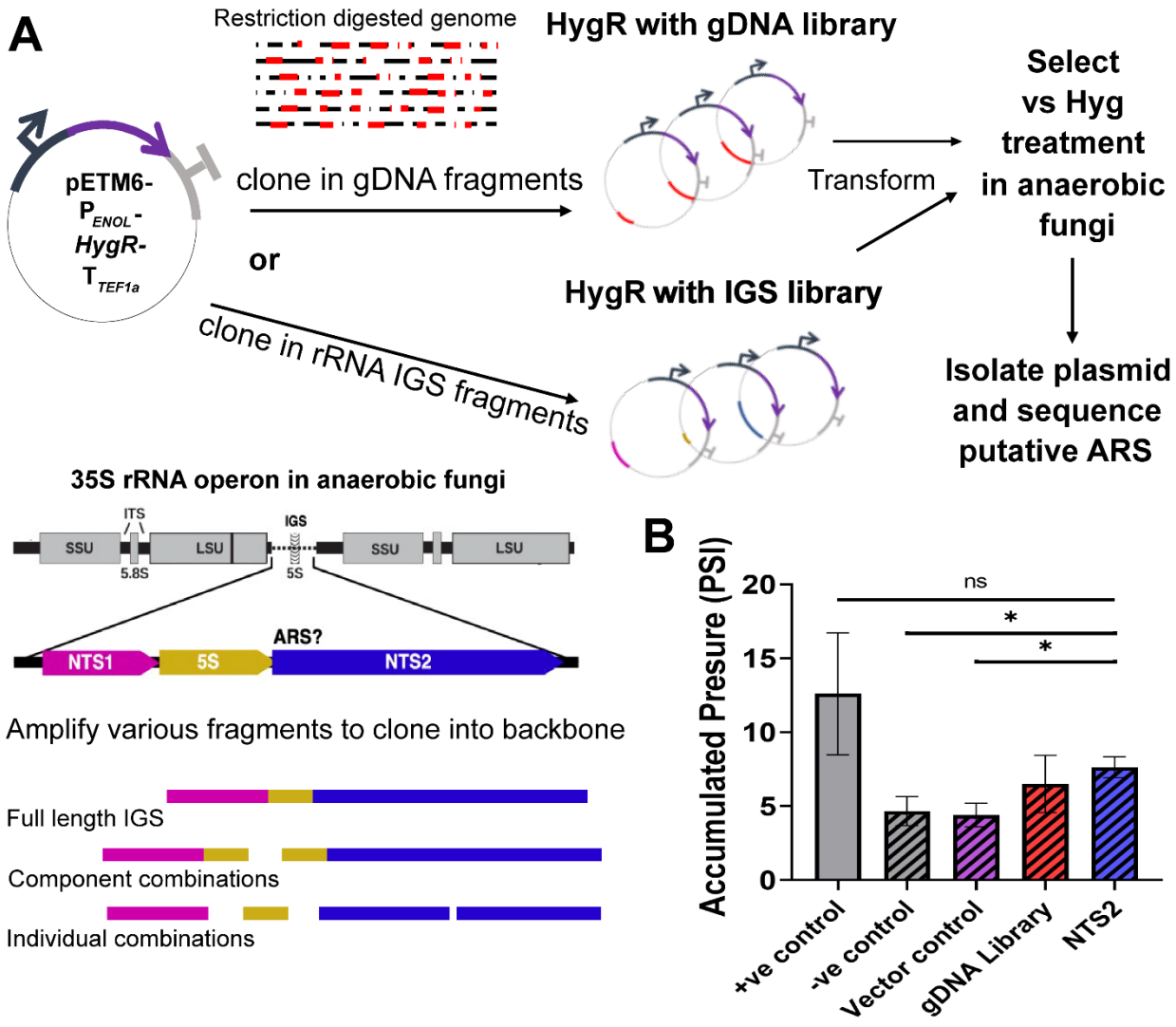


Figure 7-6: Screening for autonomously replicating sequences (ARS). A) Shotgun (top) and conserved rRNA operon loci (bottom) cloning strategies to create vectors with potential ARS. Shotgun approaches fragment genomes into small pieces that can be cloned into the *HygR* resistance backbone while the conserved rRNA loci are PCR amplified in either full or partial fragments and cloned into the resistance backbone. Both libraries of constructs are transformed into the fungi and tested for functionality vs selective pressure; after which successful constructs are isolated and sequenced. B) Results of initial ARS screen showing the accumulated pressure of the cultures. +ve control and -ve control represent the untreated and hygromycin treated cultures of *Neocallimastix sp. GfMa3-1*, respectively. The vector control contains no potential ARS sequences, whereas the gDNA library contains the shotgun genome fragments and the NTS2 contains the non-transcribed spacer region 2 of the rRNA operon. * = unpaired t-test $P < 0.05$. SSU = small subunit, LSU = large subunit, ITS = internal transcribed spacer, 5S and 5.8S = 5S and 5.8S rRNA subunits (respectively), IGS = intergenic spacer, NTS = non-transcribed spacer.

7.3.5 Establishing selection systems for genome engineering

Establishing selection for URA3 / URA5 knockouts.

Like plasmids with antibiotic resistance markers, applying selection pressure to a population of cells allows efficient selection of desired knockout mutants. So, as the first step in creating a gene-editing platform, the ability to select against the essential *URA3* and *URA5* genes that are part of the uracil biosynthesis pathway (Figure 7-7A) was evaluated. Using a fluorinated-substrate (5-FOA, 5-fluoroorotic acid), these genes were leveraged to create a toxic product in anaerobic fungi that will kill the fungi as long as these genes are functional. However, successfully knocking out these genes prevents the conversion of substrate into toxic product and the mutants will survive. Therefore, isolates can be selected and evaluated them for the desired mutations with 5-FOA. When this selection mechanism was tested for anaerobic fungi, 5-FOA was very effective and allowed selection against anaerobic fungi at concentrations as low as 100 µg/ml (Figure 7-7B). At this time however, the ability to knock out these genes with CRISPR-Cas9 has not been developed, but methods like homologous recombination that do not require gene-editing tools have also not yet been evaluated for anaerobic fungi.

Building homologous recombination cassette for anaerobic fungi

Because fungi often reproduce asexually, they need to uptake and incorporate exogenous nucleic acids to diversify their genetic pool. Many fungi use homologous recombination to do this where a piece of DNA that has similar DNA content will be incorporated into the genome. Because anaerobic fungi naturally take up DNA in their zoospore lifecycle stage, it is likely they will also incorporate DNA with significant homology into their genome like other fungi [261]. Since 5-FOA can select against *URA3/5*, *ura*5 homology arms were developed with truncations of the enzyme and multiple cloning sites that allow the insertion of reporters or selection markers (Figure 7-7C). To do this, the *ura5* gene was PCR-amplified from the genome and cloned into the pGEM backbone. After sequencing this gene for *Neocallimastix sp. GfMa3-1*, the cassette was made by amplifying the two halves of the gene with restriction sites that introduce a stop codon and loci for inserting an additional reporter or selection marker (Figure 7-7D). To date, however, the desired *URA3/5* knockouts have not been identified using the 5-FOA selection system. Because the *ura*5 gene is only ~700 bp, the homology arms are relatively small (~350bp each), which maybe too

short for efficient recombination compared to the 1 kb required for recombination efficiencies around 10% in Aspergillus [261]. Previously, the additional sequence beyond the coding region could not obtained due to a lack of high-quality genomic information; however, this sequence can now be obtained from the gapless genomes for these isolates (Chapter 4). If increasing the homology arms enables recombination, these designs and selection systems will be key components for making the first gene knock-ins in anaerobic fungi.

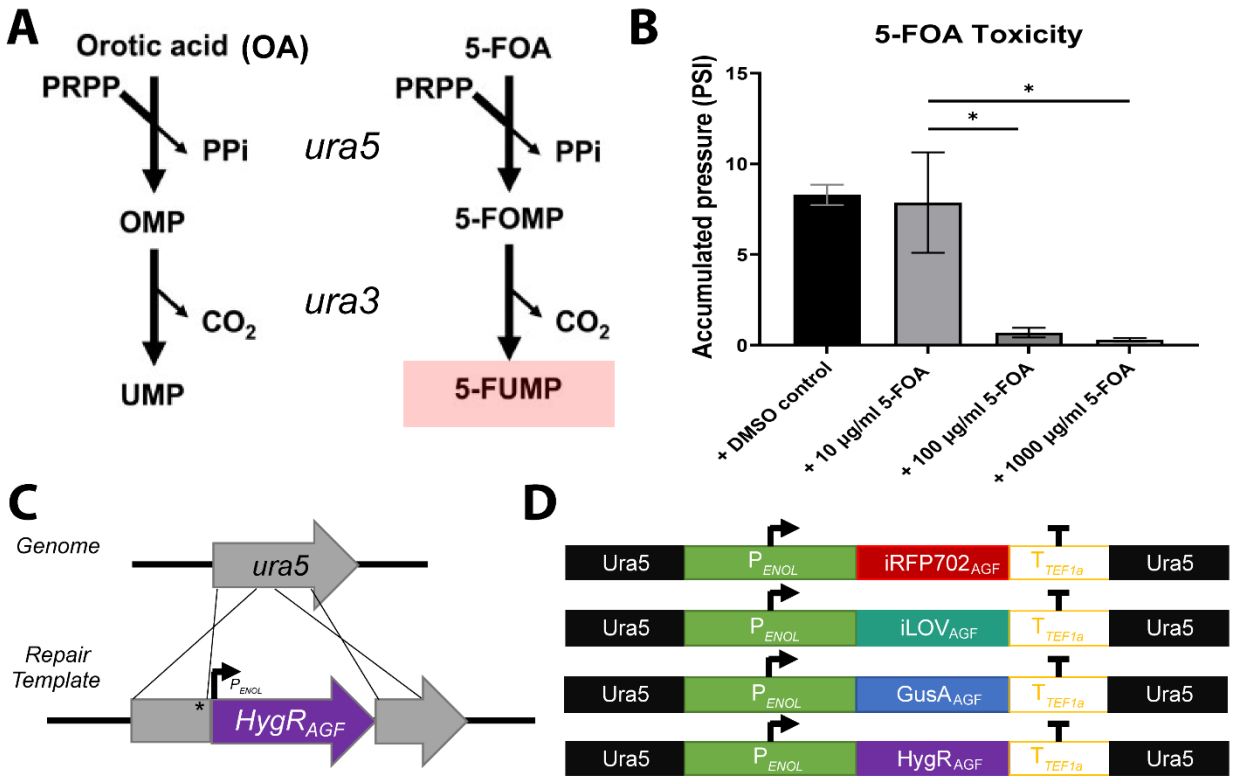


Figure 7-7: Developing the 5-FOA selection system in anaerobic fungi. A) Uracil biosynthesis pathway showing the toxic pathway culminating in the generation of 5-fluorinated uridine monophosphate for 5- fluoroorotic acid (5-FOA). B) Accumulated pressure of 5-FOA-treated cultures showing its toxicity to *Neocallimastix sp. GfMa3-1*. C) Example of homology arm design using the genomic sequence for *ura5* including D) the various recombination cassettes constructed * = unpaired t-test $P < 0.05$. PRPP = phosphoribosyl pyrophosphate, PPI = inorganic pyrophosphate, OMP = Orotidine 5'-monophosphate.

7.3.6 Constructing a gene editing toolbox for anaerobic fungi

Building the Cas9 toolbox

The development of CRISPR-Cas9 gene editing systems was paramount for metabolic engineering and now allows genes to be efficiently knocked out in a wide variety of organisms. The components of this system, a DNA endonuclease and a targeting guide RNA (Figure 7-8A), can be heterologously expressed to cut DNA at specific locus depending on the guide sequence. Additionally, alternative endonucleases, such as Cpf1 that is better at targeting AT-biased genes, are being developed but are not yet as well-characterized as Cas9 is. Therefore, expression of Cas9 in anaerobic fungi was chosen for genome editing using P_{ENOL} from our genetic toolbox. However, because Cas9 was adapted for a eukaryotic organism, the NLS tag (Figure 7-3A) needs to be incorporated to direct the protein to the nucleus for genome cleavage activity. The guide RNA, on the other hand, requires a different promoter as the P_{ENOL} is used to generate mRNA. Several strategies have been used to express these guides including using genomics to identify the U6 promoter, 5S rRNA , or even promoters like P_{ENOL} followed by the hammerhead ribozyme (HHR) that cleaves itself out of the mRNA transcript to yield a mature guide RNA [204], [262]. Here, a system was constructed to express a codon-optimized Cas9 flanked with NLS sequences in combination with two designs that utilize either the 5S rRNA or P_{ENOL}-HHR to drive expression of the guide RNA. Because the 5-FOA selection system had already been developed, guides were designed to target either the *ura3* or *ura5*. However, like the homologous recombination system, no mutants that contain the desired knockout have yet been identified (data not shown). Because the same selection system is being used for both homologous recombination and Cas9 gene editing systems, resolving some of its underlying issues will likely improve this gene-editing system as well.

7.3.7 Enhancing the genetic toolbox for anaerobic fungi

The genetic toolbox developed here represents the basic parts needed for engineering anaerobic fungi, however, future work is needed to make these tools robust enough for genome engineering. Specifically, this fundamental toolbox could be enhanced by improving transformation methods, identifying and characterizing stronger promoters, expanding the types

of quantitative and qualitative reporters, broadening the types of selection pressures and the related selection markers, identifying fundamental plasmid replication components, and fine- tuning the parameters to introduce and knockout genes in the anaerobic fungal genomes. By optimizing each one of these components, the identification and characterization of subsequent parts will become more efficient and provide a richer set of tools for engineering anaerobic fungi.

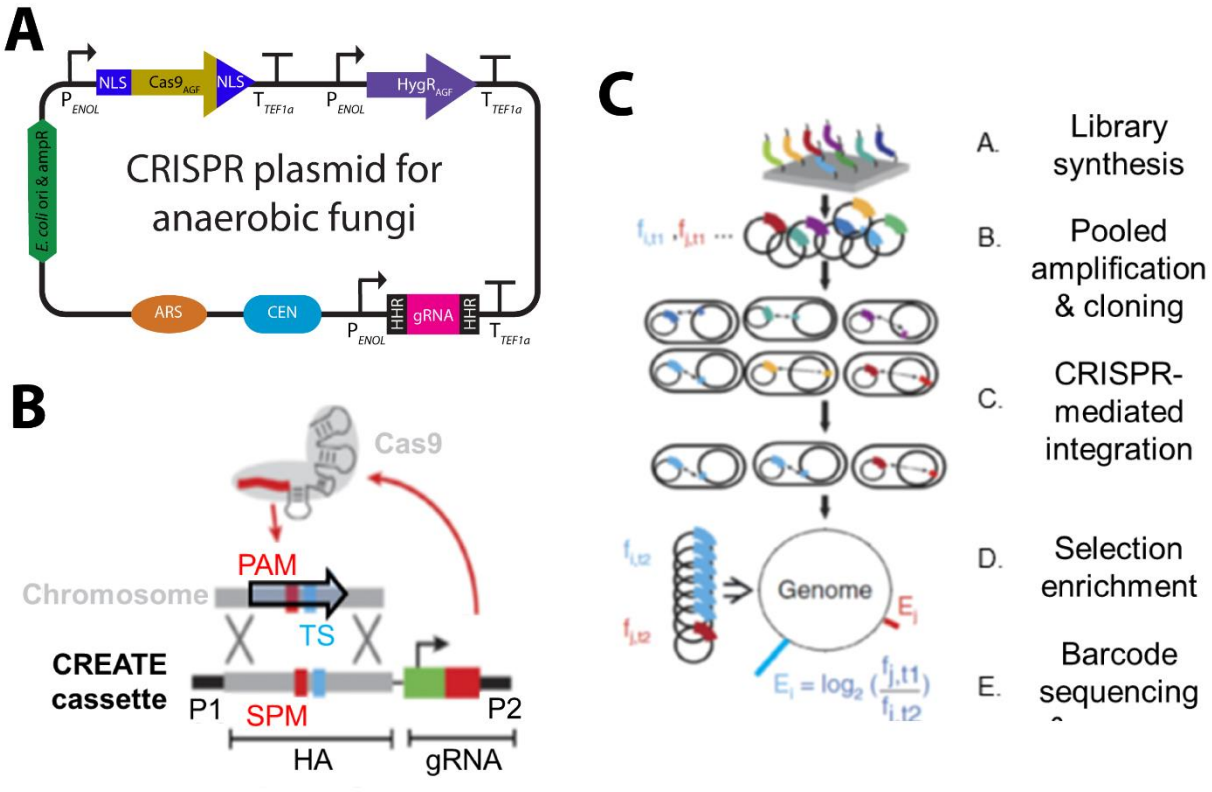


Figure 7-8: CRISPR-Cas9 gene editing toolbox. A) Example of the CRISPR-Cas9 gene-editing vector being developed with the basic P_{ENOL}, T_{tef}, NLS-tag, *HygR*, ARS, and CEN components of our toolbox. The hammerhead ribozyme (HHR) based guide RNA (gRNA)expression system from P_{ENOL} is shown, though 5S rRNA sequences are also being evaluated for their ability to drive guide expression. B) Example of the CREATE cassette that harnesses Cas9, gRNA, and homology arms (HA) to target and replace the native gene (arrow) in the chromosome with a non-functional allele and barcode. To make this system high-throughput, priming sites (P1 and P2) enable multiplexed amplification of the cassettes for cloning in our expression vectors. C) Using the CREATE cassettes, genome-wide mutant libraries can be generated and enriched on various lignocellulosic feedstocks to determine genes associated with these phenotypes. ($f_{j,\,t1}$ – frequency before enrichment, $f_{j,\,t2}$ – frequency after enrichment). B & C adapted from [255].

Optimizing transformation, for example, would theoretically increase the number of cells that uptake and express the reporter genes or resistance markers enabling easier identification of promoters and selection of cell populations, respectively. However, electroporation of anaerobic fungi is limited because it must be performed in an anaerobic chamber or using an oxygen impermeable barrier like mineral oil. Additionally, anaerobic water typically has a high ionic strength because of the cysteine salt used to reduce it; this ionic strength increases the current during electrical transformation, and ultimately, may lead to zoospore death. Similarly, validating whether the fungi that took up these nucleic acids are viable was not possible at this time. However, as more genetic tools are developed for selection and stable plasmid replication, transformation efficiencies can be calculated with colony counts on selective media, and the parameters for transformation that are currently a bottleneck in genetic engineering can be fine-tuned.

Similarly, by expanding the anaerobic fungal toolbox to include low-, medium-, and high-expression promoters, bottlenecks in heterologous expression of reporters and selection markers may be overcome. While the fluorescence of iLOV may be weak as noted above (see section 7.3.3), stronger induction of the enzyme may improve the fluorescent signal strength making it easier to identify colonies of interest. Improving expression of selection markers may also enhance the ability of cells to resist antibiotic selection. While this has not been explicitly determined to be confounding issue of the ARS screen, stronger constitute promoters may help identify ARS carrying strains and propagate them.

Optimization of the current reporters and assays will also assist the assessment of promoter strength. For example, to more accurately evaluate the exogenous *GusA* activity, understanding the expression of the seven native GH2 genes and the conditions under which they are expressed may help resolve issues when high background signal in the *GusA* assay. Additionally, more reporters are being explored in collaboration with the JGI. Three other fluorescent proteins [CreiLOV (flavin-binding), unaG [bilirubin-binding], and IFP2.0 [biliverdin-binding][257]), and up to three genes from the lycopene (red pigment) biosynthetic pathway are going to be synthesized and provided for testing in anaerobic fungi. While fluorescent proteins like GFP cannot fold under anaerobic conditions, other have shown they can fold and fluoresce when exposed to oxygen, and perhaps, can be used as reporters in a destructive assay if their fluorescent signal can be harnessed for quantifying promoter strengths [263], [264].

While optimizing the promoters, reporters, and selection markers will be instrumental in building our tool kit, finding functional components for stable heterologous expression will be equally if not more important. Although repeatedly dosing plasmid for zoospore uptake has provided a method to transiently express proteins, having self-replicating plasmids will enable propagation of the transformants and, theoretically, improved phenotype expression of the population. In addition to ARS, adding centromere binding sequences (CEN) to extra chromosomal plasmids can stabilize expression by ensuring a copy of the plasmid it retained in both parent and daughter cells during replication [254]. However, these sequences cannot enhance heterologous expression without ARS, and ultimately, are still hindered by the inefficient selection system. Because these can be identified using the HiC data [171], which was recently generated for each genome (See Chapter 4), they can rapidly be identified and cloned into the vector once the selection system is improved.

In order to improve our selection system, the selection system should be modified to mimic what has worked for other fungi. One such modification is the use of a solid agar medium instead of liquid cultures [265]. Our group has worked to create agar petri plates for anaerobic fungi, however the fungi do not seem to grow as well on this medium as they do in liquid culture, so future work will need to evaluate whether this can be optimized or if it is the system will have to be adapted to this aspect of their physiology. Working with agar also presents challenges related to plasmid dosing and delayed introduction of selective pressure. In addition to agar plates, future iterations could switch selection methods to auxotrophic knockouts paired with a vector-based selection marker gene instead of antibiotics selection systems that globally inhibit protein translation like hygromycin. Auxotrophic knockouts are widely used in yeast and may be a less severe selection pressure for evaluating genetic variants. However, this would require the ability to knockout an essential gene, such as *ura3* or *ura5*, which is also hindered without robust genome engineering tools. While genome engineering tools are being further developed, these tools will be further enhanced by the improved selection system that is a major focus of future work for the genetic toolbox.

To evaluate issues related to 5-FOA selection of knockout mutants, the parameters of the assay are also being revisiting. For example, because the anaerobic fungi require uracil for cell growth, this selection system may require additional exogenous uracil to be supplemented to the media. At this time, it is unclear if other components of the gene editing system are successfully

making knockouts but are ultimately being limited by media that cannot support their growth. Determining whether and how much uracil should be added to the media will be one of the first steps in resolving these issues.

Additionally, because the homology constructs are in circular plasmids, they may need to be linearized or PCR amplified to be incorporated into the genome [261]. The transformation efficiency of linear and circularized constructs have not yet been compared, but the size and shape the DNA takes is likely to affect the transformation efficiency [266]. Similarly, these homologous recombination experiments have been carried out using natural competency transformations, however, using an optimized electrical transformation may allow more efficient delivery of the homologous DNA. The additive inefficiencies of DNA uptake and recombination with small homology arms may also be one of the reasons for our failures. Therefore, the focus of future work should revolve around optimizing these selection and recombination parameters to generate knockout strains and demonstrate the first gene-editing capabilities of our toolbox.

Independent of the issues underlying the selection system issues, the Cas9 toolbox needs to be optimized. Specifically, the expression of the Cas9 toolbox should be validated and optimized as a first step to troubleshooting this system. As suggested for the reporters and selection markers, evaluating expression from promoters other than P_{ENOL} may improve heterologous expression of Cas9. Using anti-Cas9 antibody or more characterized anti-His antibody, western blot analysis could confirm the expression of this vital gene-editing component. Similarly, expression and validation of guide RNAs could validate that the targeting guide is being produced and that they correctly target the gene of interest. Additionally, because Cas9 is a large protein, vectors containing the Cas9 gene tend to be large and thus transformation of these vectors may be hindered because of their size [266]. Like the other genetic tools, this system will become more efficient as the issues relating to transformation, heterologous expression, plasmid replication, and selection are resolved.

7.3.8 Perspective on gene editing in anaerobic fungi

Despite the wealth of genomic information available for anaerobic fungi, the majority of this sequence is poorly annotated. However, the development of gene editing tools will allow study of these genes by overexpressing or knockout them out. Because of the interest around the lignocellulose degrading abilities of these fungi, their wealth of CAZymes need to be assessed for

functionality and activity. The construction of high-throughput genotype to phenotype systems like CREATE (see ref [255], Figure 7-8B & C) will enable creation of genome-wide maps of enzymes related to the degradation of various lignocellulosic feedstocks by enriching genes required for feedstock deconstruction. Specifically, this platform can be used to shed light on the mechanisms by which anaerobic fungi detect and respond to lignin [14]. Additionally, the gene-function maps created can elucidate the fitness of various genes and reveal systems-level insight about the physiology of anaerobic fungi. While standard physiological and biochemical characterization will help to confirm the gene functions, they will allow annotation of currently unknown genes from anaerobic fungi and, ultimately, links anaerobic fungal genotypes to lignocellulosic phenotypes. Finally, these genetic tools will allow the identification of interesting gene targets for genetic and metabolic engineering to create strains that are able to more efficiently degrade lignocellulose to produce biofuels and other bioproducts. By creating these genetic tools for anaerobic fungi, we will be able to leverage the efficient and broad substrate utilization that current model organisms lack, and that makes them particularly attractive as an emerging model organism for sustainable bioproduction.

7.4 Conclusions

While the genetic toolbox of anaerobic fungi was previously lacking, this work demonstrates for the first time the heterologous expression of various fluorescent reports, a semi-quantitative enzyme activity assay, a plasmid-based selection marker, and the basic components of a gene-editing system. Although these systems will continue to be improved and optimized, the fundamental parts of a genetic engineering toolbox were developed including the basic components for extra-chromosomal plasmid expression, homologous recombination cassettes, and CRISPR-Cas9 gene disruptions. The next steps for engineering anaerobic fungi will focused on improving the transformation efficiency of anaerobic fungi that will help with the development of additional promoters and autonomously replicating sequences and, subsequently, the improvement of heterologous expression for gene-editing. As these tools are iterated on and the parameters of the activity assays and selection systems are optimized, these fundamental toolbox components will allow genetic engineering of the anaerobic fungi. Successful development of these tools will result in deeper systems understanding of anaerobic fungal physiology and allow the advancement

of anaerobic fungi as emerging model microbes for lignocellulosic biofuels and bioproducts production.

7.5 Methods

7.5.1 Transformation

Similar to the roll tube isolation (Chapter 3 and 4), Calkins et al [256] described an agarose-based media for anaerobic fungi and then showed that they could obtain zoospores from it. By layering sterile anaerobic water (SAW) on top of the agar layer, they were able to collect zoospore of *Pecoramyces* that swam into the (SAW). Here this media and flooding approach were adapted to evaluate if *Piromyces* and *Neocallimastix* zoospores could similarly be obtained for transformation. Both 50 ml serum bottles of agar-containing Medium C and Medium B were inoculated with either *Piromyces sp. UH3*-1 or *Neocallimastix sp. GfMa3-1* and were incubated for three days at 39 ˚C. Growth on Medium C-based media typically was superior similar to what was previously observed (Chapter 5, Figure D1). On day 3, approximately 30 ml of sterile anaerobic water (~1.25 g/L cysteine, 1.5 g/L PIPES, pH ~7) was flooded on top of the agar medium and incubated at 39 ˚C for 1 hour. After 1 hour, ssRNA or annealed dsDNA probes were injected into the SAW flooded media to a final concentration of 20 nM and were allowed to incubate for 30 minutes. After 30 minutes, the SAW was evaluated with confocal microscopy for uptake. ssDNA ATGCATGCCTGATATCGATAGCGATATAGCGATAGCTCTATAGCGATAGC[Cyanine3] and ssRNA (GCAUCAGUAUUAUAGCCUAU[Cyanine5]) fluorescent probes were purchased from Sigma (St. Louis, MO) and dsDNA probes were created by annealing to a complementary probe using the Sigma Aldrich protocol (https://www.sigmaaldrich.com/technical-documents/protocols/biology/annealing-oligos.html). Confocal microscopy was performed on the Nikon EclipseTi Microscope and A1-multiphoton imaging system as described in Chapter 3 using DAPI (Final concentration ~300 nM) as a counterstain of nucleic acid content.

Because this method liberates almost exclusively zoospores, the same agar media and SAW flooding approach was used to obtain zoospores for electrical transformation. However, to reduce the scale and potentially further concentrate the zoospore, these experiments were performed in 10 ml agar in a large Hungate (~20ml total volume). These were inoculated the same way as Medium

C roll tubes, however, they were only slightly mixed and not rolled to coat the tube, resulting in cultures that were more similar to bacterial slants tubes. These smaller tubes were then flooded with approximately 3 ml of SAW instead of 30 ml, and instead of incubating with the probe for 30 minutes, they were incubated with the probe for only a few minutes while the rest of the experiment was set up. Briefly, 400 µl aliquots of the SAW were then placed in an electroporation cuvette (2mm gap) containing 100 µl of mineral oil to decrease oxygen permeation and then probe was added to approximately 20 nM. Various voltages were evaluated using the Eppendorf™ Eporator™ and 1000 V provided the most reproducible, though not quite optimal (time constant ~1.8 ms), transformation. These parameters were later iterated on and in combination with the reducing the ionic strength of the SAW (lowering cysteine concentrations) were able to reach time constants close to 5 milliseconds (data not shown). However, despite the suboptimal parameters, the transformation was efficient enough to transform both DNA and RNA. Ongoing work is evaluating the parameters to improve efficiency, though that is partially hindered by the absence of genetic tools.

7.5.2 Promoter identification

Bioinformatics

The sequence of the enolase promoter from *Neocallimastix frontalis* was obtained from Fischer et al [267], and evaluated for int conservation across the available genomes on Mycocosm. The entire coding region and the preceding 1000bp were aligned using both the genome browser in Mycocosm and manually in Mega7 [156]. Because conservation of the *N. frontalis* promoter was most similar to the *N. californiae* promoter, and the two *Piromyces* promoters were most similar to each other, these genes alignments were used to create degenerate primers that targeted highly conserved portions of this promoter region.

Molecular cloning of P_{enol} and the basic tools

Primers for the enolase promoter P_{enol} were designed to achieve an optimal annealing temperature of 50 ˚C and ultimately required them to be around 35nt. However, the GC content was between 15-20% (fwd: 5’-ACACAATCAATATTATTTTTAGTTATAATAAACT-3’, rev: 5’gatcTTTTATTAATTATTGTTTGTTACAAAA-3’). Because this region is so AT-rich (~12%

GC), this was an issue at nearly all loci along the promoter. As a result, initial PCR amplification using the standard Phusion settings was unsuccessful. Instead, the most optimal PCR attempts required that the annealing temp be 45 ˚C and the extension temp lowered to 64.5 ˚C (consequently increasing the extension temperature to about 1 minute to ensure the lower efficiency at a lower temperature did not prematurely end amplification) for 30 cycles. After PCR purification with Zymogen DNA Clean & Concentrator (Zymo Research, Irvine, CA,USA), this sequence was amplified with flanking restriction sites (5' XbaI, 3' BglII) and then cloned into the pETM6 plasmid backbone [238]. This plasmid formed the base of our vector that was then subcloned with various reporters and selection markers. Specifically, iLOV was cloned into this backbone by restriction enzyme digest of PCR products from the pQE80L plasmid [258] flanked with BglII and BcuI restriction sites. Similarly, *HygR* was cloned PCR amplified from the pTKIP-hph plasmid (Addgene # #41066). Codon-optimized iRFP, iLOV, *GusA*, and *HygR* flanked with these restriction sites were ordered from IDT with universal M13 barcodes flanking the entire gBlock for simple PCR amplification; amplicons were then cloned as the plasmid-based genes. Additionally, the promoter region was sequenced using the universal primers for the pETM6 backbone (submitted to GeneWiz - South Plainfield, NJ, USA). After obtaining the full promoter sequence for *Neocallimastix sp. GfMa3-1*, additional truncations of the promoter were amplified using internal primers. This same pipeline was used to identify and clone the putative TEF1a terminator (T_{tef}) and it will also be used to truncate the new potential promoters and terminators identified in our genomes in coordination with the JGI (Chapter 4).

7.5.3 Evaluating transient expression of reporters and selection markers

Fluorescent reporters: iLOV and iRFP

For reporter experiments, natural competency was leveraged because it allows supplementation of multiple plasmid doses and is not subject to the yet-to-be-optimized electroporation parameters that might jeopardize the experiment. Using the same agar growth media and SAW flooding as for the transformation, expression plasmids were dosed in instead of labeled oligos. The pETM6-P_{enol}-iLOV, pETM6-P_{enol}-NLS-iLOV (NLS sequenced pulled from N-terminus of *Neocallimastix californiae*'s Histone 2B [268]), and pETM6-P_{enol}-iRFP were each using the same basic protocol. Briefly, agar containing Medium C tubes containing

chloramphenicol were inoculated with *Neocallimastix sp. GfMa3-1* and grown for three days. On the third day, the tubes were flooded with 3 ml of SAW and incubated for 60 minutes at 39 ˚C. After 60 minutes, ~2 µg of DNA diluted in SAW was added and incubator for an additional 30 minutes. After 30 minutes, the entire flooding volume was removed and used to inoculate the fresh 10 ml Medium C glucose tubes, the 3 ml was split such that each 1 ml was used to inoculate a separate replicate. Tubes were then vented to 0 PSI gauge pressure, incubated at 39 ˚C, and pressure/growth/ biomass monitored every 12-24 hours. Plasmids were subsequently dosed as shown in the timeline in Figure 7-5B. At each time point, tube caps were flame-sterilized, pressures were recorded, gauge pressure brought back to zero, and 2 µg of plasmid DNA was added via syringe. Gauge pressures were zeroed once again and then placed back in the incubator. After 72 hours, cells were stained with DAPI or Syto17 and imaged as previously described (Chapter 3).

GusA semi-quantitative reporter assay

Because there are no stably replicating sequence (ARS), any current quantitative reporter assays are also dependent on the dosing experimental setup demonstrated above. Here, the optimized *GusA* gene was evaluated as a reporter of P_{enol} activity using the pETM6-P_{enol}-*GusA*-T_{tef} construct. Similarly, 2 µg of plasmid was dosed into *Neocallimastix sp. Gf-Ma3-1* according to the dosing timeline above (Figure 7-5B) in Media B. However, after about 60 hours, the fluorogenic substrate MUG (4-Methylumbelliferyl-β-D-galactopyranoside) was added at a final concentration of 1 mg/ml in DMSO and then incubated for an additional two hours. To assay activity, cells were spun down in Eppendorf tubes, washed twice with, 1x PBS, resuspended in 500ul of 1x PBS, and lysed with zirconia beads on the MP Biomedical FastPrep homogenizer (1 minute lysis, Power = 6). After lysis, debris and beads were pelleted and 200 ul of lysate was added to separate wells of a 96 well plate. Fluorescence was read using 361 nm to excite, and 447 for emission on a Biotek Synergy Neo plate reader. Additionally, cells were visualized on an EVOS FL Imaging System as an additional way to verify activity.

HygR dosing and selection

Like the previous reporters, resistance marker gene expression is hindered without a replication sequence. Therefore, the same dosing experimental setup was used with the pETM6-

P_{ENOL}-*HygR* plasmid and hygromycin B selection. As above, 2 µg of the plasmid was dosed approximately every 12 hours to enable heterologous expression by *Neocallimastix sp. GfMa3-1* zoospore that uptake the plasmid. However, in this experiment, hygromycin B is added after approximately 50 hours and an additional plasmid dose was given shortly after selection (Figure 7-5B). As when plasmid is dosed, the tube caps were flame-sterilized, pressures were recorded, gauge pressure brought back to 0, and the antibiotic added to 50ug/ml (diluted in SAW), and then the tubes were re-zeroed. As above, pressure/growth/biomass were monitored every 12 hours for a total of 120-144 hours.

7.5.4 Building and testing ARS plasmid libraries

To overcome the issues associated with plasmid dosing and create stable expression vectors, a library approach was used to identify autonomously replicating sequences (ARS) [260]. In order to be able to apply selective pressure and enrich the ARS-positive organisms in anaerobic fungi, a variant of the pETM6-P_{ENOL}-*HygR* with a high-copy number and a multiple cloning site was used for ARS insertion. To construct the library, 2 µg vector DNA was digested separately with BamHI (complementary to BglII), EcoRI, and NdeI for approximately 60 minutes with CIAP treatment (1U/20 µl) for an additional15 minutes to ensure re-ligation of backbone or partial digestions are eliminated. Simultaneously, 6 µg of genomic DNA was digested with either BglII, EcoRI, or NdeI for about 90 minutes. After digestion, both the digests were cleaned up using the Zymogen DNA Clean & Concentrator (Zymo Research, Irvine, CA,USA) – final elution in 10 ul. One 50 ml ligation was set up for each vector and gDNA pair digested with the same enzymes using the T4 Rapid Ligation kit (Thermo Fisher Scientific). Reactions were prepared according to the manufacturer's instructions and incubated at room temperature for two hours and then subsequently denaturing the ligase by heating the reaction to 65 ˚C for 10 min. In attempt to capture the entire library, each 50 ul reaction was transformed 10 ul at a time into highly competent DH5α *E. coli* by heat shock at 42 ˚C for 45 seconds and then grown out in 1 ml rich SOC media. To capture the entire library, each of the five transformations for all three digestion pairs (15 total) were combined in a 250 ml shake flask with 40 ml of fresh SOC. The whole library was incubated for 60 minutes at 37 ˚C, 250 rpm. After 1 hour, 100 ul of the library was diluted 1:10, plated on an LB/amp, and incubate overnight at 37 ˚C. The pooled transformations were then added to a 2L flask with 500 ml LB/Amp and 0.2 sterile filtered glucose. This was incubated until the OD reached

2 (~4 hrs), and a fresh 2L flask with 400 ml LB/Amp and 0.2 sterile filtered glucose was inoculated using 100 ml of the previous culture and incubated at 37 ˚C overnight. The remaining 400 ml of the initial library was centrifuged at 3000g for 15 min. Pellets were resuspended in 10 ml of SOB + 15% glycerol, and then split into cryovials and frozen at -80 ˚C for stocks. The following morning plates were counted to determine the library size. Finally, plasmid library DNA was isolated from the overnight culture using six midi preps (Zymo Research). The IGS libraries were constructed in a similar manner using PCR amplicons of the various fragments (See figure 7-6) flanked with different combinations of the BamHI, EcoRI, or NdeI restriction sites because some fragments contain internal restriction sites. These libraries were evaluated in anaerobic fungi using the same natural competency protocol as the *HygR* dosing experiment (see section 7.5.3), but plasmids are only dosed once during the initial zoospore harvest, and once at 24 hours (~6 µg each to allow the whole library to be evaluated). Additionally, after ~4-5 days, the cultures were sub-cultured into both fresh media and hygromycin-treated media. However, none of the library-containing cultures propagated on either suggesting that they did not resist the selection. Some of the controls also show no-weak growth suggesting the fungi may have been in a later stationary phase or dead already due to lack of nutrients.

7.5.5 Building a gene-editing system

Evaluating 5-FOA toxicity

Because 5-FOA (5-fluoroorotic acid) takes advantage of a conserved essential pathway in anaerobic fungi, it can be leveraged to select for anaerobic fungi that have a knockout in the pathway. Specifically, the 5-FOA concentration was evaluated to determine the level at which selection pressure should be applied. 5-FOA was dissolved in DMSO at 100g/L and then added to Medium C roll tubes while still molten at concentrations from 0 µg/ml to 1000 µg/ml. Like roll tubes, *Neocallimastix sp. GfMa3-1* was inoculated into the tube once it cooled to approximately 45 ˚C. Accumulated pressure and visible signs of growth were monitored for 120-144 hours.

Building ura5 homologous recombination cassettes

At the time of this work, the genomes for our isolates were still being isolated and sequenced. Therefore, the available genomes available on Mycocosm were leveraged in order to

get insight into our isolates' genomes. Specifically, the conservation (% ID) of the *ura5* and *ura3* genes of all isolates were evaluated in order to design primers to amplify these loci as repair cassettes given they can be selected against with 5-FOA (Figure 7-7). Degenerate primers were designed to amplify the *ura5* locus first because it did not contain any introns, which are less conserved across the available anaerobic fungal genomes. Using Phusion polymerase, amplicons were generated by annealing at 50 ˚C and extending at 64.5 ˚C for 1 minute, and then were cloned into pGEM using their standard protocol. After screening and verifying colonies positive for *ura5* gene, this was used as the base plasmid for designing recombination cassettes given that these sequences had been obtained for *Neocallimastix sp. GfMa3-1*. Primers were then designed to amplify the first and second half of the gene separately with primers containing restriction site overhang that could be digested and ligated together. The two *ura5* halves were then digested, cleaned up, ligated together, and then cloned into the pGEM vector (Promega). Additionally, these overhangs formed a stop site to truncate the *ura5* gene and introduce a multi-cloning site that the basic reporter and selections marker parts could be cloned into for gene knock-ins, These reporter, selection marker cassettes have been made, but are still hindered at this time by the lack of gene knockout system.

Constructing a Cas9 toolbox and initial ura5 knockout attempts

The CRISPR-Cas9 tool kit relies on both the expression of the Cas9 endonuclease and the guide RNA. The tools developed in this chapter enable expression of the Cas9 protein with P_{ENOL}; however, P_{ENOL} cannot drive guide RNA expression and requires the consideration of different strategies. Firstly, guide RNAs flanked with the hammerhead ribozyme sequence that can themselves out of messenger RNA [262] were designed and synthesized (TWIST biosciences). Specifically, targets were designed to cleave the *ura5* locus near our repair cassette. These gBlocks were flanked with the universal M13 sequences and then cloned into the pETM6-P_{ENOL} base plasmid. Similarly, the 5S rRNA locus has been used because this sequence can form viable guides even when still attached to the 5S subunit. Again, guides were designed and synthesized (TWIST Biosciences), however after these 5S driven guides were amplified, they were cloned into the pETM6 base plasmid because they do not require the P_{ENOL} to drive their expression. Using the ePathBrick cloning scheme [238], guides constructs were then cloned into the pETM6-P_{ENOL}-NLS-Cas9-NLS-Tt_{ef} that was constructed by ligating three gBlocks together to form the large Cas9

gene and then cloning that into the pETM6-P_{ENOL} NLS-cassette-T_{tef} designed to target desired proteins to the nucleus. After these constructs were assembled, they were evaluated using the dosing experimental design demonstrated above. Briefly, 2 µg of plasmid was transformed into harvested zoospores and then subsequently every 12 hours for the first 48-60 hours. Constructs targeting a random sequence or the *ura5* loci were dosed with and without the repair cassette to evaluate homologous recombination alongside DNA cleavage activity. After about 60 hours, transformed cultures were sub-cultured into 5-FOA-treated roll tubes or serum bottles and evaluated for growth. This selection system, however, still require improvement, because variable growth was observed in both the positive controls and the negative controls without any repair cassette or Cas9 expression.

8. CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

As advances in sequencing technologies continue to help us uncover the biodiversity of the planet, it is important that we also develop the tools necessary to exploit this potential for sustainability. Anaerobic fungi are a prime example of microbes whose potential is just beginning to be realized thanks to a handful of recently sequenced genomes. Because they are responsible for approximately 50% of the biomass degradation [93] in the rumen microbiome, their ability to degrade agricultural crop waste into simpler forms of accessible carbon can play a substantial role in advancing the emerging bioeconomy [1], [12], [13]. While many strains have been isolated since their discovery over 50 years ago, they have not been fully adopted in agricultural and industrial applications [11]. Advances in DNA sequencing and synthesis, however, are helping to close the gap between their potential and utility by allowing us to overcome substantial challenges in genome sequencing and heterologous expression. By leveraging these advances in technology, we are now closer than ever to being able to leverage the full potential of anaerobic fungi through genetic engineering. The advances made by this dissertation are summarized below.

First, in order to evaluate anaerobic fungi as a platform for efficient plant biomass hydrolysis, three novel anaerobic fungi were isolated and characterized in Chapters 3 and 4. As demonstrated here, these organisms efficiently degrade a variety of wasted biomass from agricultural crops to food waste (Chapter 3). Additionally, the highest quality genomes for anaerobic fungi were obtained by adapting plant-based extraction protocols and pairing them with advanced chromosomal capture sequencing technologies (Chapter 4). The improved resolution of these genomes now allows their full potential to be revealed and the identification of specific genes, or clusters of genes, that can be leveraged for biotechnology. Finally, for the first time, the ploidy and chromosomal number of these anaerobic fungi was determined and will aid future engineering endeavors that target specific loci. By characterizing these degradative capacity and genomic content of these isolates in these chapters, the foundation is laid for harnessing their potential through both genetic engineering and non-genetic engineering strategies.

Next, while the suite of genetic tools to engineer anaerobic fungi was being constructed, two approaches were used to leverage their metabolic (Chapter 5) and genomic (Chapter 6) potential

of anaerobic fungi in the absence of genetic engineering tools. In Chapter 5, the efficient lignocellulose degrading abilities of anaerobic fungi was paired with another emerging model microbe to produce commodity and fine chemicals. By utilizing *K. marxianus* strains that can consume the metabolic byproducts of anaerobic fungi in a two-step bioproduction system, more recalcitrant biomass was converted to valuable metabolites, specifically ethyl acetate, 2-phenylethanol, and isoamyl alcohol. The productivity of *K. marxianus* on the hydrolysates of simple and complex feedstock was comparable to that on optimized rich yeast media. In Chapter 6, using recently acquired genomes, the genes of anaerobic fungi were shown to be poor first for heterologous expression in *E. coli* to the point that highly expressing them can severely limit growth. Through codon-optimization, however, it is possible to express these genes without growth defects. Similarly, codon-optimization enabled *E. coli* to produce up to 2.5 g/L mevalonate from genes and pathways that otherwise make negligible, if any, product. Ultimately, this case study demonstrates that the genetic potential of anaerobic fungi may be realized by studying and adapting their genes for model organisms. Although the strategies laid out in these chapters demonstrate that the potential of anaerobic fungi can be leveraged without genetic tools, their full potential for biofuel and bioproduct conversion remains limited until these tools are developed.

The genetic toolbox of anaerobic fungi to this point consists only of one poorly characterized promoter and a natural competency method for RNA uptake. Therefore, Chapter 7 focused on developing DNA uptake methods, both through natural and electroporation, as well as using available genomic information to identify and build genetic parts. Further characterization of the enolase promoter is provided, in addition to characterization of two fluorescent reporters and a semi-quantitative reporter assay for evaluating future parts. Similarly, this promoter expressed *HygR* resistance for the first antibiotic selection system in anaerobic fungi that is currently being used to identify other critical plasmid components. Finally, the 5-FOA selection system was shown to be effective against anaerobic fungi and will be instrumental in making the first gene knockouts in anaerobic fungi. While constructs for homologous recombination and CRISPR-Cas9 based gene-editing were constructed and tested, they are still being optimized alongside the selection assays to form the basis of our genome engineering toolbox. These tools form the foundation of the genetic toolkit that is still expanding thanks to collaborations with the Joint Genome Institute and the Department of Energy. Ultimately, these tools will enable the creation of genetically engineered strains that enhance our ability to produce biofuels from otherwise wasted biomass.

8.2 Future work and perspectives

As more genomic information, metabolic characterization, and genetic tools come to fruition, anaerobic fungi have enormous potential to form the foundation of sustainable bioproduction platforms. Whether as platforms for production of CAZyme cocktails for enzymatic hydrolysis of feedstocks, lignocellulosic deconstruction specialists in microbial bioproduction consortia, or engineered one-pot strains for biofuel and chemical production, the potential of anaerobic fungi will be actualized in a variety of ways. Although three closed genomes are provided here, there are hundreds of strains and new genera whose genomic potential is yet to be evaluated [269], [270]. While the initial anaerobic fungal genomes sequenced have the largest repertoire of CAZymes, unevaluated genomes from other genera may contain larger repertoires or unique enzymes acquired from diverse environments that have not yet been explored.

Similarly, we are only beginning to scratch the surface of the biosynthetic potential of anaerobic fungi that may contain new interesting compounds and chemistries beneficial to both chemical and pharmaceutical production. For this reason, additional studies that equip model organisms to express the biosynthetic gene clusters of anaerobic fungi will be paramount in actualizing this potential. Recently, the HEx system for heterologous expression specifically of fungal natural product pathways was developed in *S. cerevisiae* [271], though clusters of anaerobic fungi remain to be tested. As shown by the HEx pipeline and in agreement with our work on codon optimization, these endeavors are driven by large-scale synthesis efforts including the use of a variety of promoters to fully evaluate the way these products are produced in a high-throughput system. In addition to gene expression, demonstration of fungal natural products formation and function will require multi-systems collaborations that can identify chemical structure, activity, and effectiveness [272]. While the discovery of natural products is historically challenging [163], the need for finding new medicines, namely antibiotics, will only continue to grow as population and antibiotic resistance continue to grow [59]. Genetic tools for anaerobic fungi, however, could eliminate the need for heterologous expression altogether by allowing us to study their biosynthetic pathways in the native fungal environment.

The genetic tools developed here will not only enable anaerobic fungi to be adapted for a variety of processes, but they will also allow deeper study of their physiology and gene function. For example, anaerobic fungi produce a variety of fermentation products including H_2, CO_2, acetate, lactate, formate, and ethanol, among some other minor products like succinate [12]. It is

not clear, however, what drives the production of these metabolites, and which, if any, of the associated genes are essential to their physiology. Creating targeted genetic modifications through the use of CRISPR will ultimately enable us to answer these questions in route to creating strains for bioproduction. Further, efficiency of the two-stage bioproduction platforms could be improved if genetic engineering tools were developed to regulate CAZyme expression or flux of metabolic pathways in anaerobic fungi. Knocking out genes that create otherwise lost carbon, like CO_2 and formate, could results in higher product conversions in two-stage bioproduction platforms depending on how much more usable carbon is liberated. Similarly, adapting systems like CREATE [255] that enable high throughput genotype to phenotype maps of the genome will lead to advances in the understanding of how anaerobic fungi tailor their secretomes to degrade complex feedstocks and annotation of the associated genes [14], [18], [251]. Because systems like CREATE use top-down engineering, they can rapidly create strains that perform more efficiently simply by tuning the selective feedstock conditions rather than the cumbersome one-knockout-at-a-time strategies associated with bottom-up engineering. Bottom-up engineering strategies, however, will be equally important for creating the foundation auxotrophic strains needed for synthetic biology in anaerobic fungi. From such foundations, the basic tools being developed will lead to more complex synthetic circuits such as metabolic valves [273] and kill switches [274] that can regulate the genome in order maximize CAZyme expression and reduce substrate utilization. Overall, genetic tools for these emerging non-model anaerobic fungi will allow their efficient and broad utilization of robust lignocellulosic substrates to be leveraged for sustainable bioproduction in ways that current model organisms cannot accommodate.

APPENDIX A. COMPOSITIONAL ANALYSIS OF THE PLANT BIOMASS USED IN THIS STUDY

This appendix is adapted from [14] and the Master's thesis of C.A. Hooker [100] but is reproduced here in its entirety for full context and interpretation.

Table A.1 NREL compositional analysis of the renewable plant biomass used in this study

Component	Corn Stover	Switchgrass	Sweet Sorghum	Grain Sorghum	Forage Sorghum	Orange Peel	Fresh Cut Alfalfa	Dry Bale Alfalfa
Glucan (%)	37.7 ± 1.1	34.8 ± 0.28	50.8 ± 0.40	44.7 ± 1.53	39.5 ± 1.37	22.8 ± 5.5	15.2 ± 4.4	15.6 ± 3.1
Xylan (%)	22.1 ± 0.42	22.1 ± 0.42	11.2 ± 0.04	14.0 ± 0.88	20.5 ± 0.61	14.7 ± 0.3	5.1 ± 1.8	6.3 ± 1.2
Arabinan (%)	3.4 ± 0.09	2.7 ± 0.01	2.2 ± 0	3.3 ± 0.25	2.8 ± 0.09	6.1 ± 1.2	2.1 ± 0.1	2.3 ± 0.6
Acetyl (%)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Total Lignin (%)	17.7 ± 1.5	23.3 (NR)	16.1 ± 0.34	14.8 0.01	20.5 ± 0.38	ND	ND	ND
Water Extractives (%)	11 (NR)	6.8 (NR)	22.3 (NR)	34.0 (NR)	27.5 (NR)	0.556 (NR)	30.4 (NR)	35.7 (NR)
Ethanol Extractives (%)	4.5 (NR)	2.4 (NR)	7.1 (NR)	5.5 (NR)	3.8 (NR)	0.024 (NR)	6.0 (NR)	4.4 (NR)
NR- No Replicates, ND- No data								

Table A.2: NREL compositional analysis of 2014 harvested poplar constructs used.

Component	INRA 717	NM 6	0998-45	1036-73	1020-44	1035-41	F5H 37	F5H 64
Glucan	0.43 ± 0.00	0.41 ± 0.00	0.43 ± 0.00	0.45 ± 0.00	0.43 ± 0.00	0.45 ± 0.00	0.43 ± 0.00	0.44 ± 0.00
Xylan	0.21 ± 0.01	0.20 ± 0.00	0.22 ± 0.00	0.22 ± 0.00	0.22 ± 0.00	0.21 ± 0.00	0.22 ± 0.00	0.21 ± 0.01
Arabinan	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Acetyl	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00
Acid Insoluble Residue	0.15 ± 0.00	0.17 ± 0.00	0.14 ± 0.00	0.17 ± 0.00	0.15 ± 0.00	0.17 ± 0.00	0.14 ± 0.00	0.17 ± 0.00
Acid Soluble Lignin	0.06 ± 0.00	0.06 ± 0.00	0.04 ± 0.00	0.05 ± 0.00	0.08 ± 0.00	0.05 ± 0.00	0.09 ± 0.00	0.06 ± 0.01
Ash	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Glucose	0.01 ± N/A	0.01 ± N/A	0.01 ± N/A	0.01 ± N/A	0.00 ± N/A	0.01 ± N/A	0.01 ± N/A	0.00 ± N/A
Xylose	0.01 ± N/A	0.00 ± N/A	0.01 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A
Arabinose	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A
Acetic Acid	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A	0.00 ± N/A
Other Water Extractives	0.05 ± N/A	0.05 ± N/A	0.04 ± N/A	0.02 ± N/A	0.00 ± N/A	0.04 ± N/A	0.02 ± N/A	0.00 ± N/A
Ethanol Extractives	0.02 ± N/A	0.05 ± N/A	0.05 ± N/A	0.04 ± N/A	0.05 ± N/A	0.04 ± N/A	0.04 ± N/A	0.04 ± N/A
NA- No replicates								

Table A.3 : Syringyl lignin content of the 2014 poplar constructs used in this study.

Construct	Syringyl lignin molar percentage
0998-45	5
1036-73	28
1065-41	34
1020-44	42
INRA 717	64
NM-6	68
F5H-37	77
F5H 64	98

Table A.4: NREL compositional analysis of the 2017 harvested poplar

	Control Lines				Low Syringyl Lignin/ Hydroxy Guaicyl Lignin				Low Syringyl Lignin/ High Guaicyl Lignin				High Syringyl Lignin			
	INRA 717		NM 6		0998-45		1036-73		1020-44		1035-41		F5H 37		F5H 64	
Component	% dry mass	Std. Dev.	% dry mass	Std. Dev.	% dry mass	Std. Dev.	% dry mass	Std. Dev.	% dry mass	Std. Dev.	% dry mass	Std. Dev.	% dry mass	Std. Dev.	% dry mass	Std. Dev.
Glucan	43.40%	0.30%	41.46%	0.41%	43.49%	0.41%	44.96%	0.17%	43.26%	0.21%	44.54%	0.21%	43.22%	0.25%	44.49%	0.34%
Xylan	21.20%	0.66%	19.52%	0.20%	22.14%	0.26%	21.76%	0.11%	22.14%	0.32%	20.97%	0.26%	22.04%	0.18%	21.48%	0.56%
Arabinan	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Acetyl	1.80%	0.06%	1.62%	0.03%	1.83%	0.03%	1.79%	0%	1.83%	0%	1.79%	0.14%	1.82%	0%	1.87%	0.03%
Acid Insoluble Residue	15.46%	0.18%	16.91%	0.18%	14.25%	0.06%	16.51%	0.28%	15.01%	0.18%	16.78%	0.39%	13.97%	0.12%	16.67%	0.07%
Acid Soluble Lignin	5.90%	0.28%	5.96%	0.30%	4.08%	0.16%	4.66%	0.21%	7.99%	0.26%	4.64%	0.26%	9.05%	0.28%	5.62%	0.54%
Ash	0.70%	0.05%	0.81%	0.01%	0.69%	0.18%	0.61%	0.11%	0.61%	0.05%	0.53%	0.06%	0.58%	0.02%	0.49%	0.07%
Glucose	0.99%	-	0.74%	-	1.37%	-	0.68%	-	0%	-	0.69%	-	0.70%	-	0%	-
Xylose	0.79%	-	0%	-	0.53%	-	0.40%	-	0%	-	0%	-	0%	-	0%	-
Arabinose	0%	-	0%	-	0%	-	0%	-	0%	-	0%	-	0%	-	0%	-
Acetic Acid	0%	-	0.12%	-	0%	-	0%	-	0%	-	0%	-	0%	-	0%	-
Other Water Extractives	5.41%	-	4.86%	-	3.76%	-	1.95%	-	0%	-	4.29%	-	1.57%	-	0.31%	-
Ethanol Extractives	2.43%	-	4.73%	-	4.55%	-	3.70%	-	4.55%	-	3.64%	-	4.29%	-	3.94%	-
Mass Balance	98.07%	-	96.74%	-	96.68%	-	97.03%	-	95.39%	-	98.24%	-	97.24%	-	94.88%	-

Table A.5: Lignin analysis of the 2017 harvested poplar

	Control Lines				Low Syringyl Lignin/ Hydroxy Guaicyl Lignin				Low Syringyl Lignin/ High Guaicyl Lignin				High Syringyl Lignin			
	INRA 717		NM 6		0998-45		1036-73		1020-44		1035-41		F5H 37		F5H 64	
Lignin Amount (nmol mg $^{-1}$ cell wall)	Amount	Std. Dev.	Amount	Std. Dev.	Amount	Std. Dev.	Amount	Std. Dev.	Amount	Std. Dev.	Amount	Std. Dev.	Amount	Std. Dev.	Amount	Std. Dev.
H	1.7	0	1.2	0.4	0	0	1.5	0.1	2.2	0.6	0.7	1	1.7	0.4	1.6	0.2
G	80.5	0.3	75.1	0.6	65.3	2.6	88.3	0.7	101.3	9.3	100.7	2.4	17.2	0.2	35.4	1
S	164.1	0.9	152.2	12.2	17.2	1.3	107.3	0.5	136.1	6.4	111.8	0.8	253.4	13.8	221.6	0.5
total	246.3	1.2	228.5	11.2	82.5	3.9	197.1	0.1	238.7	16.4	213.2	2.5	272.2	14	258.5	1.6
Percent	%	Std. Dev.	%	Std. Dev.	%	Std. Dev.	%	Std. Dev.	%	Std. Dev.	%	Std. Dev.	%	Std. Dev.	%	Std. Dev.
H	0.7	0%	0.5	0.2%	0	0.0%	0.8	0.0%	0.9	0.2%	0.3	0.5%	0.6	0.1%	0.6	0.1%
G	32.7	0%	32.9	1.9%	79.2	0.6%	44.8	0.3%	42.2	1.0%	47.2	0.6%	6.3	0.4%	13.7	0.3%
S	66.6	0%	66.5	2.1%	20.8	0.6%	54.4	0.3%	56.8	12.0%	52.4	1.0%	93.1	0.3%	85.7	0.4%

Table A.6: Glucan and xylan conversion efficiencies for *Piromyces sp. UH3-1* when grown on untreated corn stover for 168 hours

	Corn Stover
Glucan in Raw (g)	0.377 ± 0.011
Glucan in Spent (g)	0.156 ± 0.010
Xylan in Raw (g)	0.221 ± 0.004
Xylan in Spent (g)	0.157 ± 0.002
Glucan Released	58.6% ± 3.6%
Xylan Released	29.0% ± 2.0%

APPENDIX B. ADDITIONAL DATA FOR *PIROMYCES SP. UH3-1*

This appendix is adapted from [14] and the Master's thesis of C.A. Hooker [100] but is reproduced here in its entirety for full context and interpretation.

Formal Species Description

Taxonomy

Piromyces sp. UH3-1 Ethan Hillman, Adrian Ortiz-Velez, Kevin Solomon, sp. nov.

Index Fungorum number: IF554555 JMRC: SF:012426

Typification: The holotype (Figure 2.1A) derived from the following: USA, INDIANA: Independence, 40.34˚ N, 18.17˚ W, ~170m above sea level, three day old culture of isolate UH3-1, originally isolated from the feces of a donkey (*Equus africanus asinus*), July 2016, Ethan Hillman.Ex-type strain: UH3-1. GenBank: ITS1-5.8S-ITS2 = KY494854

Etymology: The epithet honors the host organisms from which this fungus was isolated.

*A*n anaerobic fungus with a determinate (finite) life cycle displaying a monocentric thallus. The fungi exhibit endogenous zoosporangial development where the encysted zoospore retain the nucleus. The encysted zoospore geminates to form a rhizoidal network and a single oval or balloon-shaped sporangium (20-75 µm long and 20-30 µm wide), which on maturity liberates many zoospores. The rhizoidal system is devoid of nuclei (as seen under DAPI staining; Figure 2.1D-E)and is highly branched. Free swimming zoospores (Figure 2.1C) are typically spherical (10 µm diameter) and the species is characterized by the presence of a single posteriorly directed flagella(~30 µm long); the flagella propels the zoospore forward toward plant material/nutrient sources (chemotaxis)[150].

The clade is defined by the sequence KY494854 for ITS1-5.8S-ITS2

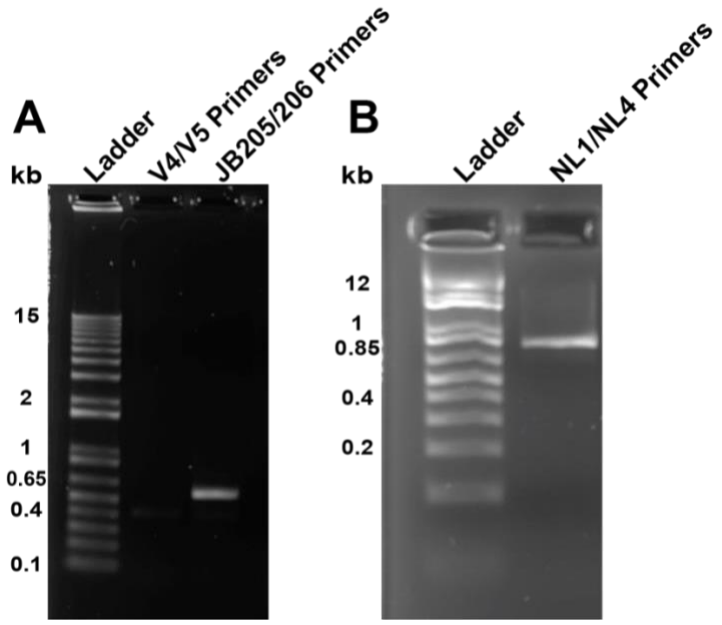


Figure B.1: *Piromyces sp. UH3-1* DNA controls: Lane 2 V4/V5 primers don't lead to amplification of *Piromyces sp. UH3-1* DNA, Lane 3 JB206/JB205 primers lead to amplificationof the ITS1 region of the *Piromyces sp. UH3-1* genome [14]

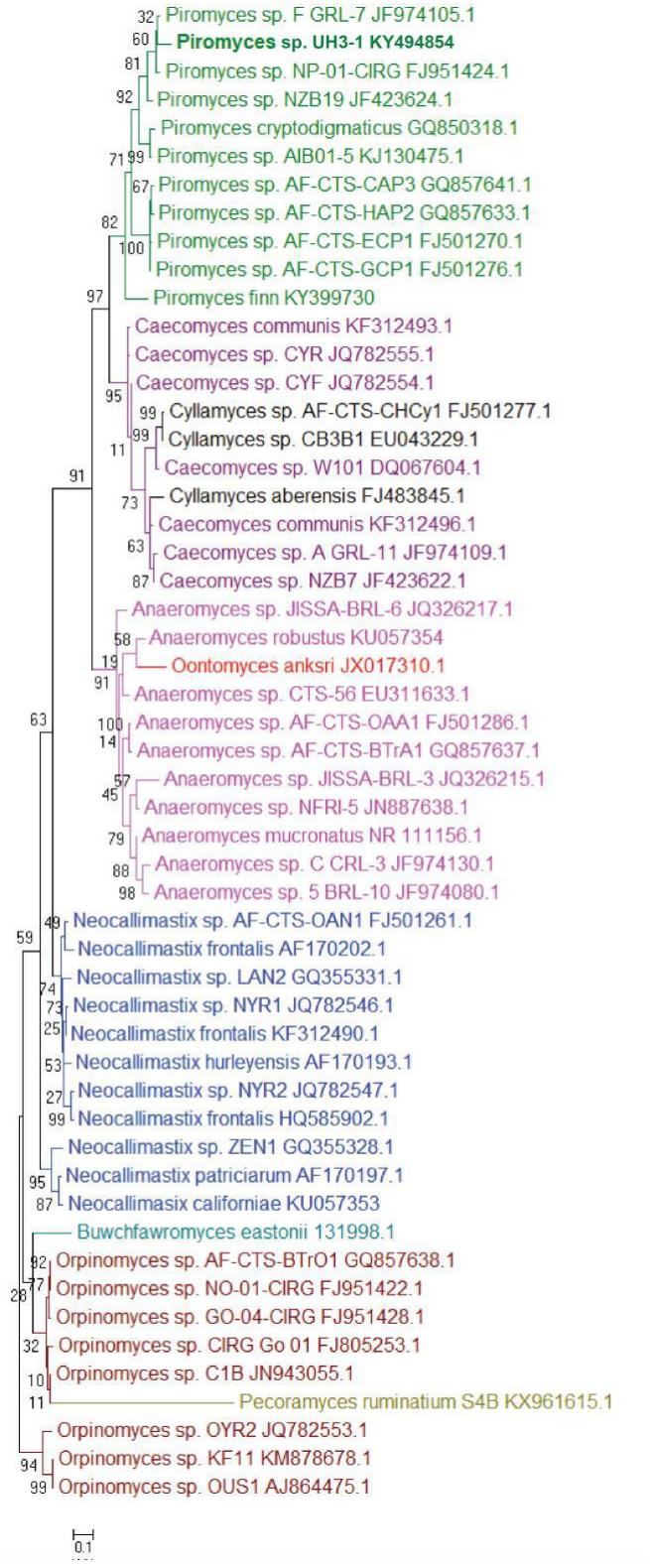


Figure B.2: Expanded *Piromyces sp. UH3-1* ITS1 phylogenetic tree with accession numbers [14].

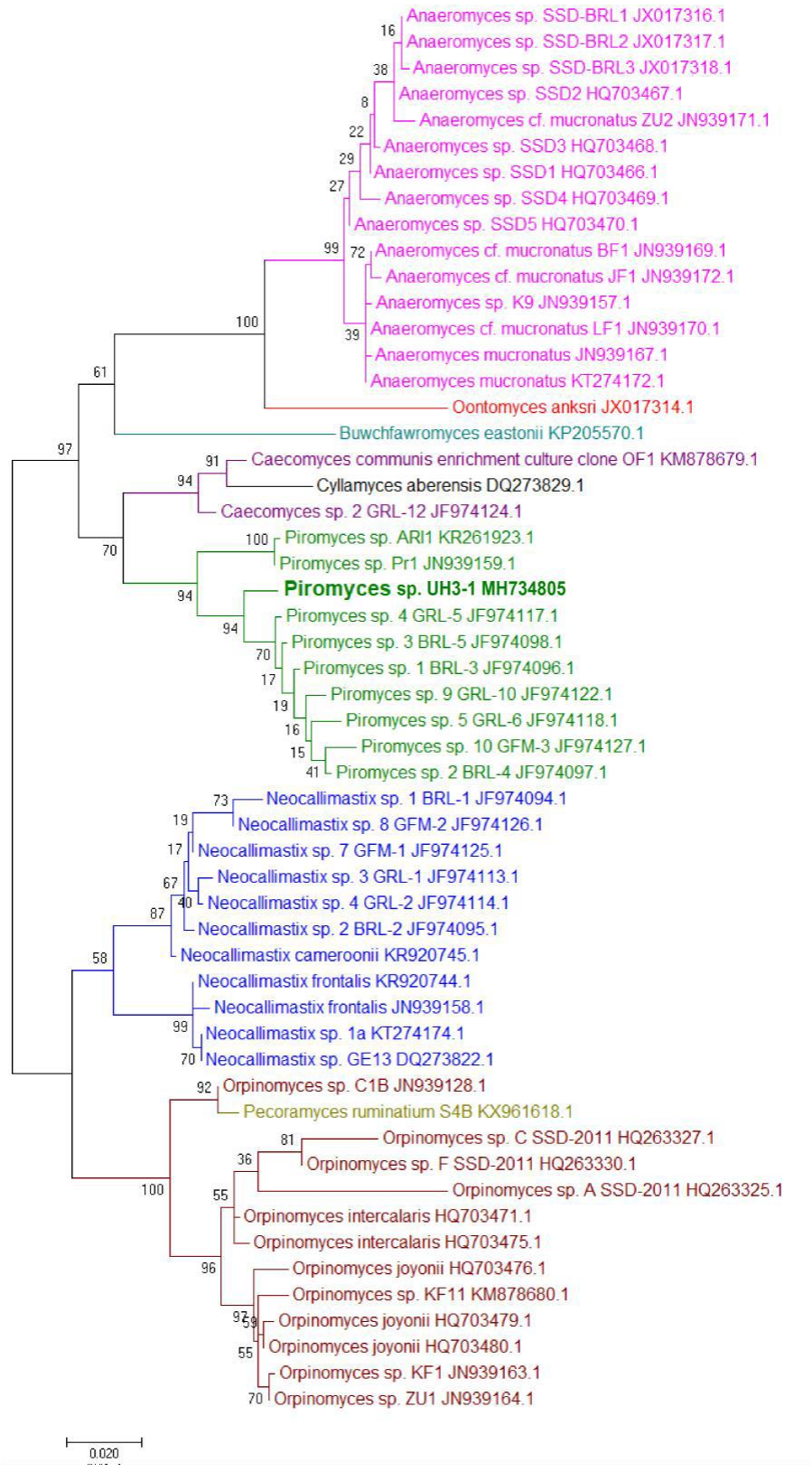


Figure B.3: Expanded *Piromyces sp. UH3-1* LSU phylogenetic tree with accession numbers [14]

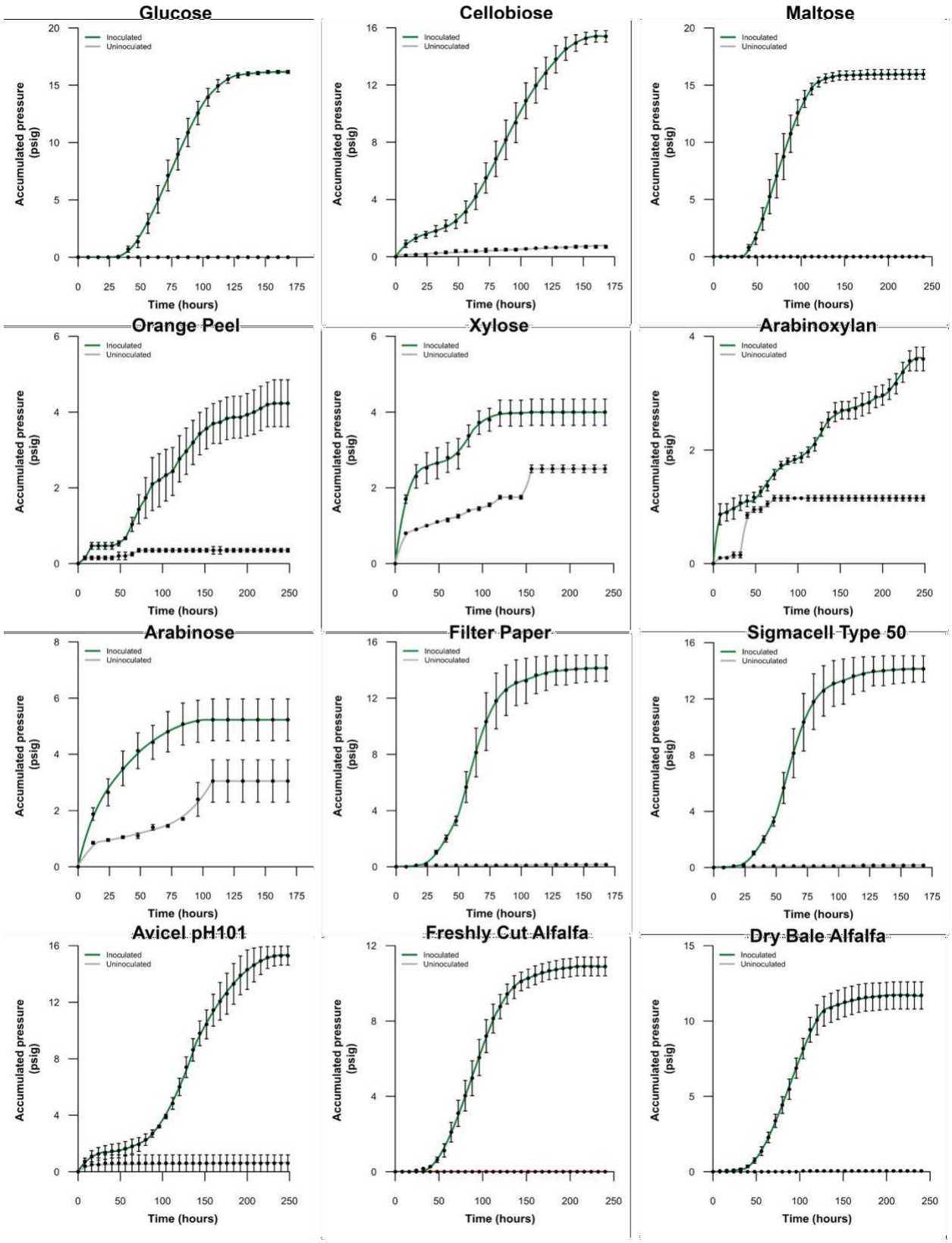


Figure B.4: Selected growth curves for *Piromyces sp. UH3-1* [14].

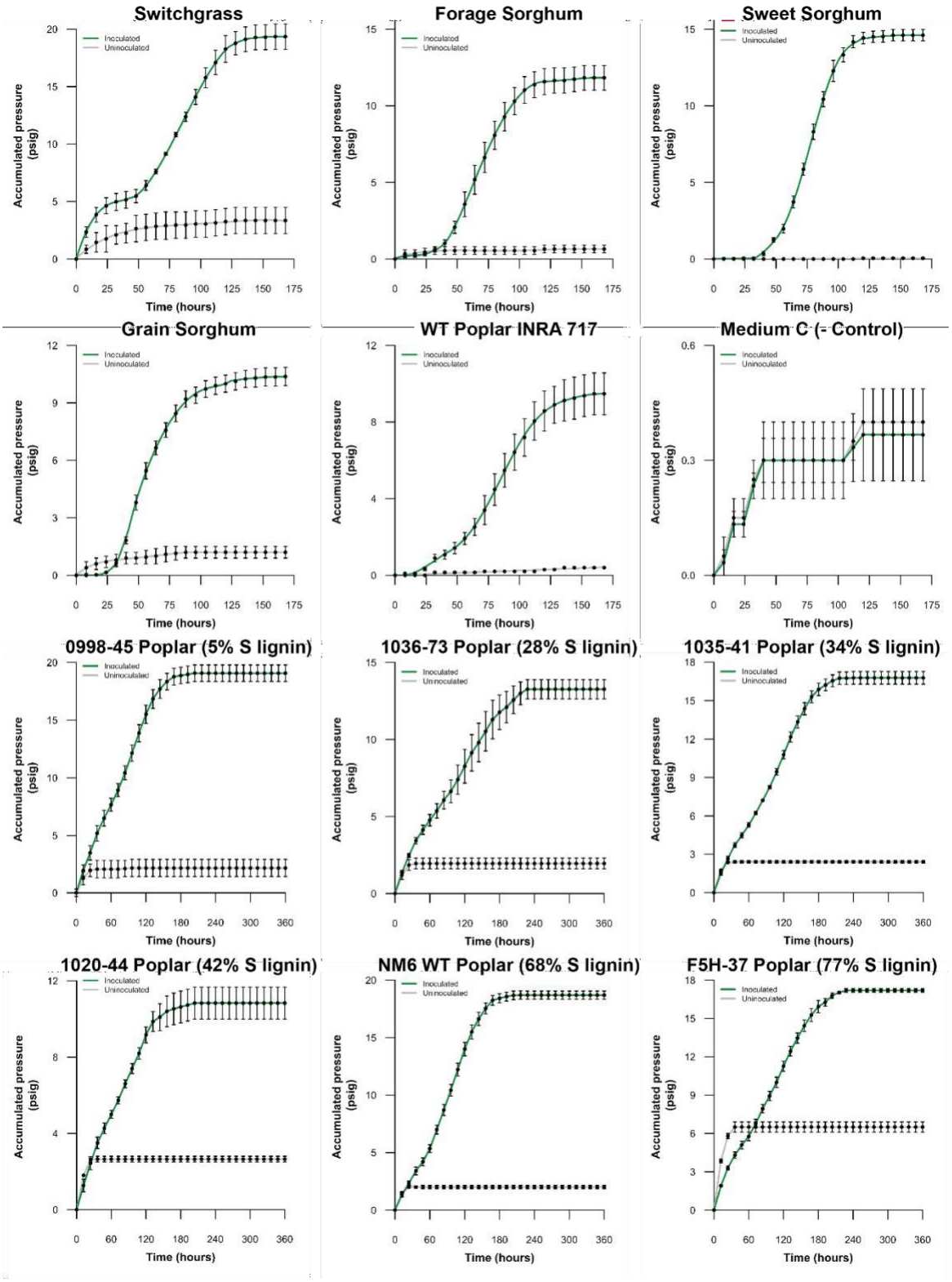


Figure B.5: Growth curves of *Piromyces sp. UH3-1* on genetically modified lines of poplar[14].

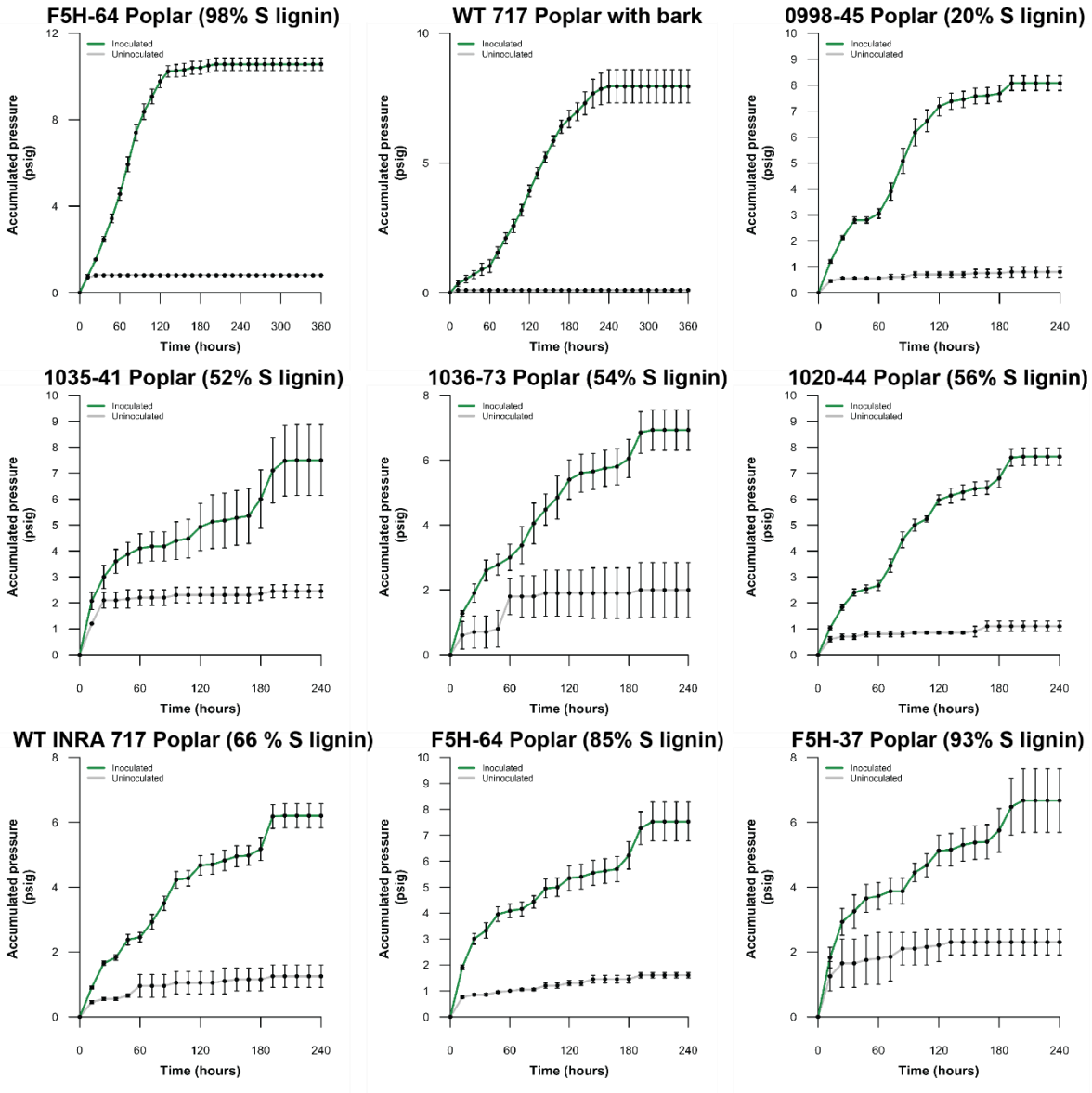


Figure B.6: Growth curves of *Piromyces sp. UH3-1* of genetically modified lines of poplar from the 2017 harvest [14].

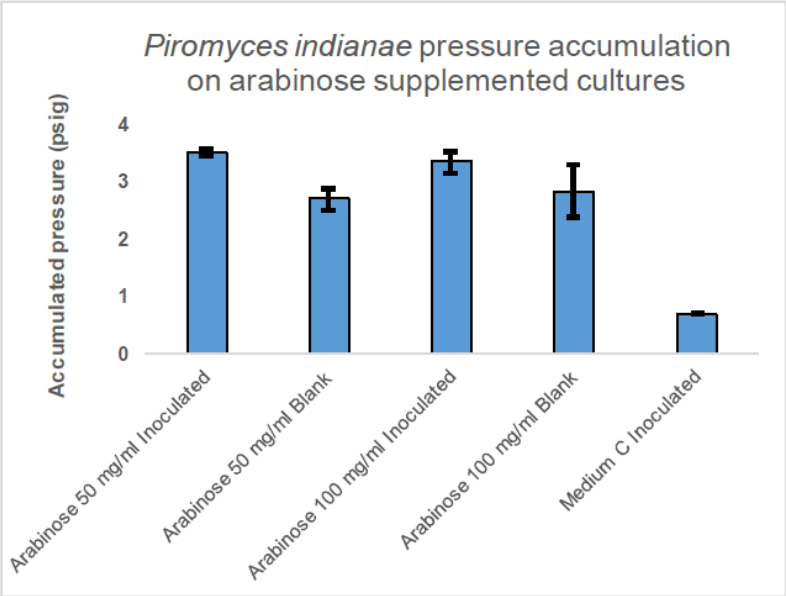


Figure B.7: *Piromyces sp. UH3-1* fails to reproducibly accumulate pressure on arabinosesupplemented cultures regardless of substrate loading.

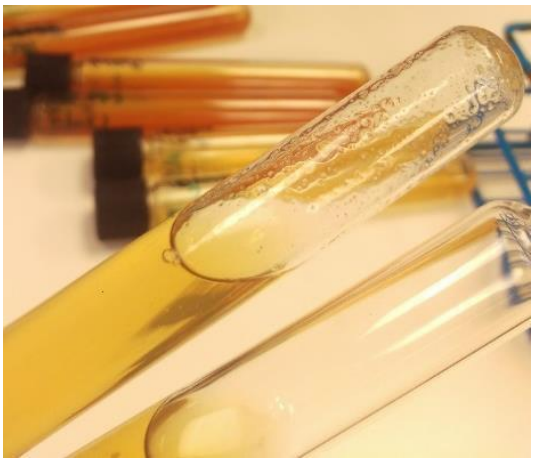


Figure B.8: *Piromyces sp. UH3-1* shows visible fungal biomass accumulation on media containing xylose. The top tube was inoculated and shows a high amount of fungal biomass,while the bottom tube was used as a negative control, and was not inoculated [14].

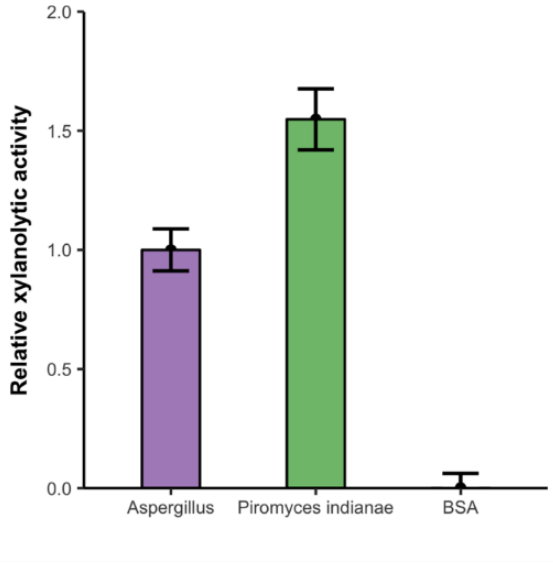


Figure B.9: *Piromyces sp. UH3-1* shows strong xylanolytic activity on xylan from beechwood at 50 °C, pH 7 for six hours of hydrolysis. Values normalized to Viscozyme (*Aspergillus spp.*) [14].

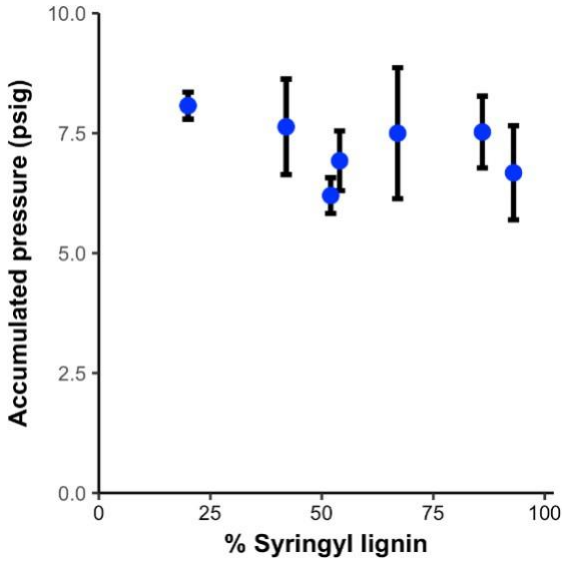


Figure B.10: *Piromyces sp. UH3-1* degrades untreated poplar to a similar extent on the 2017 harvested poplar having distinct S lignin molar ratios to those of the 2014 harvested lines. R^2=0.008, p = 0.647.

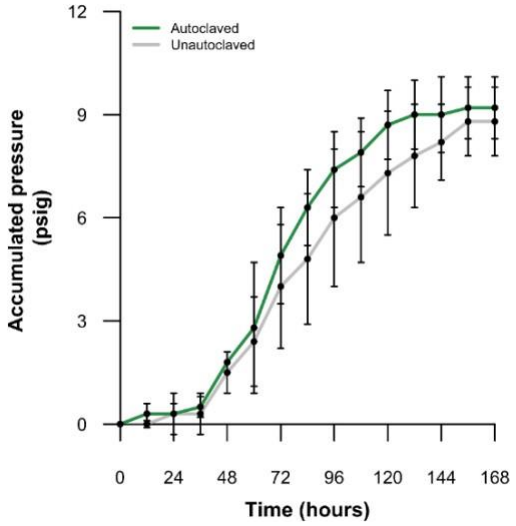


Figure B.11: Autoclaving corn stover at 120 °C for 30 minutes does not significantly enhance fungal growth rate or total accumulated pressure for *Piromyces sp. UH3-1* This autoclaved corn stover was not washed to remove any potential fermentation inhibitors that would be expected to reduce fungal growth. N=4 [14]

APPENDIX C. ADDITIONAL GENOME INFORMATION

Isolate: *Neocallimastix giraffae*

Etymology: The specific epithet refers to the giraffe; the animal from which the fungus was isolated.

Holotype: Neocallimastix sp. *Gf-Ma3-1* (Solomon Lab, Purdue University)

An obligate anaerobic fungus isolated from the feces of a giraffe (*Giraffa reticulata*) housed at the Indianapolis Zoo (www.indianapoliszoo.com) in 2017. The species is monocentric and has a determinate (finite) life cycle. The fungus exhibits endogenous zoosporangial development (i.e., the encysted zoospore retains the nucleus). The encysted zoospore germinates to form a rhizoidal system and a single typically spherical zoosporangium (30-100 µm diameter) that on maturity liberates up to 100 zoospores. The rhizoidal system is highly branched and devoid of nuclei (as seen under DAPI staining). The zoosporangium attaches to the rhizoidal system via one main sporangiophore. Free swimming zoospores are spherical (10 µm diameter) with multiple posteriorly directed flagella that are up to 30 µm in length.

The reference culture is maintained by continual passage at Purdue University (Gf-Ma3-1 – holotype), and under cryopreservation in repositories at the Solomon Lab, Purdue University and University of Jena and Leibniz Institute for Natural Product Research and Infection Biology, Jena, Germany.

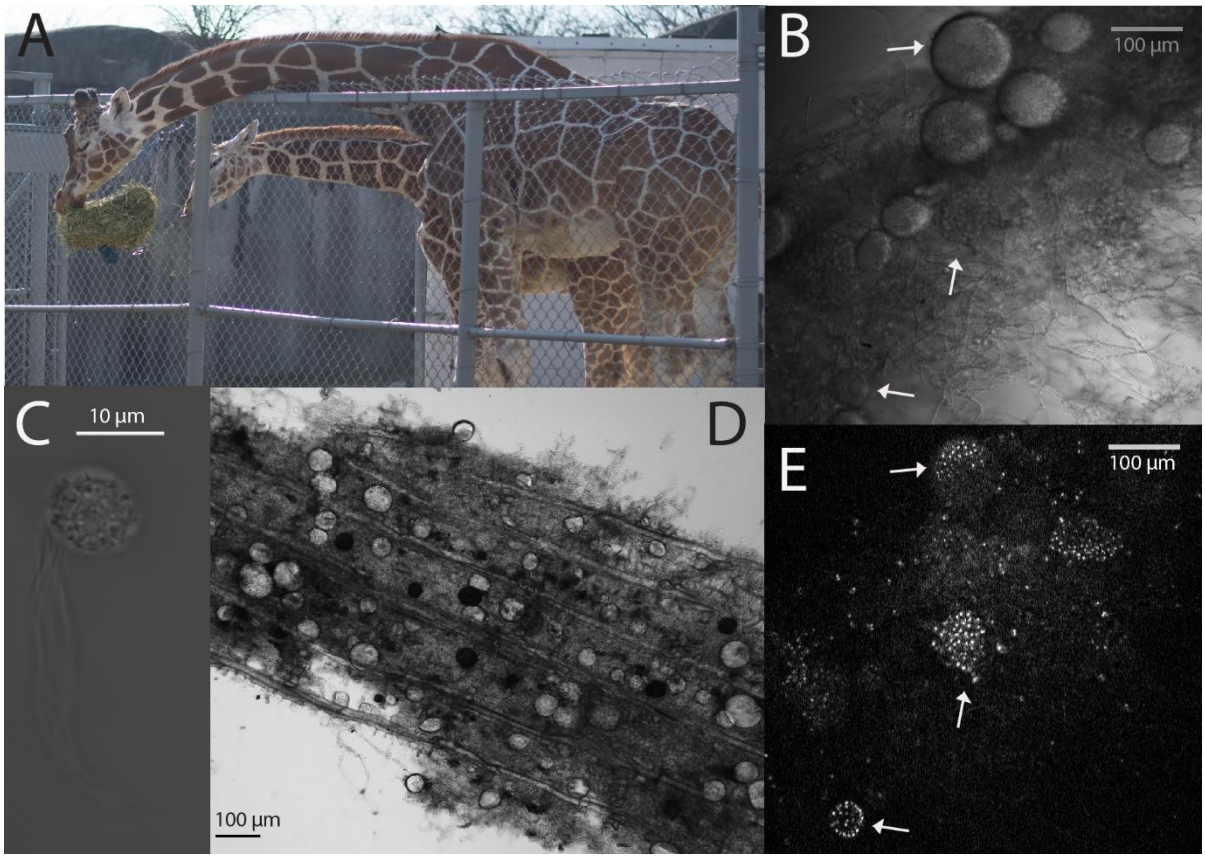


Figure C.1: Isolate images of *Neocallimastix sp. GfMa3-1*. A) Giraffe host that this species was isolated from. B) Mature spherical to ovoid, monocentric sporangia (arrows) and rhizomycelia of GfMa3-1 under transmitted light and E) DAPI stain. C) Zooflagellate zoospore. D) GfMa3-1 growing on corn stover biomass.

Isolate: *Neocallimastix stellae*

Etymology: The specific epithet refers to the name (stella) of the wildebeest; the name of animal from which the fungus was isolated.

Holotype: Neocallimastix sp. WI3-B (Solomon Lab, Purdue University)

An obligate anaerobic fungus isolated from the feces of a Wildebeest (*Giraffa reticulata*) housed at the Indianapolis Zoo (www.indianapoliszoo.com) in 2017. The species is monocentric and has a determinate (finite) life cycle. The fungus exhibits endogenous zoosporangial development (i.e., the encysted zoospore retains the nucleus). The encysted zoospore germinates to form a rhizoidal system and a single typically spherical zoosporangium (30-100 µm diameter) that on maturity liberates up to 100 zoospores. The rhizoidal system is highly branched and devoid of nuclei (as seen under DAPI staining). The zoosporangium attaches to the rhizoidal system via one main sporangiophore. Free swimming zoospores are spherical (10 µm diameter) with multiple posteriorly directed flagella that are up to 30 µm in length.

The reference culture is maintained by continual passage at Purdue University (Gf-Ma3-1 – holotype), and under cryopreservation in repositories at the Solomon Lab, Purdue University, and University of Jena and Leibniz Institute for Natural Product Research and Infection Biology, Jena, Germany

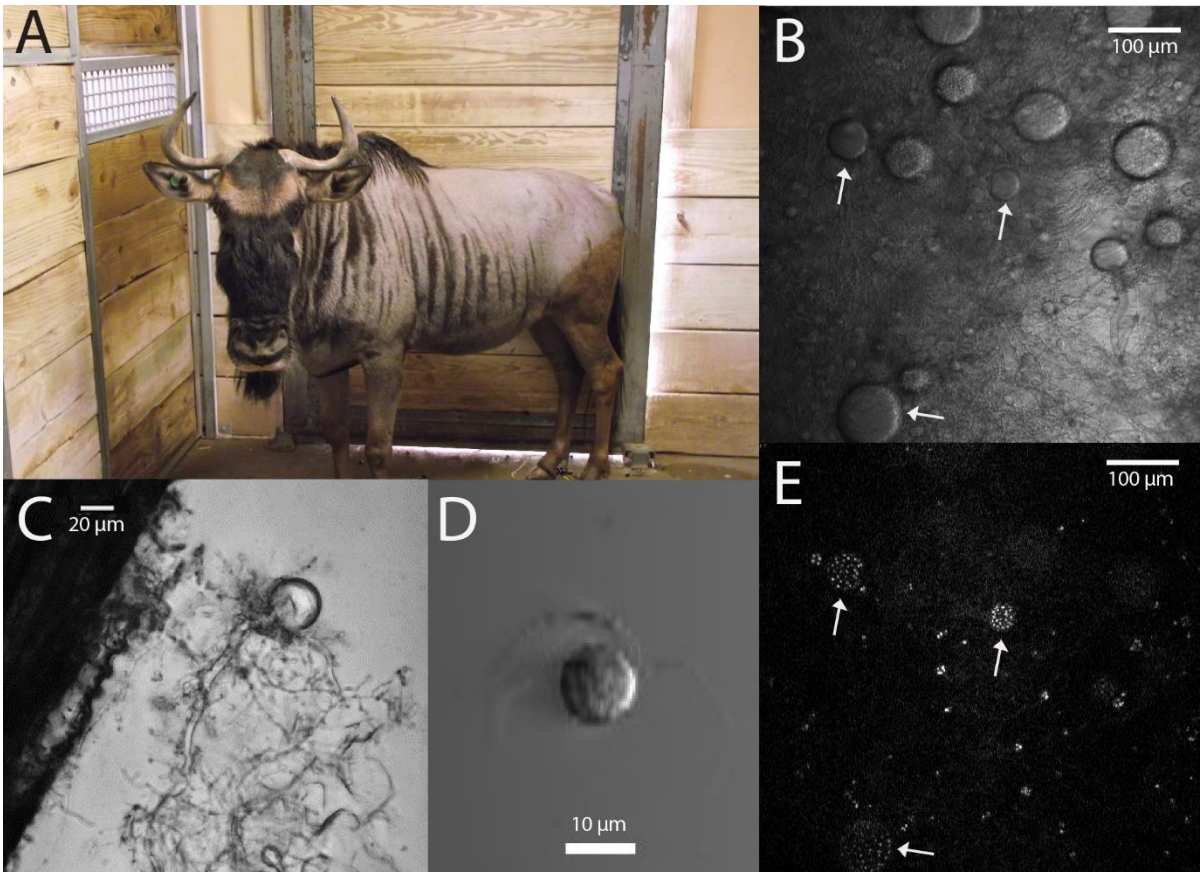


Figure C.2: Isolate images of *Neocallimastix sp. WI3-B*. A) Wildebeest host that this species was isolated from. B) Mature monocentric sporangia (arrows) and rhizomycelia of WI3-B under transmitted light and E) DAPI stain. C) GfMa3-1 growing on corn stover biomass. C) Zooflagellate zoospore.

Table C.1: Additional genome assembly stats

Isolate	# misjoins	Unscaffolded DNA	Chromosome sizes (Mb)
Piromyces sp. UH3-1	153	0.69 Mb (0.8%)	11.0, 10.9, 8.3, 8.2, 7.3, 7.0, 6.4, 5.2, 5.2, 4.9, 4.7, 4.2
Neocallimastix sp. GfMa3-1	264	0.69 MB (0.3%)	18.2, 17.4, 15.0, 14.9, 14.1, 14.0, 10.4, 9.6, 9.2, 8.7, 8.5, 8.4, 8.0, 7.5, 7.2, 6.9, 6.3, 5.7, 4.1, 4.0, 3.5, 3.5, 1.4, 0.7
Neocallimastix sp. WI3-B	188	0.10 Mb (>0.1%)	15.0, 15.0, 15.0, 14.7, 13.5, 12.2, 11.5, 10.1, 10.1, 8.4, 8.3, 8.2, 8.0, 7.2, 6.7, 6.7, 6.6, 5.6, 4.2, 4.0, 3.3, 3.1, 2.3, 0.8, 0.7

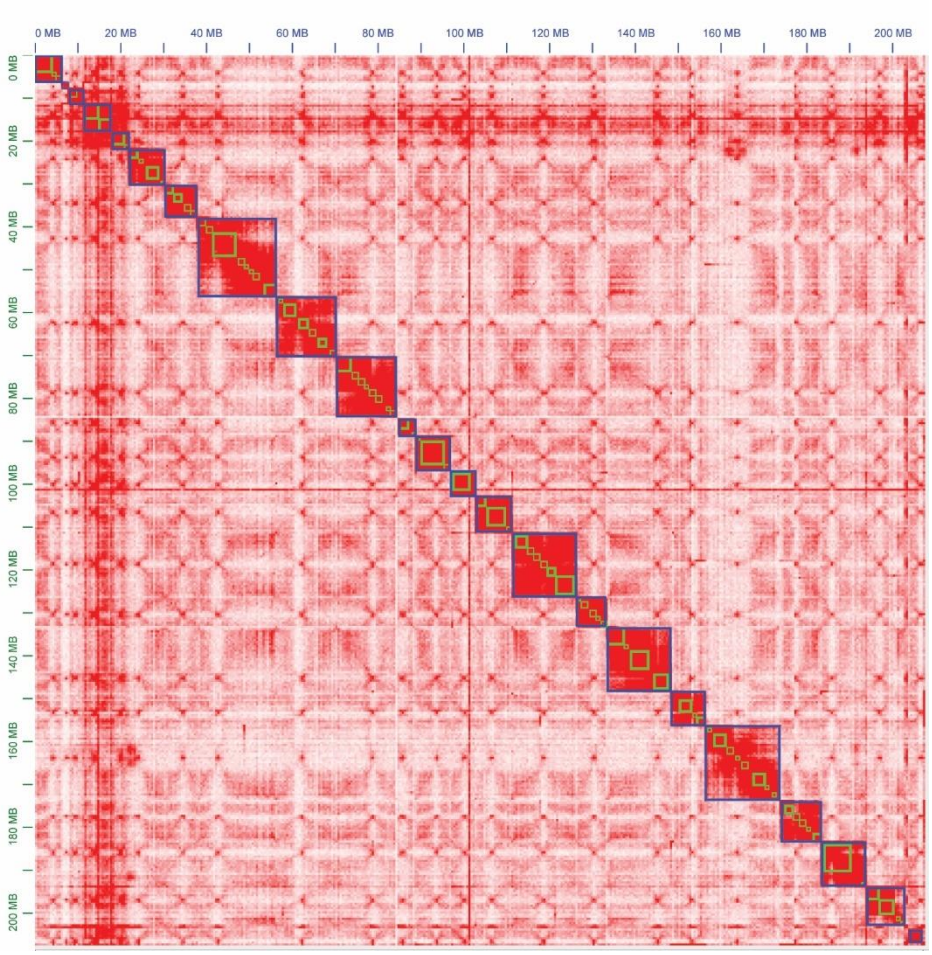


Figure C.3: Interaction plot of *Neocallimastix sp. GfMa3-1* assemblies**.** after HiC scaffolding. Tick marks indicate 20 Mbp. blue boxes represent scaffolds/chromosomes; green boxes represent contigs

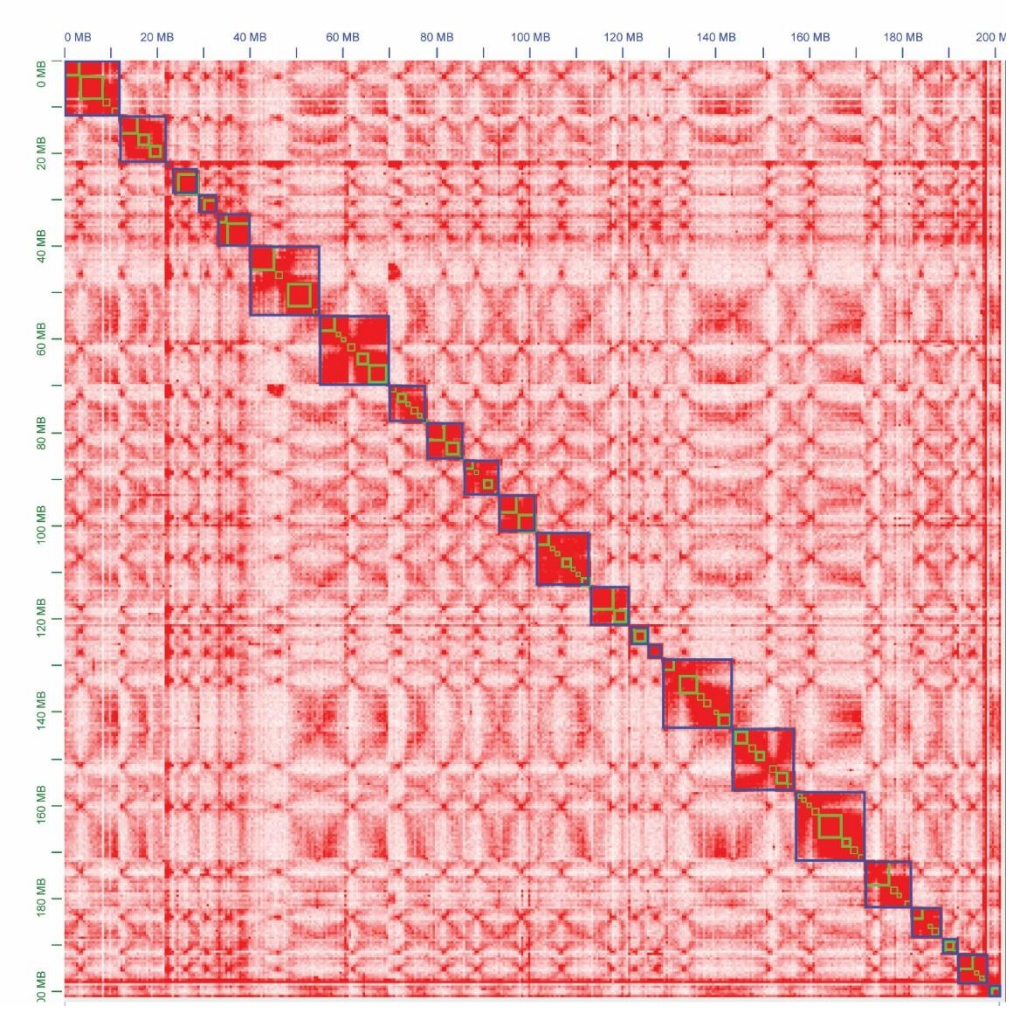


Figure C.4: Interaction plot of *Neocallimastix sp. WI3-B* assemblies**.** after HiC scaffolding. Tick marks indicate 20 Mbp. blue boxes represent scaffolds/chromosomes; green boxes represent contigs

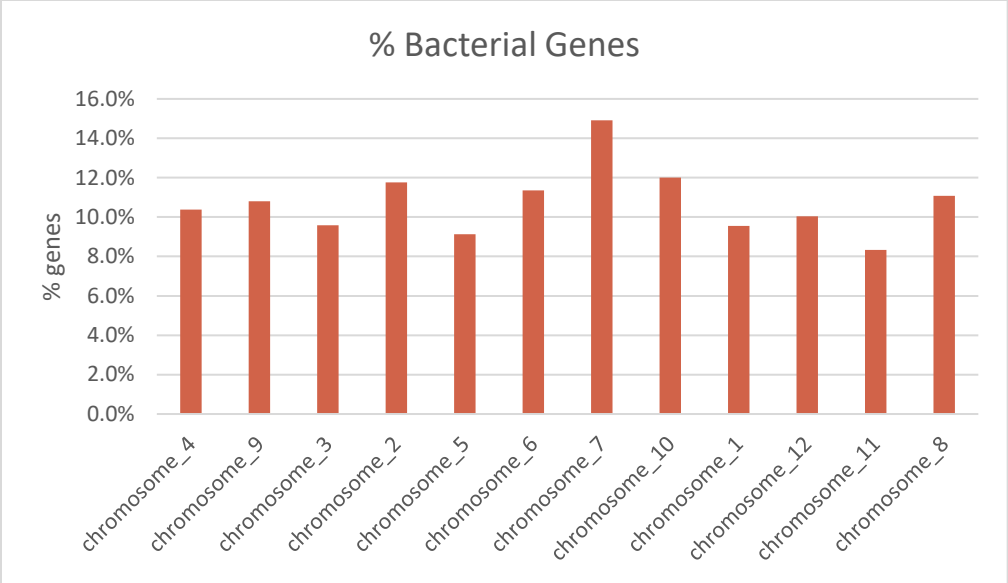


Figure C.5: Percentage of genes on each chromosome of the *Piromyces sp. UH3-1* genome whose best non-Neocallimastigomycota BLAST hit is of bacterial origin.

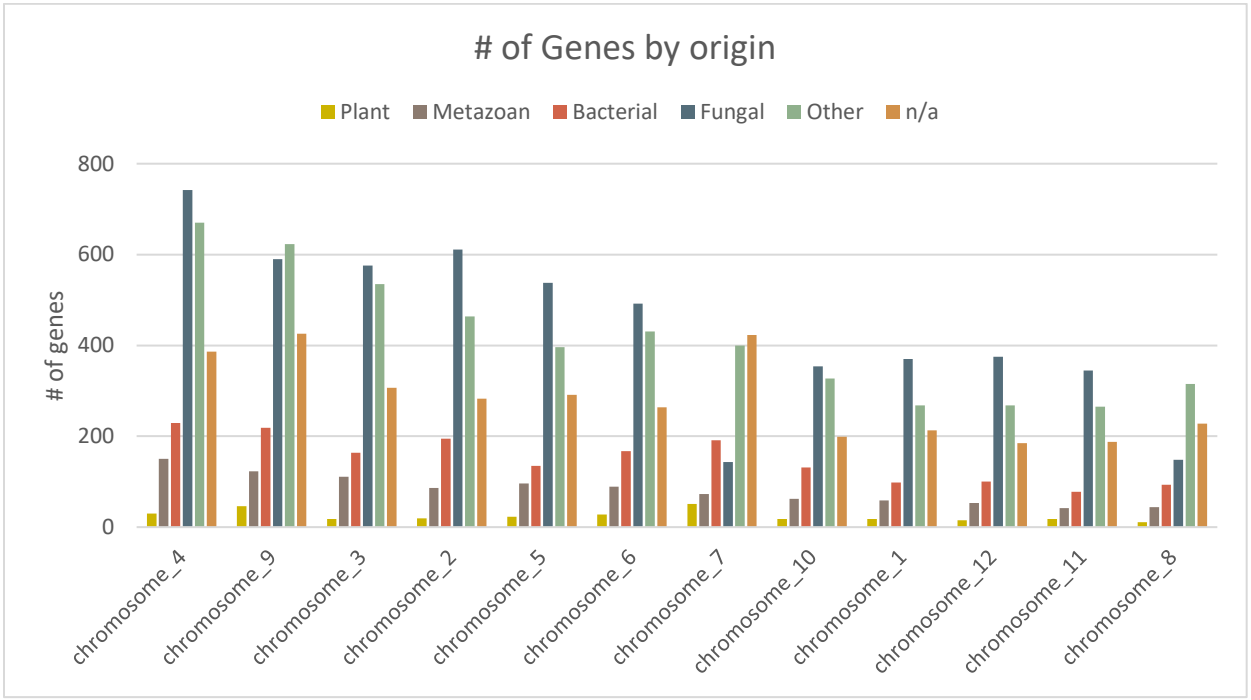


Figure C.6: Number of genes on each chromosome of the *Piromyces sp. UH3-1* genome whose best non-Neocallimastigomycota BLAST hit is belongs to Plant, Metazoan, Bacterial, Fungal, Other (misc Eukaryotes, Viruses, Archaea, etc), or n/a (only hit is to Neocallimastigomycota) .

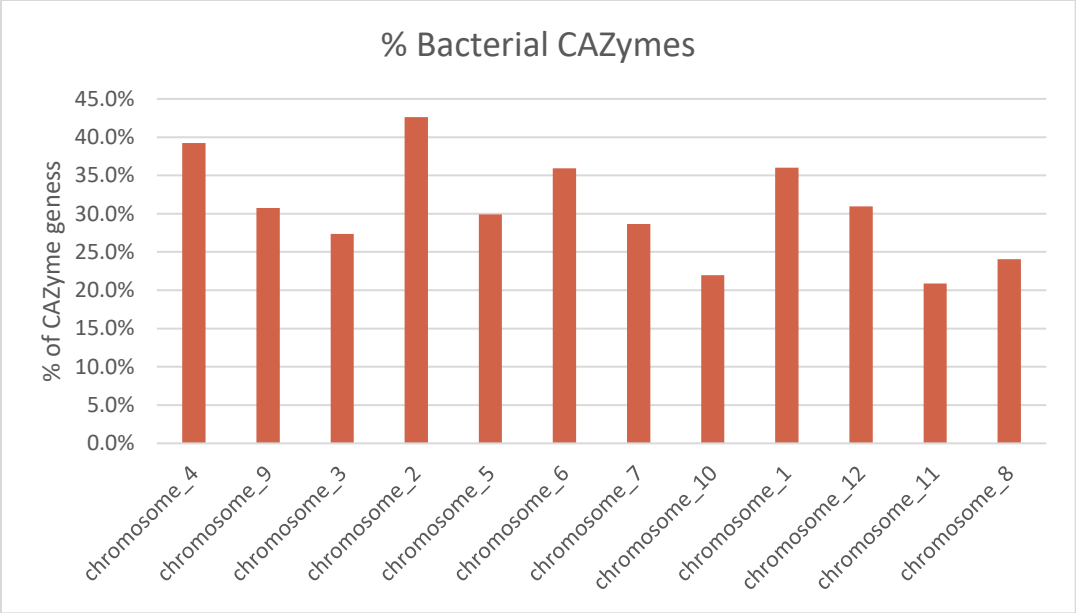


Figure C.7: Percentage of CAZyme genes on each chromosome of the *Piromyces sp. UH3-1* genome whose best non-Neocallimastigomycota BLAST hit is of bacterial origin.

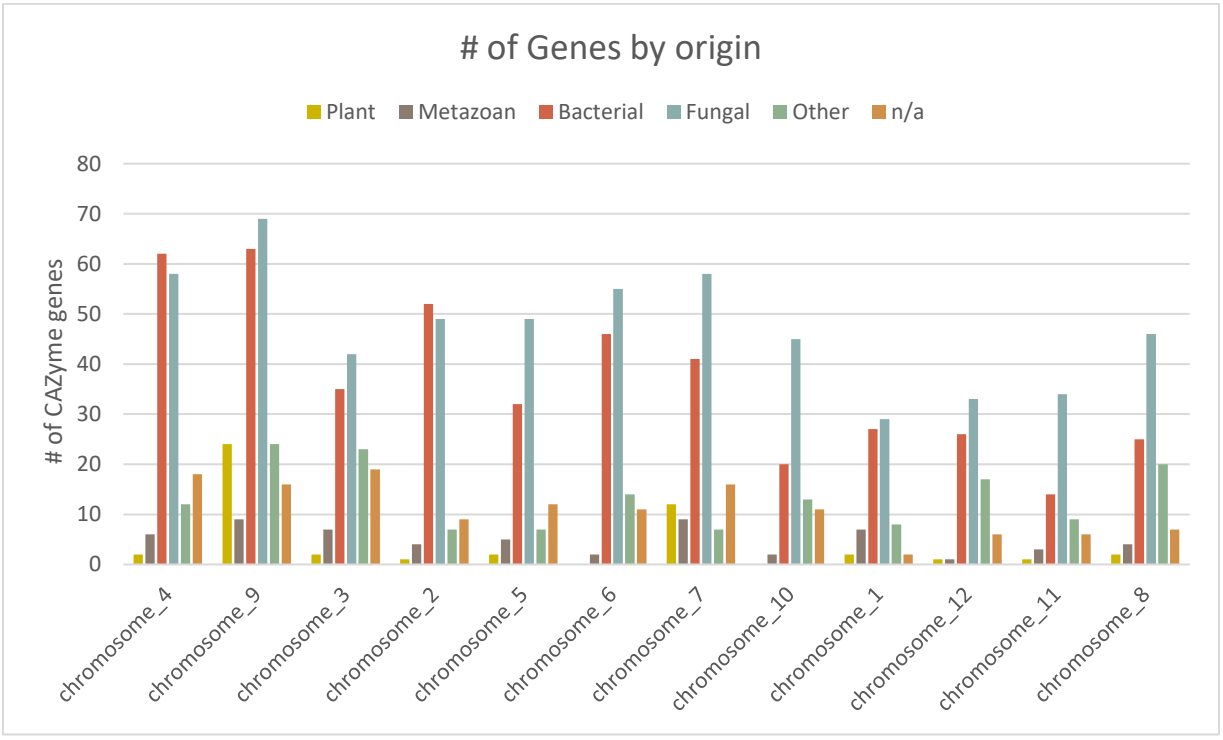
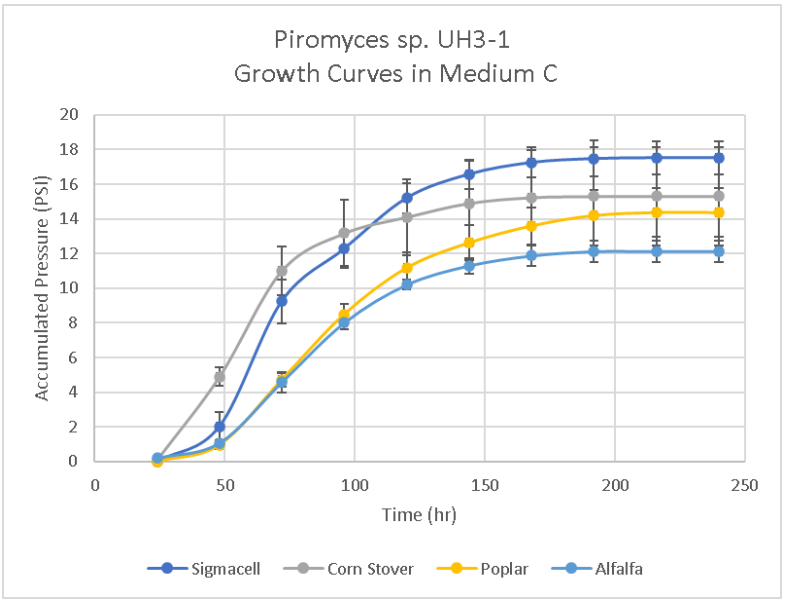


Figure C.8: Number of CAZyme genes on each chromosome of the *Piromyces sp. UH3-1* genome whose best non-Neocallimastigomycota BLAST hit is belongs to Plant, Metazoan, Bacterial, Fungal, Other (misc Eukaryotes, Viruses, Archaea, etc), or n/a (only hit is to Neocallimastigomycota) .

APPENDIX D. SUPPORTING INFORMATION FOR TWO STAGE PLATFORM



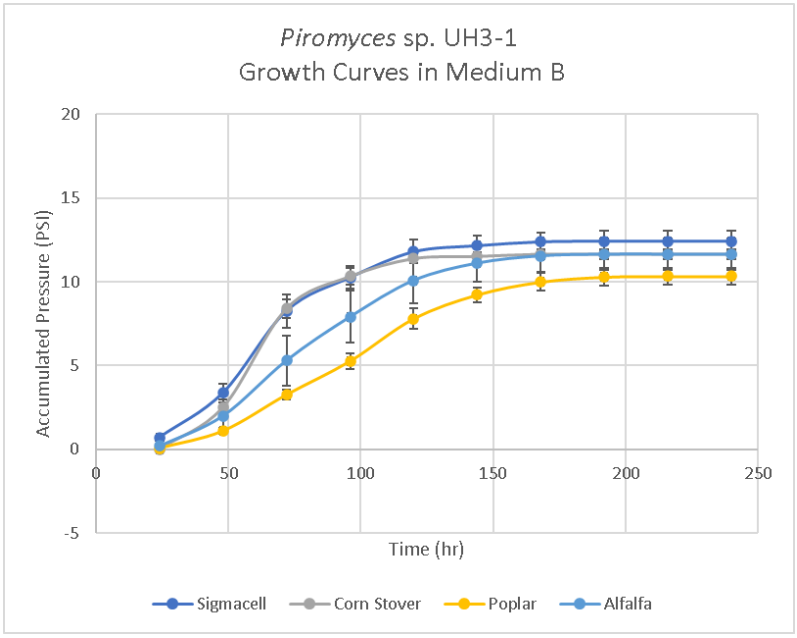


Figure D.1 – Growth curves of *P. indianae* on various feedstocks in Medium C (top) and Medium B (bottom). N=3, error bars = standard deviation.

Table D.1 – Stage 1 *P. indianae* conversion of feedstock to sugars, acids, and biomass across media types. Units are in milligrams (mg) except where noted otherwise. ND = not detected.

		Dry weight			% Utilized			Mass (mg) secreted in supernatant							Yields				
Substrate	**Media**	Start	End	% Lost	Glucan	Xylan	Arabinan	Glucose	Xylose	Arabinose	Lactate	Formate	Acetate	Ethanol	Fungal Biomass	Total sugars	g sugar / g biomass	Free Glucose yield	Free Xylose Yield
Sigmacell	*C*	600	69.85	91.9%	99.5%	98.3%	ND	292.89	11.33	ND	57.24	26.00	43.62	215.32	21.41	303.92	14.26	0.52	29.0%
Corn Stover	*C*	600	413.00	35.1%	67.1%	62.9%	76.6%	14.75	10.28	7.62	38.39	24.26	47.10	192.01	23.84	32.65	1.37	0.05	7.8%
Alfalfa	*C*	600	331.45	47.6%	72.1%	55.6%	89.6%	0.13	1.38	1.91	11.34	18.48	40.11	176.00	17.00	3.41	0.20	0.00	4.5%
Poplar	*C*	600	520.93	16.5%	57.4%	63.2%	ND	0.13	1.47	ND	24.32	20.91	45.14	178.82	19.66	2.17	0.11	0.00	1.0%

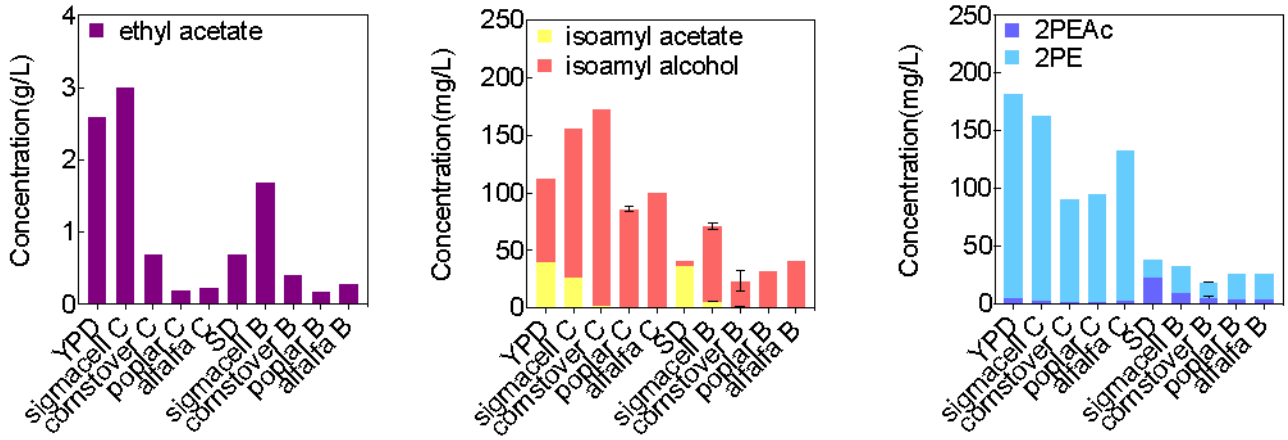


Figure D.2 – *K. marxianus* Growth and Production on spent Medium B**.** Titers of ethyl acetate (left), isoamyl alcohol/acetate (middle), and 2PE/2PEAc (right) for strain WT-u-h quantified by GC-FID.

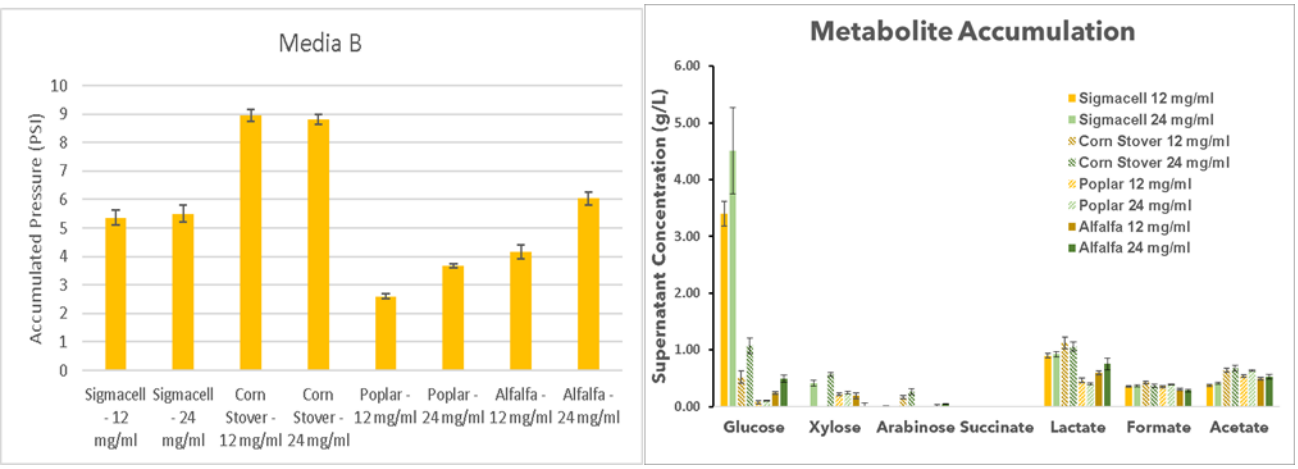


Figure D.3 – Substrate loading enhances hydrolysis in Medium B (Left) Growth of *P. indianae* on Medium B with 12 or 24 mg/ml substrate loading. (Right) Products of metabolism from growth of *P. indianae* with 12 mg/ml (yellow) and 24 mg/ml (green) loading of various substrates in Medium B.

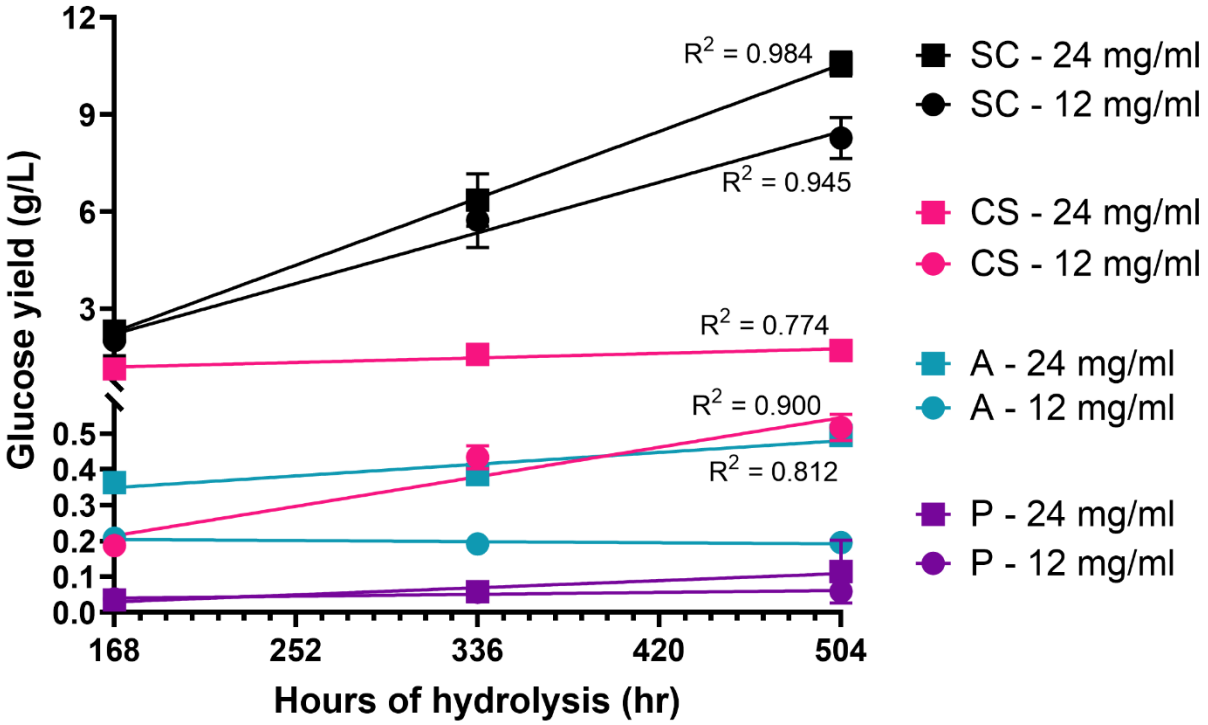


Figure D.4 - Rate of glucose hydrolysis with 12 or 24 mg/ml loading. Error bars represent the standard deviation or triplicate experiments. R^2 values greater than 0.75 are shown.

Table D.2 – Conversion Efficiency Tables (A) gram product yielded from the WT-u-h or ENGR *K. marxianus* strains per gram of feedstock loaded for hydrolysis. (B) Percent of theoretical yield achieved from each hydrolysate assuming that all feedstock glucose, xylose, and arabinose was converted into final product. (C) Ratio of actual product titers to theoretical yields if only the free sugars in the hydrolysate were converted to product. Bold values represent titers of products that the strain is engineered to produce in high amounts. Strike through values represent titers of products which have been knocked out in the strain.

A

g/g Biomass	WT-u-h				ENGR			
	SC	CS	P	A	SC	CS	P	A
EtAc	**0.205**	**0.059**	**0.016**	**0.012**	~~0.001~~	~~0.000~~	~~0.000~~	~~0.000~~
IsA	0.010	0.013	0.007	0.010	0.010	0.013	0.008	0.014
2-PE	0.013	0.008	0.008	0.011	**0.046**	**0.033**	**0.032**	**0.028**
2-PEAc	0.000	0.000	0.000	0.000	~~0.001~~	~~0.000~~	~~0.000~~	~~0.000~~
IsAc	0.001	0.000	0.000	0.000	~~0.000~~	~~0.000~~	~~0.000~~	~~0.000~~

B

% of theoretical yield	WT-u-h				ENGR			
	SC	CS	P	A	SC	CS	P	A
EtAc	**41.2%**	**16.0%**	**4.3%**	**10.4%**	~~0.1%~~	~~0.1%~~	~~0.1%~~	~~0.3%~~
IsA	2.1%	3.6%	2.0%	8.6%	2.0%	3.4%	2.0%	11.9%
2-PE	1.9%	1.5%	1.6%	7.1%	**6.7%**	**6.3%**	**6.2%**	**17.2%**

C

Actual:Theoretical Sugars only	WT-u-h				ENGR			
	SC	CS	P	A	SC	CS	P	A
EtAc	**0.55**	**1.14**	**6.18**	**2.22**	~~0.00~~	~~0.01~~	~~0.12~~	~~0.06~~
IsA	0.03	0.25	2.87	1.82	0.03	0.24	2.94	2.53
2-PE	0.03	0.18	2.19	1.07	**0.15**	**0.64**	**9.02**	**4.24**

APPENDIX E. CODON TABLES AND OTHER HETEROLOGOUS MEVALONATE PATHWAY INFORMATION

Table E.1 – Codon usage table for anaerobic fungus *Piromyces sp. UH3-1*. % Usage represents overall usage of the codon relative to the total of all codons used for all genes (coding sequences, CDS). Ratio = relative usage based on the frequency that a codon is used for the related amino acid.

	CODON	AMINO ACID	% USAGE	RATIO	CODON	AMINO ACID	% USAGE	RATIO	CODON	AMINO ACID	% USAGE	RATIO	CODON	AMINO ACID	% USAGE	RATIO	
U	UUU	Phe (F)	3.0%	0.71	UCU	Ser (S)	2.3%	0.28	UAU	Tyr (Y)	3.8%	0.82	UGU	Cys (C)	1.4%	0.86	U
	UUC	Phe (F)	1.1%	0.29	UCC	Ser (S)	0.7%	0.08	UAC	Tyr (Y)	0.8%	0.18	UGC	Cys (C)	0.2%	0.14	C
	UUA	Leu (L)	4.9%	0.60	UCA	Ser (S)	2.6%	0.30	UAA	STOP	0.1%	0.80	UGA	STOP	0.0%	0.09	A
	UUG	Leu (L)	0.7%	0.09	UCG	Ser (S)	0.2%	0.03	UAG	STOP	0.0%	0.11	UGG	Trp (W)	0.8%	1.00	G
C	CUU	Leu (L)	1.7%	0.22	CCU	Pro (P)	0.9%	0.27	CAU	His (H)	1.4%	0.86	CGU	Arg (R)	0.8%	0.26	U
	CUC	Leu (L)	0.2%	0.03	CCC	Pro (P)	0.1%	0.05	CAC	His (H)	0.2%	0.14	CGC	Arg (R)	0.0%	0.01	C
	CUA	Leu (L)	0.4%	0.05	CCA	Pro (P)	2.3%	0.65	CAA	Gln (Q)	3.0%	0.91	CGA	Arg (R)	0.2%	0.06	A
	CUG	Leu (L)	0.1%	0.01	CCG	Pro (P)	0.1%	0.03	CAG	Gln (Q)	0.3%	0.09	CGG	Arg (R)	0.0%	0.01	G
A	AUU	Ile (I)	5.4%	0.62	ACU	Thr (T)	2.6%	0.48	AAU	Asn (N)	9.0%	0.88	AGU	Ser (S)	2.3%	0.27	U
	AUC	Ile (I)	0.7%	0.09	ACC	Thr (T)	0.8%	0.15	AAC	Asn (N)	1.3%	0.12	AGC	Ser (S)	0.3%	0.03	C
	AUA	Ile (I)	2.7%	0.29	ACA	Thr (T)	1.8%	0.32	AAA	Lys (K)	7.0%	0.75	AGA	Arg (R)	1.9%	0.60	A
	AUG	Met (M)	2.0%	1.00	ACG	Thr (T)	0.2%	0.05	AAG	Lys (K)	2.0%	0.25	AGG	Arg (R)	0.2%	0.06	G
G	GUU	Val (V)	2.6%	0.57	GCU	Ala (A)	2.1%	0.55	GAU	Asp (D)	5.3%	0.90	GGU	Gly (G)	2.4%	0.58	U
	GUC	Val (V)	0.4%	0.09	GCC	Ala (A)	0.5%	0.14	GAC	Asp (D)	0.6%	0.10	GGC	Gly (G)	0.2%	0.05	C
	GUA	Val (V)	1.3%	0.28	GCA	Ala (A)	1.1%	0.28	GAA	Glu (E)	6.7%	0.91	GGA	Gly (G)	1.4%	0.34	A
	GUG	Val (V)	0.2%	0.06	GCG	Ala (A)	0.1%	0.03	GAG	Glu (E)	0.5%	0.09	GGG	Gly (G)	0.1%	0.03	G
	U				C				A				G				

Table E.2 – Codon usage table for *E. coli*. % Usage represents overall usage of the codon relative to the total of all codons used for all genes (coding sequences, CDS). Ratio = relative usage based on the frequency that a codon is used for the related amino acid.

	CODON	AMINO ACID	% USAGE	RATIO	CODON	AMINO ACID	% USAGE	RATIO	CODON	AMINO ACID	% USAGE	RATIO	CODON	AMINO ACID	% USAGE	RATIO	
U	UUU	Phe (F)	1.9%	0.51	UCU	Ser (S)	1.1%	0.19	UAU	Tyr (Y)	1.6%	0.53	UGU	Cys (C)	0.4%	0.43	U
	UUC	Phe (F)	1.8%	0.49	UCC	Ser (S)	1.0%	0.17	UAC	Tyr (Y)	1.4%	0.47	UGC	Cys (C)	0.6%	0.57	C
	UUA	Leu (L)	1.0%	0.11	UCA	Ser (S)	0.7%	0.12	UAA	STOP	0.2%	0.62	UGA	STOP	0.1%	0.30	A
	UUG	Leu (L)	1.1%	0.11	UCG	Ser (S)	0.8%	0.13	UAG	STOP	0.03%	0.09	UGG	Trp (W)	1.4%	1.00	G
C	CUU	Leu (L)	1.0%	0.10	CCU	Pro (P)	0.7%	0.16	CAU	His (H)	1.2%	0.52	CGU	Arg (R)	2.4%	0.42	U
	CUC	Leu (L)	0.9%	0.10	CCC	Pro (P)	0.4%	0.10	CAC	His (H)	1.1%	0.48	CGC	Arg (R)	2.2%	0.37	C
	CUA	Leu (L)	0.3%	0.03	CCA	Pro (P)	0.8%	0.20	CAA	Gln (Q)	1.3%	0.31	CGA	Arg (R)	0.3%	0.05	A
	CUG	Leu (L)	5.2%	0.55	CCG	Pro (P)	2.4%	0.55	CAG	Gln (Q)	2.9%	0.69	CGG	Arg (R)	0.5%	0.08	G
A	AUU	Ile (I)	2.7%	0.47	ACU	Thr (T)	1.2%	0.21	AAU	Asn (N)	1.6%	0.39	AGU	Ser (S)	0.7%	0.13	U
	AUC	Ile (I)	2.7%	0.46	ACC	Thr (T)	1.6%	0.43	AAC	Asn (N)	2.6%	0.61	AGC	Ser (S)	1.5%	0.27	C
	AUA	Ile (I)	0.4%	0.07	ACA	Thr (T)	1.4%	0.30	AAA	Lys (K)	3.8%	0.76	AGA	Arg (R)	0.2%	0.04	A
	AUG	Met (M)	2.6%	1.00	ACG	Thr (T)	1.3%	0.23	AAG	Lys (K)	1.2%	0.24	AGG	Arg (R)	0.2%	0.03	G
G	GUU	Val (V)	2.0%	0.29	GCU	Ala (A)	1.8%	0.19	GAU	Asp (D)	3.3%	0.59	GGU	Gly (G)	2.8%	0.38	U
	GUC	Val (V)	1.4%	0.20	GCC	Ala (A)	2.3%	0.25	GAC	Asp (D)	2.3%	0.41	GGC	Gly (G)	3.0%	0.40	C
	GUA	Val (V)	1.2%	0.17	GCA	Ala (A)	2.1%	0.22	GAA	Glu (E)	4.4%	0.70	GGA	Gly (G)	0.7%	0.09	A
	GUG	Val (V)	2.4%	0.34	GCG	Ala (A)	3.2%	0.34	GAG	Glu (E)	1.9%	0.30	GGG	Gly (G)	0.9%	0.13	G
	U				C				A				G				

Table E.3 – Codon usage table for *S. cerevisiae*. % Usage represents overall usage of the codon relative to the total of all codons used for all genes (coding sequences, CDS). Ratio = relative usage based on the frequency that a codon is used for the related amino acid.

	CODON	AMINO ACID	% USAGE	RATIO	CODON	AMINO ACID	% USAGE	RATIO	CODON	AMINO ACID	% USAGE	RATIO	CODON	AMINO ACID	% USAGE	RATIO	
U	UUU	Phe (F)	2.6%	0.59	UCU	Ser (S)	2.3%	0.26	UAU	Tyr (Y)	1.9%	0.56	UGU	Cys (C)	0.8%	0.63	U
	UUC	Phe (F)	1.8%	0.41	UCC	Ser (S)	1.4%	0.16	UAC	Tyr (Y)	1.5%	0.44	UGC	Cys (C)	0.5%	0.37	C
	UUA	Leu (L)	2.6%	0.28	UCA	Ser (S)	1.9%	0.21	UAA	STOP	0.1%	0.47	UGA	STOP	0.1%	0.30	A
	UUG	Leu (L)	2.7%	0.29	UCG	Ser (S)	0.9%	0.10	UAG	STOP	0.05%	0.23	UGG	Trp (W)	1.0%	1.00	G
C	CUU	Leu (L)	1.2%	0.13	CCU	Pro (P)	1.4%	0.31	CAU	His (H)	1.4%	0.64	CGU	Arg (R)	0.6%	0.14	U
	CUC	Leu (L)	0.5%	0.06	CCC	Pro (P)	0.7%	0.15	CAC	His (H)	0.8%	0.36	CGC	Arg (R)	0.3%	0.06	C
	CUA	Leu (L)	1.3%	0.14	CCA	Pro (P)	1.8%	0.42	CAA	Gln (Q)	2.7%	0.69	CGA	Arg (R)	0.3%	0.07	A
	CUG	Leu (L)	1.0%	0.11	CCG	Pro (P)	0.5%	0.12	CAG	Gln (Q)	1.2%	0.31	CGG	Arg (R)	0.2%	0.04	G
A	AUU	Ile (I)	3.0%	0.46	ACU	Thr (T)	2.0%	0.35	AAU	Asn (N)	3.6%	0.59	AGU	Ser (S)	1.4%	0.16	U
	AUC	Ile (I)	1.7%	0.26	ACC	Thr (T)	1.3%	0.22	AAC	Asn (N)	2.5%	0.41	AGC	Ser (S)	1.0%	0.11	C
	AUA	Ile (I)	1.8%	0.27	ACA	Thr (T)	1.8%	0.30	AAA	Lys (K)	4.2%	0.58	AGA	Arg (R)	2.1%	0.48	A
	AUG	Met (M)	2.1%	1.00	ACG	Thr (T)	0.8%	0.14	AAG	Lys (K)	3.1%	0.42	AGG	Arg (R)	0.9%	0.21	G
G	GUU	Val (V)	2.2%	0.39	GCU	Ala (A)	2.1%	0.38	GAU	Asp (D)	3.8%	0.65	GGU	Gly (G)	2.4%	0.47	U
	GUC	Val (V)	1.2%	0.21	GCC	Ala (A)	1.3%	0.22	GAC	Asp (D)	2.0%	0.16	GGC	Gly (G)	1.0%	0.19	C
	GUA	Val (V)	1.2%	0.21	GCA	Ala (A)	1.6%	0.29	GAA	Glu (E)	4.6%	0.70	GGA	Gly (G)	1.1%	0.22	A
	GUG	Val (V)	1.1%	0.19	GCG	Ala (A)	0.6%	0.11	GAG	Glu (E)	1.9%	0.30	GGG	Gly (G)	0.6%	0.12	G
	U				C				A				G				

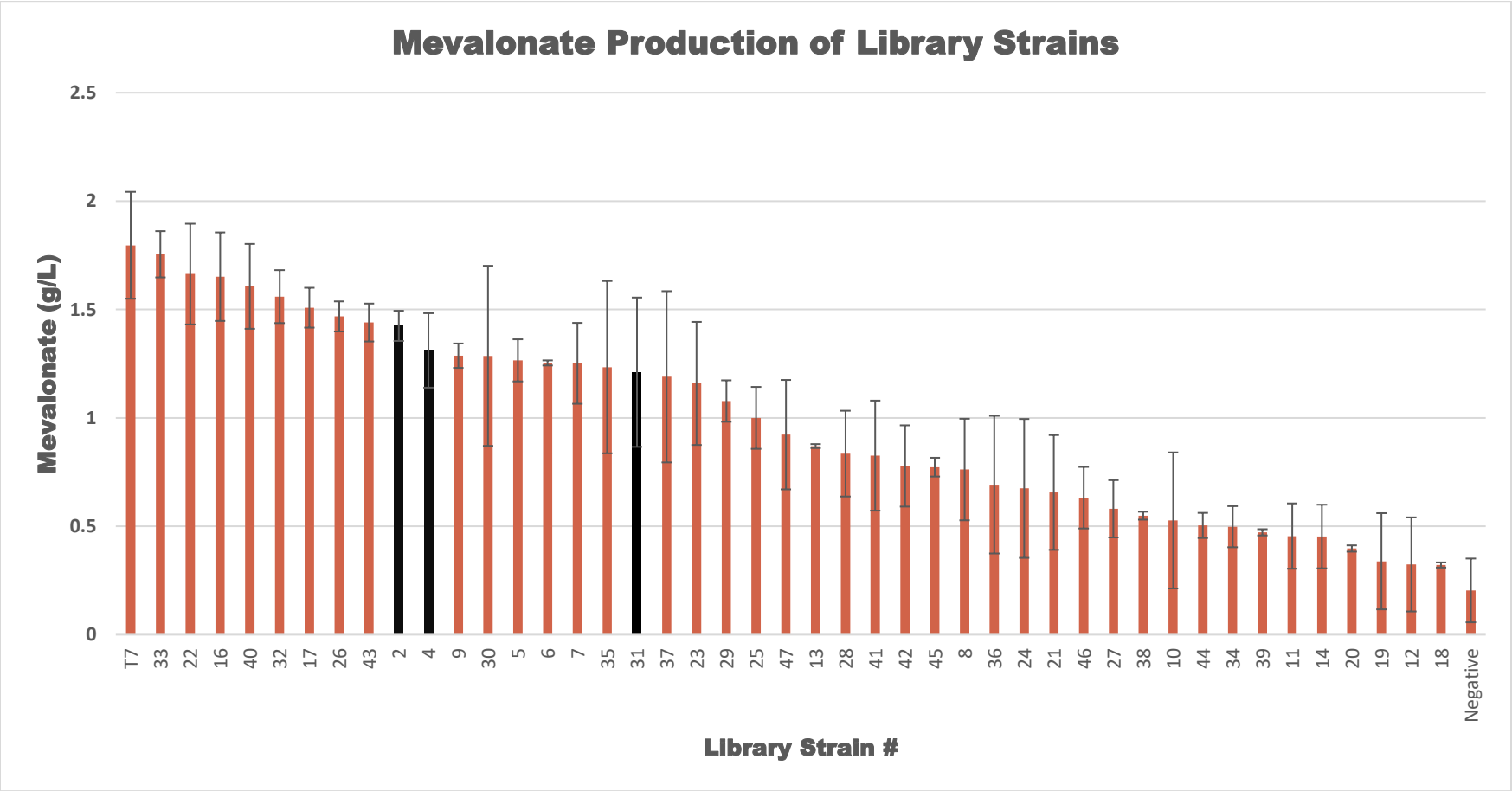


Figure E.1: Mevalonate production of various library strains with varied promoter organization for the full yeast pathway. Black bars show constructs with the *atoB*$_{H9}$-*HMGS*$_{C4}$-*HMGR*$_{H9}$ – these constructs had reasonably high production levels and were common (at least three occurrences) which made them promising constructs to work from. Error bars = standard deviation.

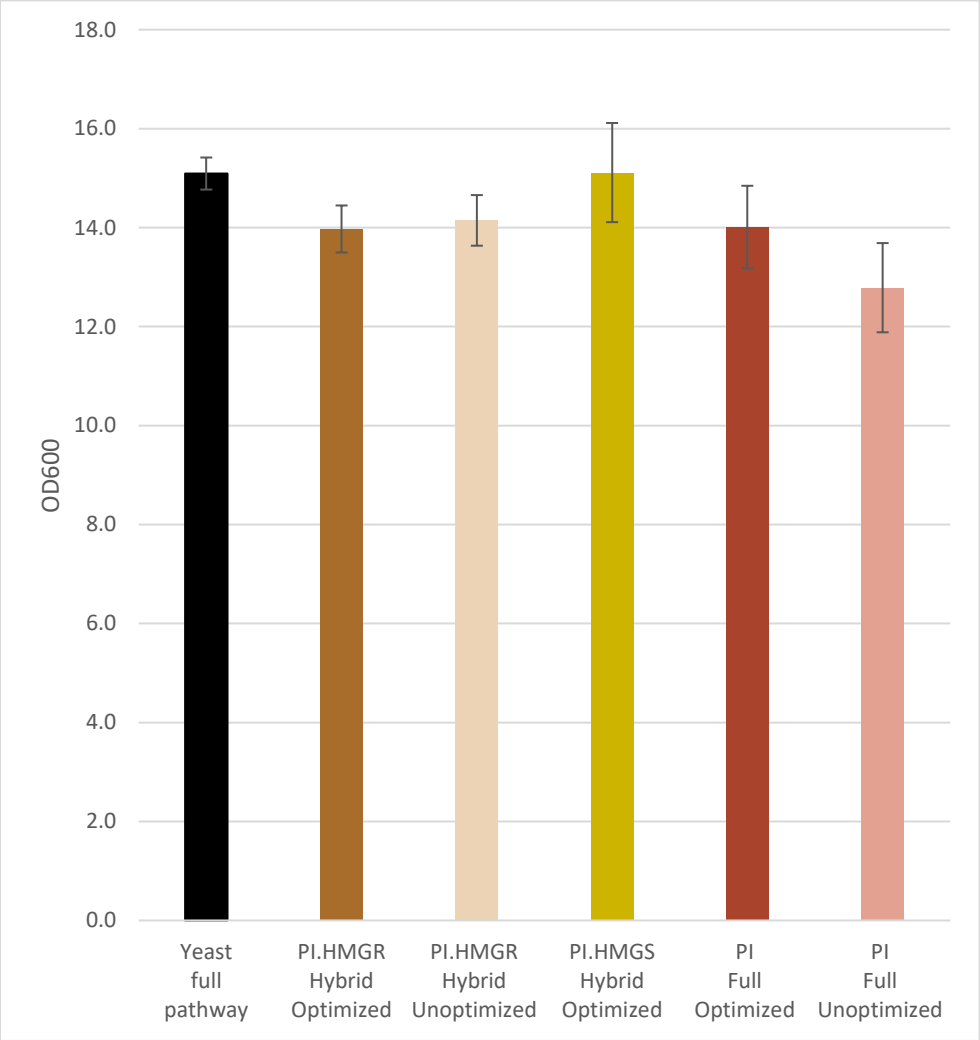


Figure E.2: OD of 50 ml cultures containing various mevalonate pathways after 20 hours of growth. Yeast pathway = the *S. cerevisiae* from Martin et al [236]. Full PI pathways = genes form *P. indianae* (Figure 3B) either native or optimized for E. coli. Hybrid pathways use the yeast construct and swap the indicated gene for the original yeast homolog. All pathways evaluated with $atoB_{H9}$-$HMGS_{C4}$-$HMGR_{H9}$ promoter organization. Error bars = standard deviation.

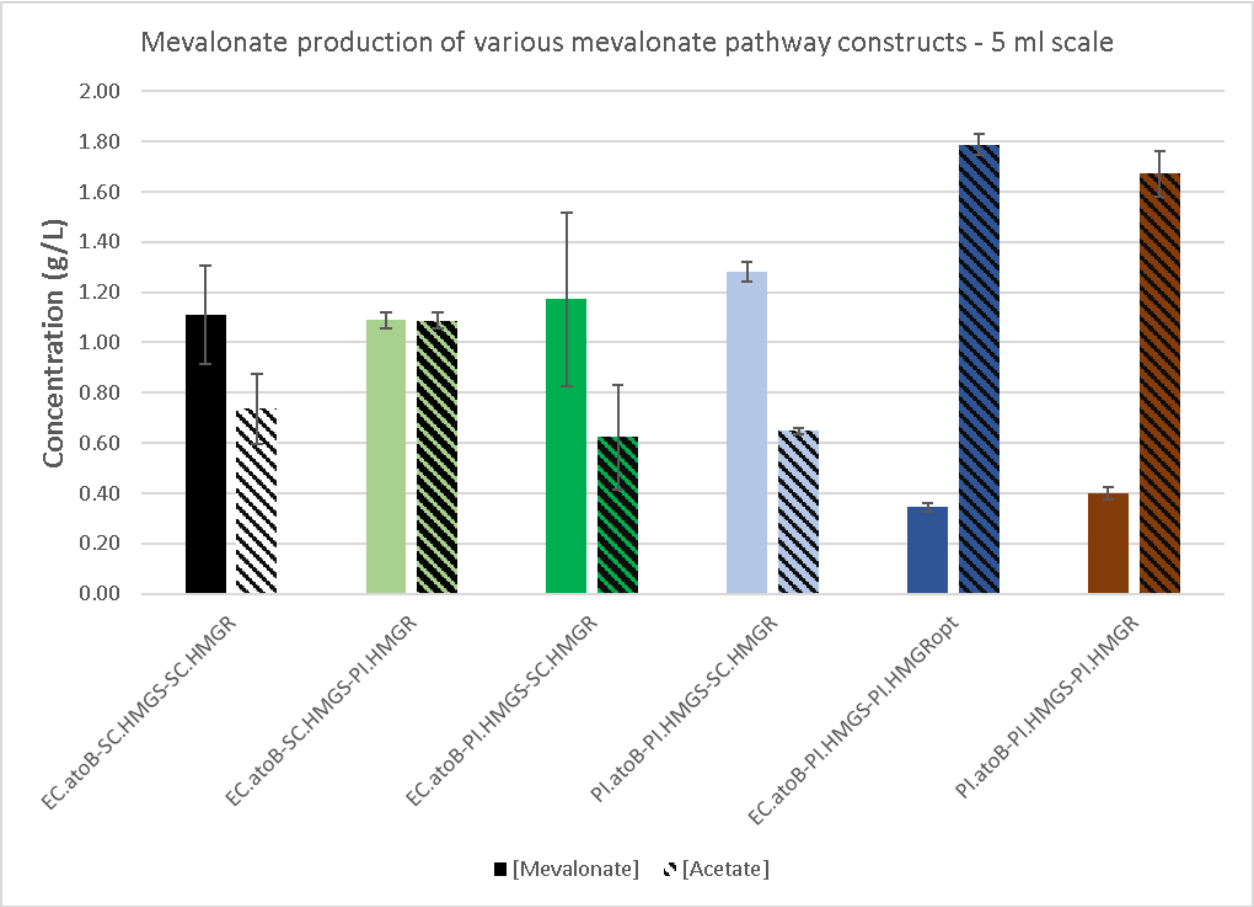


Figure E.3: Mevalonate and acetate titers of original and hybrid mevalonate pathways.at 5 ml scale. Mevalonate production (solid) and acetate accumulation (striped) from various mevalonate pathway hybrids containing Martin et al (black) or *E. coli*-codon-optimized *P. indianae* genes (various colors, see Figure 6-7) after 20 hrs of culture. All pathways are configured in the high-producing h9-promoter configuration. Errors bars represent standard deviation.

APPENDIX F. ADDITIONAL GENETIC TOOLS INFORMATION

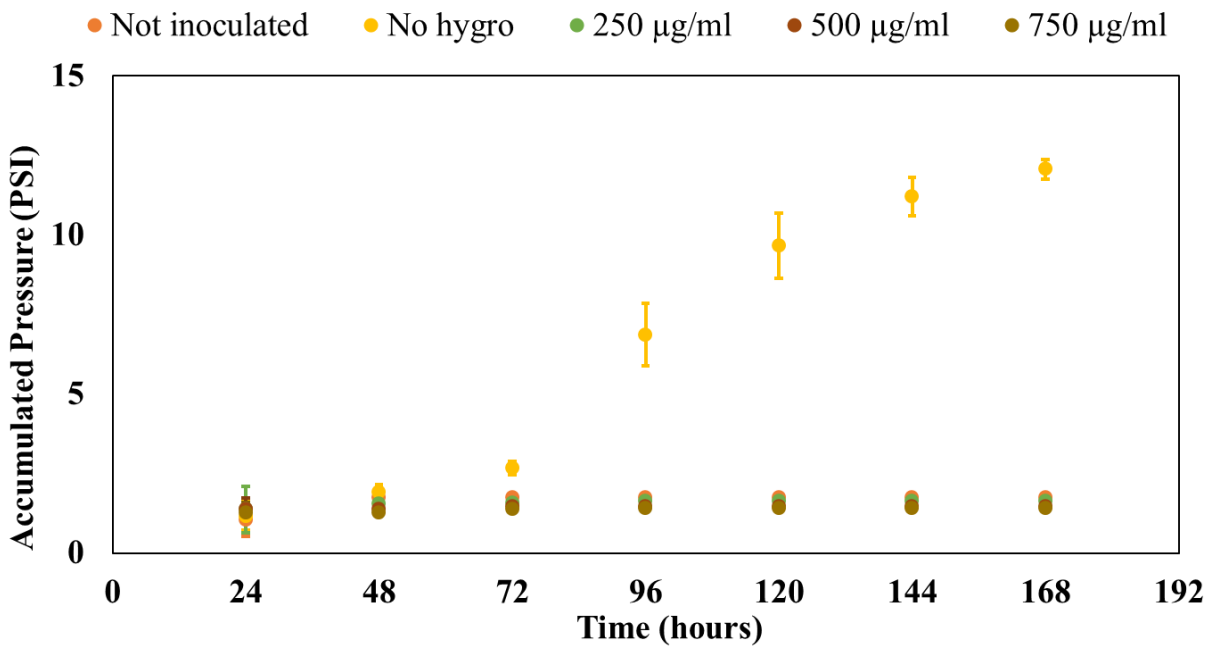


Figure F.1: Growth curve (accumulated pressure) of *Neocallimastix sp. GfMa3-1* under antibiotic selection demonstrating the susceptibility of anaerobic fungi to hygromycin B at various concentrations. Error bars = standard deviation.

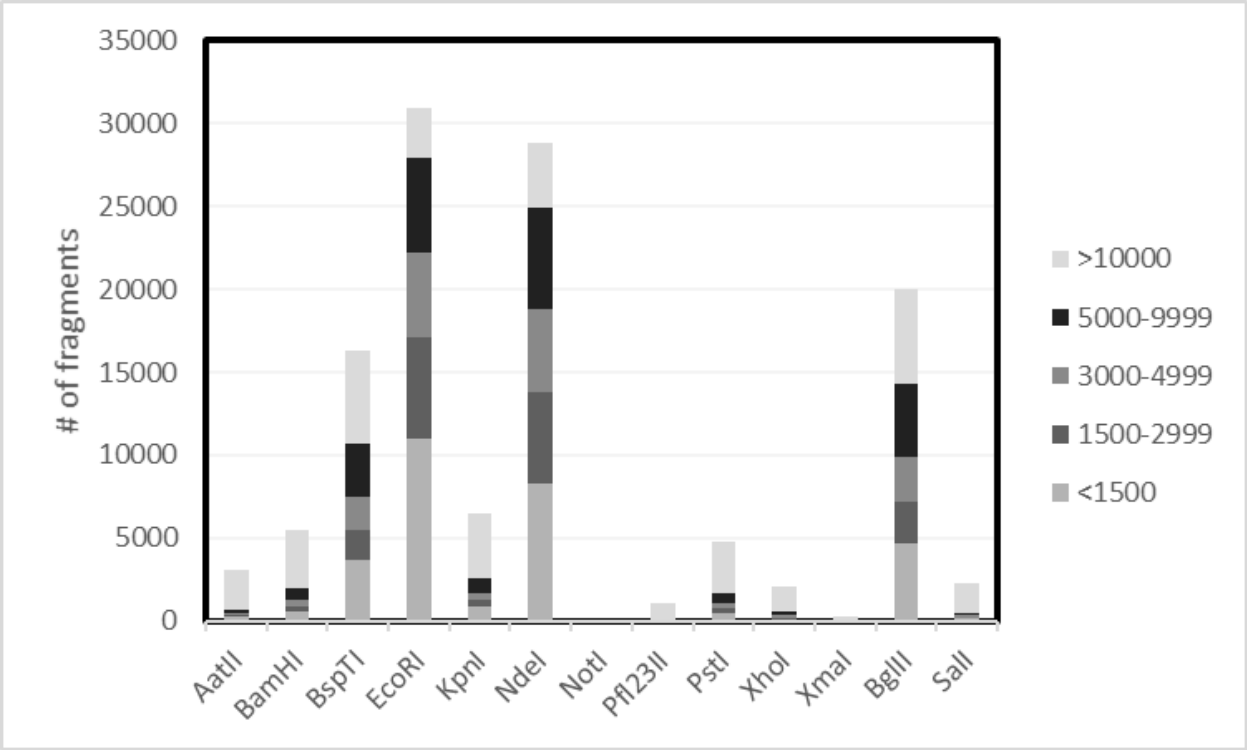


Figure F.2: *In-silico* digests of the *Piromyces sp. UH3-1* genome showing the best enzymes to use to create fragments of various sizes. EcoRI, NdeI, and BglII create the most fragments for reasonably efficient restriction enzyme-based cloning (3 kb or less).

REFERENCES

[1] J. N. . Rogers *et al.*, "An assessment of the potential products and economic and environmental impacts resulting from a billion ton bioeconomy," *Biofuels, Bioprod. Biorefining*, vol. 11, pp. 110–128, 2017.

[2] Z. Anwar, M. Gulfraz, and M. Irshad, "Agro-industrial lignocellulosic biomass a key to unlock the future bio-energy: A brief review," *J. Radiat. Res. Appl. Sci.*, vol. 7, no. 2, pp. 163–173, 2014.

[3] N. Mosier *et al.*, "Features of promising technologies for pretreatment of lignocellulosic biomass," *Bioresour. Technol.*, vol. 96, no. 6, pp. 673–686, 2005.

[4] M. Pauly and K. Keegstra, "Cell-wall carbohydrates and their modification as a resource for biofuels.," *Plant J.*, vol. 54, no. 4, pp. 559–568, May 2008.

[5] E. de Jong, A. Higson, P. Walsh, and M. Wellisch, "Task 42 Biobased Chemicals - Value Added Products from Biorefineries," *A Rep. Prep. IEA Bioenergy-Task*, p. 36, 2011.

[6] A. B. Diaz, A. Blandino, and I. Caro, "Value added products from fermentation of sugars derived from agro-food residues," *Trends Food Sci. Technol.*, vol. 71, no. October, pp. 52–64, 2018.

[7] A. M. Lopes, E. X. Ferreira Filho, and L. R. S. Moreira, "An update on enzymatic cocktails for lignocellulose breakdown," *J. Appl. Microbiol.*, vol. 125, no. 3, pp. 632–645, 2018.

[8] D. Klein-Marcuschamer, P. Oleskowicz-Popiel, B. A. Simmons, and H. W. Blanch, "The challenge of enzyme cost in the production of lignocellulosic biofuels," *Biotechnol. Bioeng.*, vol. 109, no. 4, pp. 1083–1087, 2012.

[9] Y. Sun and J. J. Cheng, "Dilute acid pretreatment of rye straw and bermudagrass for ethanol production," *Bioresour. Technol.*, vol. 96, no. 14, pp. 1599–1606, 2005.

[10] H. Palonen, F. Tjerneld, G. Zacchi, and M. Tenkanen, "Adsorption of Trichoderma reesei CBH I and EG II and their catalytic domains on steam pretreated softwood and isolated lignin," *J. Biotechnol.*, vol. 107, no. 1, pp. 65–72, 2004.

[11] V. Flad *et al.*, "The Biotechnological Potential of Anaerobic Gut Fungi," in *Genetics and Biotechnology*, J. P. Benz and K. Schipper, Eds. Cham: Springer International Publishing, 2020, pp. 413–437.

[12] A. Ranganathan, O. P. Smith, N. H. Youssef, C. G. Struchtemeyer, H. K. Atiyeh, and M. S. Elshahed, "Utilizing Anaerobic Fungi for Two-stage Sugar Extraction and Biofuel Production from Lignocellulosic Biomass," *Frontiers in Microbiology*, vol. 8. 2017.

[13] J. K. Henske *et al.*, “Metabolic characterization of anaerobic fungi provides a path forward for bioprocessing of crude lignocellulose,” *Biotechnol. Bioeng.*, vol. 115, no. 4, pp. 874–884, 2018.

[14] C. A. Hooker *et al.*, “Hydrolysis of untreated lignocellulosic feedstock is independent of S-lignin composition in newly classified anaerobic fungal isolate, Piromyces sp. UH3-1,” *Biotechnol. Biofuels*, vol. 11, no. 1, pp. 1–14, 2018.

[15] S. Seppälä, S. E. Wilken, D. Knop, K. V. Solomon, and M. A. O’Malley, “The importance of sourcing enzymes from non-conventional fungi for metabolic engineering and biomass breakdown,” *Metab. Eng.*, vol. 44, no. September, pp. 45–59, 2017.

[16] C. Wan and Y. Li, “Fungal pretreatment of lignocellulosic biomass,” *Biotechnol. Adv.*, vol. 30, no. 6, pp. 1447–1457, 2012.

[17] O. Skyba, C. J. Douglas, and S. D. Mansfield, “Syringyl-Rich lignin renders poplars more resistant to degradation by wood decay fungi,” *Appl. Environ. Microbiol.*, vol. 79, no. 8, pp. 2560–2571, 2013.

[18] K. V Solomon *et al.*, “Early-branching gut fungi possess large , comprehensive array of biomass-degrading enzymes,” *Science (80-.).*, vol. 351, no. 6278, pp. 1192–1195, 2016.

[19] S. E. Wilken *et al.*, “Genomic and proteomic biases inform metabolic engineering strategies for anaerobic fungi,” *Metab. Eng. Commun.*, vol. 10, no. November 2019, 2020.

[20] M. A. O’Malley, M. K. Theodorou, and C. A. Kaiser, “Evaluating Expression and Catalytic Activity of Anaerobic Fungal Fibroyltic Enzymes Native to Piromyces sp E2 in Saccharomyces cerevisiae,” *Environ. Sci. Technol.*, vol. 31, no. 1, pp. 37–46, 2012.

[21] S. Seppälä, J. I. Yoo, D. Yur, and M. A. O’Malley, “Heterologous transporters from anaerobic fungi bolster fluoride tolerance in Saccharomyces cerevisiae,” *Metab. Eng. Commun.*, vol. 9, no. March, 2019.

[22] C. Lambertz *et al.*, “Challenges and advances in the heterologous expression of cellulolytic enzymes: a review,” *Biotechnol. Biofuels*, vol. 7, no. 1, p. 135, 2014.

[23] E. T. Hillman, L. R. Readnour, and K. V. Solomon, “Exploiting the natural product potential of fungi with integrated -omics and synthetic biology approaches,” *Curr. Opin. Syst. Biol.*, 2017.

[24] C. Walsh, “Where will new antibiotics come from?,” *Nat. Rev. Microbiol.*, vol. 1, no. 1, pp. 65–70, 2003.

[25] M. Sharma and R. Sharma, “Drugs and drug intermediates from fungi: Striving for greener processes,” *Crit. Rev. Microbiol.*, vol. 7828, no. June, pp. 1–17, 2014.

[26] J. Avalos and M. Carmen Limón, “Biological roles of fungal carotenoids,” *Curr. Genet.*, vol. 61, no. 3, pp. 309–324, 2014.

[27] A. A. Brakhage, “Regulation of fungal secondary metabolism,” *Nat. Rev. Microbiol.*, vol. 11, no. 1, pp. 21–32, 2012.

[28] L. Tedersoo *et al.*, “Disentangling global soil fungal diversity,” *Science*, vol. 346, no. 6213, pp. 1052–1053, 2014.

[29] M. Mar Rodríguez *et al.*, “Obesity changes the human gut mycobiome,” *Sci. Rep.*, vol. 5, p. 14600, 2015.

[30] R. J. Gruninger *et al.*, “Anaerobic fungi (phylum Neocallimastigomycota): Advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential,” *FEMS Microbiol. Ecol.*, vol. 90, no. 1, pp. 1–17, 2014.

[31] H. E. O’Brien, J. L. Parrent, J. A. Jackson, J. Moncalvo, and R. Vilgalys, “Fungal Community Analysis by Large-Scale Sequencing of Environmental Samples,” *Appl. Environ. Microbiol.*, vol. 71, no. 9, pp. 5544–5550, 2005.

[32] M. Blackwell, “The fungi: 1, 2, 3 ... 5.1 million species?,” *Am. J. Bot.*, vol. 98, no. 3, pp. 426–438, 2011.

[33] I. V. Grigoriev *et al.*, “MycoCosm portal: Gearing up for 1000 fungal genomes,” *Nucleic Acids Res.*, vol. 42, no. D1, pp. 1–6, 2014.

[34] Z. Charlop-Powers *et al.*, “Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity,” *Proc. Natl. Acad. Sci.*, vol. 113, no. 51, pp. 14811–14816, 2016.

[35] C. H. Haitjema *et al.*, “A Parts List for Fungal Cellulosomes Revealed by Comparative Genomics,” *Nat. Microbiol.*, vol. 2, no. May, p. 17087, 2017.

[36] D. Hoffmeister and N. P. Keller, “Natural products of filamentous fungi: enzymes, genes, and their regulation,” *Nat. Prod. Rep.*, vol. 24, no. 2, pp. 393–416, 2007.

[37] M. Gressler, P. Hortschansky, E. Geib, and M. Brock, “A new high-performance heterologous fungal expression system based on regulatory elements from the Aspergillus terreus terrein gene cluster,” *Front. Microbiol.*, vol. 6, p. 184, 2015.

[38] H. Chai *et al.*, “Sesterterpene ophiobolin biosynthesis involving multiple gene clusters in Aspergillus ustus,” *Sci. Rep.*, vol. 6, no. 1, p. 27181, 2016.

[39] N. P. Keller, G. Turner, and J. W. Bennett, “Fungal secondary metabolism — from biochemistry to genomics,” *Nat. Rev. Microbiol.*, vol. 3, no. 12, pp. 937–947, 2005.

[40] M. Subhan, R. Faryal, and I. Macreadie, “Exploitation of Aspergillus terreus for the production of natural statins,” *J. Fungi*, vol. 2, p. 13, 2016.

[41] M. Hashimoto, T. Nonaka, and I. Fujii, “Fungal type III polyketide synthases,” *Nat. Prod. Rep.*, vol. 31, no. 10, pp. 1306–1317, 2014.

[42] P. K. Chang and K. C. Ehrlich, “Genome-wide analysis of the Zn(II)2Cys6 zinc cluster-encoding gene family in Aspergillus flavus,” *Appl. Microbiol. Biotechnol.*, vol. 97, no. 10, pp. 4289–4300, 2013.

[43] P. Wiemann and N. P. Keller, “Strategies for mining fungal natural products,” *J. Ind. Microbiol. Biotechnol.*, vol. 41, no. 2, pp. 301–313, 2014.

[44] M. A. van den Berg, I. Westerlaken, C. Leeflang, R. Kerkman, and R. A. L. Bovenberg, “Functional characterization of the penicillin biosynthetic gene cluster of Penicillium chrysogenum Wisconsin54-1255,” *Fungal Genet. Biol.*, vol. 44, no. 9, pp. 830–844, 2007.

[45] T. Weber and H. U. Kim, “The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production,” *Synth. Syst. Biotechnol.*, vol. 1, no. 2, pp. 69–79, 2016.

[46] N. Khaldi *et al.*, “SMURF: Genomic mapping of fungal secondary metabolite clusters,” *Fungal Genet. Biol.*, vol. 47, no. 9, pp. 736–741, 2010.

[47] T. Weber *et al.*, “AntiSMASH 3.0-A comprehensive resource for the genome mining of biosynthetic gene clusters,” *Nucleic Acids Res.*, vol. 43, no. W1, pp. W237–W243, 2015.

[48] D. O. Inglis *et al.*, “Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of Aspergillus nidulans, A. fumigatus, A. niger and A. oryzae,” *BMC Microbiol.*, vol. 13, no. 1, p. 91, 2013.

[49] J. C. Nielsen *et al.*, “Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in Penicillium species,” *Nat. Microbiol.*, vol. 2, no. April, p. 17044, 2017.

[50] C. W. Johnston *et al.*, “An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products,” *Nat. Commun.*, vol. 6, p. 8421, 2015.

[51] C. A. Dejong *et al.*, “Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching,” *Nat. Chem. Biol.*, vol. 12, no. 12, pp. 1007–1014, 2016.

[52] R. A. Hughes and A. D. Ellington, “Synthetic DNA Synthesis and Assembly : Putting the Synthetic in Synthetic Biology,” *Cold Spring Harb. Perspect. Biol.*, vol. 9, no. 1, p. a023812, 2017.

[53] Y. M. Chiang *et al.*, “An efficient system for heterologous expression of secondary metabolite genes in Aspergillus nidulans,” *J. Am. Chem. Soc.*, vol. 135, no. 20, pp. 7720–7731, 2013.

[54] M. N. Heneghan *et al.*, “First heterologous reconstruction of a complete functional fungal biosynthetic multigene cluster,” *ChemBioChem*, vol. 11, no. 11, pp. 1508–1512, 2010.

[55] J. W. Bok *et al.*, “Fungal artificial chromosomes for mining of the fungal secondary metabolome,” *BMC Genomics*, vol. 16, p. 343, 2015.

[56] M. T. Nielsen *et al.*, "Heterologous Reconstitution of the Intact Geodin Gene Cluster in Aspergillus nidulans through a Simple and Versatile PCR Based Approach," *PLoS One*, vol. 8, no. 8, p. e72871, 2013.

[57] A. Endo, "A historical perspective on the discovery of statins," *Proc. Japan Acad. Ser. B*, vol. 86, no. 5, pp. 484–493, 2010.

[58] A. L. Demain, "MINI-REVIEW Pharmaceutically active secondary metabolites of microorganisms," *Appl. Microbiol. Biotechnol.*, vol. 52, pp. 455–463, 1999.

[59] C. L. Ventola, "The antibiotic resistance crisis: part 1: causes and threats.," *P T A peer-reviewed J. Formul. Manag.*, vol. 40, no. 4, pp. 277–83, 2015.

[60] P. Bhosale, "Environmental and cultural stimulants in the production of carotenoids from microorganisms," *Appl. Microbiol. Biotechnol.*, vol. 63, no. 4, pp. 351–361, 2004.

[61] A. M. Beekman and R. A. Barrow, "Fungal metabolites as pharmaceuticals," *Aust. J. Chem.*, vol. 67, no. 6, pp. 827–843, 2014.

[62] S. B. Singh *et al.*, "Nodulisporic acids D-F: Structure, biological activities, and biogenetic relationships," *J. Nat. Prod.*, vol. 67, no. 9, pp. 1496–1506, 2004.

[63] L. Mata-Gómez, J. Montañez, A. Méndez-Zavala, and C. Aguilar, "Biotechnological production of carotenoids by yeasts: an overview," *Microb. Cell Fact.*, vol. 13, no. 1, p. 12, 2014.

[64] R. Liu, L. Chen, Y. Jiang, Z. Zhou, and G. Zou, "Efficient genome editing in filamentous fungus Trichoderma reesei using the CRISPR/Cas9 system," *Cell Discov.*, vol. 1, p. 15007, 2015.

[65] C. Pohl, J. A. K. W. Kiel, A. J. M. Driessen, R. A. L. Bovenberg, and Y. Nygard, "CRISPR/Cas9 Based Genome Editing of Penicillium chrysogenum," *ACS Synth. Biol.*, vol. 5, no. 7, pp. 754–764, 2016.

[66] T. Matsu-ura, M. Baek, J. Kwon, and C. Hong, "Efficient gene editing in Neurospora crassa with CRISPR technology," *Fungal Biol. Biotechnol.*, vol. 2, no. 1, p. 4, 2015.

[67] L. Richter *et al.*, "Engineering of Aspergillus niger for the production of secondary metabolites," *Fungal Biol. Biotechnol.*, vol. 1, no. 1, p. 4, 2014.

[68] E. K. Shwab, W. B. Jin, M. Tribus, J. Galehr, S. Graessle, and N. P. Keller, "Histone deacetylase activity regulates chemical diversity in Aspergillus," *Eukaryot. Cell*, vol. 6, no. 9, pp. 1656–1664, 2007.

[69] W. H. Meijer *et al.*, "Peroxisomes are required for efficient penicillin biosynthesis in Penicillium chrysogenum," *Appl. Environ. Microbiol.*, vol. 76, no. 17, pp. 5702–5709, 2010.

[70] S. H. Akone *et al.*, "Inducing secondary metabolite production by the endophytic fungus Chaetomium sp. through fungal-bacterial co-culture and epigenetic modification," *Tetrahedron*, vol. 72, no. 41, pp. 6340–6347, 2016.

[71] C. Wu, B. Zacchetti, A. F. J. Ram, G. P. van Wezel, D. Claessen, and Y. Hae Choi, "Expanding the chemical space for natural products by Aspergillus-Streptomyces co-cultivation and biotransformation," *Sci. Rep.*, vol. 5, no. March, p. 10868, 2015.

[72] V. Schroeckh *et al.*, "Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in Aspergillus nidulans," *Proc. Natl. Acad. Sci.*, vol. 106, no. 34, pp. 14558–14563, 2009.

[73] W. Yin and N. P. Keller, "Transcriptional regulatory elements in fungal secondary metabolism," *J. Microbiol.*, vol. 49, no. 3, pp. 329–339, 2011.

[74] H. Nützmann, Y. Reyes-dominguez, K. Scherlach, V. Schroeckh, and F. Horn, "Bacteria-induced natural product formation in the fungus Aspergillus nidulans requires Saga / Ada-mediated histone acetylation," *Proc. Natl. Acad. Sci.*, vol. 108, no. 34, pp. 14282–14287, 2011.

[75] T. Netzker *et al.*, "Microbial communication leading to the activation of silent fungal secondary metabolite gene clusters," *Front. Microbiol.*, vol. 6, no. MAR, pp. 1–13, 2015.

[76] D. Krug and R. Müller, "Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products," *Nat. Prod. Rep.*, vol. 31, no. 6, p. 768, 2014.

[77] M. Vansteelandt *et al.*, "Patulin and secondary metabolite production by marine-derived Penicillium strains," *Fungal Biol.*, vol. 116, no. 9, pp. 954–961, 2012.

[78] R. Tautenhahn, K. Cho, W. Uritboonthai, Z. Zhu, G. J. Patti, and G. Siuzdak, "An accelerated workflow for untargeted metabolomics using the METLIN database," *Nat. Biotechnol.*, vol. 30, no. 9, pp. 826–828, 2012.

[79] M. Wang *et al.*, "Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking," *Nat. Biotechnol.*, vol. 34, no. 8, pp. 828–837, 2016.

[80] J. Fang and P. C. Dorrestein, "Emerging mass spectrometry techniques for the direct analysis of microbial colonies," *Curr. Opin. Microbiol.*, vol. 19, no. 1, pp. 120–129, 2014.

[81] X. Wang *et al.*, "Genomic and transcriptomic analysis of the endophytic fungus Pestalotiopsis fici reveals its lifestyle and high potential for synthesis of natural products," *BMC Genomics*, vol. 16, no. 1, p. 28, 2015.

[82] P. Daly *et al.*, "Transcriptomic responses of mixed cultures of ascomycete fungi to lignocellulose using dual RNA-seq reveal inter-species antagonism and limited beneficial effects on CAZyme expression," *Fungal Genet. Biol.*, vol. 102, pp. 4–21, 2016.

[83] A. H. Smits and M. Vermeulen, “Characterizing Protein-Protein Interactions Using Mass Spectrometry: Challenges and Opportunities,” *Trends Biotechnol.*, vol. 34, no. 10, pp. 825–834, 2016.

[84] L. F. C. Ribeiro, C. L. Chelius, S. D. Harris, and M. R. Marten, “Insights Regarding Fungal Phosphoproteomic Analysis,” *Fungal Genet. Biol.*, vol. 104, pp. 38–44, 2017.

[85] S. B. Bumpus, B. S. Evans, P. M. Thomas, I. Ntai, and N. L. Kelleher, “A proteomics approach to discovering natural products and their biosynthetic pathways,” *Nat. Biotechnol.*, vol. 27, no. 10, pp. 951–956, 2009.

[86] E. Gutierrez, A. García-Villaraco, J. A. Lucas, A. Gradillas, F. J. Gutierrez-Mañero, and B. Ramos-Solano, “Transcriptomics, Targeted Metabolomics and Gene Expression of Blackberry Leaves and Fruits Indicate Flavonoid Metabolic Flux from Leaf to Red Fruit,” *Front. Plant Sci.*, vol. 8, no. April, pp. 1–15, 2017.

[87] K. Karničar *et al.*, “Integrated omics approaches provide strategies for rapid erythromycin yield increase in Saccharopolyspora erythraea,” *Microb. Cell Fact.*, vol. 15, p. 93, 2016.

[88] F. Y. Lim, J. F. Sanchez, C. C. C. Wang, and N. P. Keller, “Toward Awakening Cryptic Secondary Merabolite Gene Clusters in Filamentous Fungi,” *Methods Enzymol.*, vol. 517, pp. 303–324, 2012.

[89] Z. Büttel *et al.*, “Unlocking the potential of fungi: the QuantFung project,” *Fungal Biol. Biotechnol.*, vol. 2, p. 6, 2015.

[90] J. F. Imhoff, “Natural products from marine fungi - Still an underrepresented resource,” *Mar. Drugs*, vol. 14, p. 19, 2016.

[91] C. G. Orpin, “Studies on the Rumen Flagellate Neocallimastix frontalis,” *J. Gen. Microbiol.*, vol. 91, no. 2, pp. 249–262, 1975.

[92] C. H. Haitjema, K. V Solomon, J. K. Henske, M. K. Theodorou, and M. A. O. Malley, “Anaerobic Gut Fungi: Advances in Isolation, Culture, and Cellulolytic Enzyme Discovery for Biofuel Production,” *Biotechnol. Bioeng.*, vol. 111, no. 8, pp. 1471–1482, 2014.

[93] M. K. Theodorou, G. Mennim, D. R. Davies, W. Y. Zhu, a P. Trinci, and J. L. Brookman, “Anaerobic fungi in the digestive tract of mammalian herbivores and their potential for exploitation.,” *Proc. Nutr. Soc.*, vol. 55, pp. 913–926, 1996.

[94] I. V. Grigoriev *et al.*, “MycoCosm portal: Gearing up for 1000 fungal genomes,” *Nucleic Acids Res.*, vol. 42, no. D1, pp. 699–704, 2014.

[95] I. V. Grigoriev *et al.*, “Fueling the future with fungal genomics,” *Mycology*, vol. 2, no. 3, pp. 192–209, 2011.

[96] H. Nordberg *et al.*, “The genome portal of the Department of Energy Joint Genome Institute: 2014 updates,” *Nucleic Acids Res.*, vol. 42, no. D1, pp. 26–32, 2014.

[97] E. M. Musiol-Kroll *et al.*, "Polyketide Bioderivatization Using the Promiscuous Acyltransferase KirCII," *ACS Synth. Biol.*, vol. 6, no. 3, pp. 421–427, 2017.

[98] S. Yuzawa, J. D. Keasling, and L. Katz, "Insights into polyketide biosynthesis gained from repurposing antibiotic-producing polyketide synthases to produce fuels and chemicals," *J. Antibiot. (Tokyo).*, vol. 69, pp. 494–499, 2016.

[99] S. Yuzawa *et al.*, "Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production," *ACS Synth. Biol.*, vol. 6, no. 1, pp. 139–147, 2017.

[100] C. A. Hooker, "Developing Anaerobic Fungi As a platform for Efficient lignocellulose hydrolysis," 2019.

[101] E. M. Rubin, "Genomics of cellulosic biofuels," *Nature*, vol. 454, no. 7206, pp. 841–5, 2008.

[102] J. R. Porter, F. M. Howell, P. B. Mason, and T. C. Blanchard, "Existing biomass infrastructure and theoretical potential biomass production in the US," *J Maps*, vol. 5, 2009.

[103] T. Kawai *et al.*, "Analysis of the saccharification capability of high-functional cellulase JN11 for various pretreated biomasses through a comparison with commercially available counterparts," *J Ind Microbiol Biotechnol*, vol. 39, 2012.

[104] E. Dondelinger, N. Aubry, F. B. Chaabane, C. Cohen, J. Tayeb, and C. Rémond, "Contrasted enzymatic cocktails reveal the importance of cellulases and hemicellulases activity ratios for the hydrolysis of cellulose in presence of xylans," *AMB Express*, vol. 6, p. 24, 2016.

[105] D. Klein-Marcuschamer, P. Oleskowicz-Popiel, B. A. Simmons, and H. W. Blanch, "The challenge of enzyme cost in the production of lignocellulosic biofuels," *Biotechnol Bioeng*, vol. 109, 2012.

[106] A. Harkki *et al.*, "Genetic engineering of Trichoderma to produce strains with novel cellulase profiles," *Enzym. Microb Technol*, vol. 13, 1991.

[107] R. H. Bischof, J. Ramoni, and B. Seiboth, "Cellulases and beyond: the first 70 years of the enzyme producer Trichoderma reesei," *Microb Cell Fact*, vol. 15, no. 1, p. 106, 2016.

[108] M. Schülein, "Cellulases of Trichoderma reesei," in *Methods in enzymology*, 1st ed., London: Academic Press, 1988, pp. 234–42.

[109] H. Nakazawa *et al.*, "Construction of a recombinant Trichoderma reesei strain expressing Aspergillus aculeatus β-glucosidase 1 for efficient biomass conversion," *Biotechnol Bioeng*, vol. 109, no. 1, pp. 92–9, 2012.

[110] M. J. Nicholson, M. K. Theodorou, and J. L. Brookman, "Molecular analysis of the anaerobic rumen fungus Orpinomyces - Insights into an AT-rich genome," *Microbiology*, vol. 151, no. 1, pp. 121–133, 2005.

[111] S. S. Dagar, S. Kumar, P. Mudgil, R. Singh, and A. K. Puniya, "D1/D2 domain of large-subunit ribosomal DNA for differentiation of Orpinomyces spp," *Appl Env. Microb*, vol. 77, no. 18, pp. 6722–5, 2011.

[112] C. L. Schoch *et al.*, "Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 16, pp. 1–6, 2012.

[113] W. Walters *et al.*, "Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Analysis," *mSystems*, vol. 1, no. 1, pp. e0009-15, 2016.

[114] A. S. Liggenstoffer, N. H. Youssef, M. B. Couger, and M. S. Elshahed, "Phylogenetic diversity and community structure of anaerobic gut fungi (phylum Neocallimastigomycota) in ruminant and non-ruminant herbivores," *ISME J*, vol. 4, no. 10, pp. 1225–35, 2010.

[115] V. Dollhofer, S. M. Podmirseg, T. M. Callaghan, G. W. Griffith, and K. Fliegerová, "Anaerobic fungi and their potential for biogas production," *Adv Biochem Eng Biotechnol*, vol. 151, pp. 41–61, 2015.

[116] M. K. Theodorou, B. A. Williams, M. S. Dhanoa, A. B. McAllan, and J. France, "A simple gas production method using a pressure transducer to determine the fermentation kinetics of ruminant feeds," *Anim. Feed Sci. Technol.*, vol. 48, no. 3, pp. 185–197, 1994.

[117] T. Juhász, A. Egyházi, and K. Réczey, "β-Glucosidase production by Trichoderma reesei," *Appl Biochem Biotechnol*, vol. 121, no. 1, p. 243, 2005.

[118] D. Mikkelsen, B. M. Flanagan, S. M. Wilson, A. Bacic, and M. J. Gidley, "Interactions of arabinoxylan and (1,3)(1,4)-β-glucan with cellulose networks," *Biomacromolecules*, vol. 16, 2015.

[119] M. C. Edwards and J. Doran-Peterson, "Pectin-rich biomass as feedstock for fuel ethanol production," *Appl. Microbiol. Biotechnol.*, vol. 95, no. 3, pp. 565–575, 2012.

[120] D. Mohnen, "Pectin structure and biosynthesis," *Curr Opin Plant Biol*, vol. 11, no. 3, pp. 266–77, 2008.

[121] D. Mohnen, "Pectins and their manipulation," in *Biosynthesis of pectins*, G. S. A. J. Knox, Ed. Oxford: Blackwell Publishing, Ltd, 2002.

[122] R. M. Teather and P. J. Wood, "Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen," *Appl Env. Microbiol*, vol. 43, no. 4, pp. 777–80, 1982.

[123] D. Kumar and G. S. Murthy, "Impact of pretreatment and downstream processing technologies on economics and energy in cellulosic ethanol production," *Biotechnol Biofuels*, vol. 4, 2011.

[124] S. Park, J. O. Baker, M. E. Himmel, P. A. Parilla, and D. K. Johnson, "Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance," *Biotechnol Biofuels*, vol. 3, 2010.

[125] M. Hall, P. Bansal, J. H. Lee, M. J. Realff, and A. S. Bommarius, "Cellulose crystallinity a key predictor of the enzymatic hydrolysis rate," *FEBS J*, vol. 277, 2010.

[126] A. Peciulyte, K. Karlström, P. T. Larsson, and L. Olsson, "Impact of the supramolecular structure of cellulose on the efficiency of enzymatic hydrolysis," *Biotechnol Biofuels*, vol. 8, 2015.

[127] R. Barbehenn and E. Bernays, "Relative nutritional quality of -C3 and -C4 grasses for a graminivorous lepidopteran, Paratrytone melane (Hesperiidae). Oecologia," *Oecologia*, vol. 92, no. 1, pp. 97–103, 1992.

[128] S. Mathur, A. V Umakanth, V. A. Tonapi, R. Sharma, and M. K. Sharma, "Sweet sorghum as biofuel feedstock: recent advances and available resources," *Biotechnol Biofuels*, vol. 10, no. 1, p. 146, 2017.

[129] P. S. Rao, C. G. Kumar, R. S. Prakasham, A. U. Rao, and B. V. S. Reddy, "Sweet sorghum: breeding and bioproducts," in *Industrial crops: breeding for bioenergy and bioproducts*, V. M. V Cruz and D. A. Dierig, Eds. New York: Springer, 2015, pp. 1–28.

[130] D. J. Parrish and J. H. Fike, "The biology and agronomy of switchgrass for biofuels," *Crit Rev Plant Sci*, vol. 24, no. 5–6, pp. 423–59, 2005.

[131] J. Littlewood, M. Guo, W. Boerjan, and R. J. Murphy, "Bioethanol from poplar: a commercially viable alternative to fossil fuel in the European Union," *Biotechnol Biofuels*, vol. 7, p. 113, 2014.

[132] J. J. K. D. Balatinecz, *Properties and utilization of poplar wood*. Ontario: Canadian Science Publishing NRC Research Press, 2001.

[133] A. Wiselogel, J. Tyson, and D. E. Johnsson, "Biomass feedstock resources and composition," in *Handbook on bioethanol: production and utilization*, C. Wyman, Ed. Taylor & Francis, 1996.

[134] Y. Kim, N. S. Mosier, and M. R. Ladisch, "Enzymatic digestion of liquid hot water pretreated hybrid poplar," *Biotechnol Prog*, vol. 25, pp. 340–8, 2009.

[135] L. Christersson, "Wood production potential in poplar plantations in Sweden," *Biomass Bioenergy*, vol. 34, 2010.

[136] H. Jørgensen, J. B. Kristensen, and C. Felby, "Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities," *Biofuel Bioprod Biorefin*, vol. 1, 2007.

[137] K. N. Joblin and G. E. Naylor, "Fermentation of woods by rumen anaerobic fungi," *FEMS Microbiol Lett*, vol. 65, no. 1, pp. 119–22, 1989.

[138] J. R. Erb-Downward *et al.*, "Analysis of the lung microbiome in the 'healthy' smoker and in COPD," *PLoS One*, vol. 6, no. 2, 2011.

[139] R. Franke, C. M. McMichael, K. Meyer, A. M. Shirley, J. C. Cusumano, and C. Chapple, "Modified lignin in tobacco and poplar plants over-expressing the Arabidopsis gene encoding ferulate 5-hydroxylase," *Plant J*, vol. 22, no. 3, pp. 223–34, 2000.

[140] L. J. Jönsson and C. Martín, "Pretreatment of lignocellulose: formation of inhibitory by-products and strategies for minimizing their effects," *Bioresour Technol*, vol. 199, 2016.

[141] T. H. Kim, J. S. Kim, C. Sunwoo, and Y. Y. Lee, "Pretreatment of corn stover by aqueous ammonia," *Bioresour Technol*, vol. 90, 2003.

[142]B. Yang and C. E. Wyman, "Pretreatment: the key to unlocking low-cost cellulosic ethanol," *Biofuel Bioprod Biorefin*, vol. 2, no. 1, pp. 26–40, 2008.

[143] S. M. Kim, B. S. Dien, M. E. Tumbleson, K. D. Rausch, and V. Singh, "Improvement of sugar yields from corn stover using sequential hot water pretreatment and disk milling," *Bioresour Technol*, vol. 216, pp. 706–13, 2016.

[144] A. C. Rodrigues, M. Ø. Haven, J. Lindedam, C. Felby, and M. Gama, "Celluclast and -Cellic® CTec2: saccharification/fermentation of wheat straw, solid–liquid partition and potential of enzyme recycling by alkaline washing," *Enzym. Microb Technol*, vol. 70, no. 7, pp. 79–80, 2015.

[145] J. Hu, V. Arantes, and J. N. Saddler, "The enhancement of enzymatic hydrolysis of lignocellulosic substrates by the addition of accessory enzymes such as xylanase: is it an additive or synergistic effect?," *Biotechnol Biofuels*, vol. 4, p. 36, 2011.

[146] D. S. Chahal, S. McGuire, H. Pikor, and G. Noble, "Production of cellulase complex by Trichoderma reesei Rut-C30 on lignocellulose and its hydrolytic potential," *Biomass*, vol. 2, 1982.

[147] B. S. Montenecourt and D. E. Eveleigh, "Selective screening methods for the isolation of high yielding cellulase mutants of Trichoderma reesei," in *American Chemical Society*, 1979, pp. 289–301.

[148] K. M. Nevalainen, "Induction, isolation, and characterization of Aspergillus niger mutant strains producing elevated levels of beta-galactosidase," *Appl Envir Microb*, vol. 41, 1981.

[149] S. Mohanram, D. Amat, J. Choudhary, A. Arora, and L. Nain, "Novel perspectives for evolving enzyme cocktails for lignocellulose hydrolysis in biorefineries," *Sustain Chem Proc*, vol. 1, 2013.

[150] S. E. Lowe, G. G. Griffith, A. Milne, M. K. Theodorou, and A. P. J. Trinci, "The life cycle and growth kinetics of an anaerobic rumen fungus," *J. Gen. Microbiol.*, vol. 133, no. 7, pp. 1815–1827, 1987.

[151] R. Meilan and C. Ma, "Poplar (Populus spp.)," in *Agrobacterium protocols, vol. 2*, K. Wang, Ed. Totowa: Humana Press, 2007, pp. 143–151.

[152] N. A. Anderson *et al.*, "Manipulation of Guaiacyl and Syringyl Monomer Biosynthesis in an Arabidopsis Cinnamyl Alcohol Dehydrogenase Mutant Results in Atypical Lignin Biosynthesis and Modified Cell Wall Structure," *The Plant Cell*, vol. 27, no. 8. pp. 2195–2209, Aug-2015.

[153] R. P. Overend and E. Chornet, "Fractionation of lignocellulosics by steam-aqueous pretreatments," *Phil Trans Roy Soc L. A*, vol. 321, 1987.

[154] G. Cao *et al.*, "Bioabatement with hemicellulase supplementation to reduce enzymatic hydrolysis inhibitors," *Bioresour Technol*, vol. 190, pp. 412–5, 2015.

[155] D. S. Tuckwell, M. J. Nicholson, C. S. McSweeney, M. K. Theodorou, and 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*, vol. 151, no. 5, pp. 1557–1567, 2005.

[156] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.," *Mol. Biol. Evol.*, no. April, p. msw054, 2016.

[157] M. M. Bradford, "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding," *Anal Biochem*, vol. 72, 1976.

[158] M.-J. Tseng, M.-N. Yap, K. Ratanakhanokchai, K. L. Kyu, and S.-T. Chen, "Purification and characterization of two cellulase free xylanases from an alkaliphilic Bacillus firmus," *Enzyme Microb. Technol.*, vol. 30, no. 5, pp. 590–595, 2002.

[159] Z. Xiao, R. Storms, and A. Tsang, "Microplate-based filter paper assay to measure total cellulase activity," *Biotechnol. Bioeng.*, vol. 88, no. 7, pp. 832–837, 2004.

[160] A. Sluiter *et al.*, "Determination of structural carbohydrates and lignin in Biomass," *Natl. Renew. Energy Lab.*, vol. August, no. TP-510-42618, p. 17, 2012.

[161] B. Hames, R. Ruiz, C. Scarlata, A. Sluiter, J. Sluiter, and D. Templeton, "Preparation of Samples for Compositional Analysis," *Natl. Renew. Energy Lab.*, vol. August, no. TP-510-42620, 2008.

[162] A. Sluiter, R. Ruiz, C. Scarlata, J. Sluiter, and D. Templeton, "Determination of Extractives in Biomass," *Natl. Renew. Energy Lab.*, vol. January, no. TP-510-42619, pp. 838–852, 2008.

[163] D. J. Newman and G. M. Cragg, "Natural Products as Sources of New Drugs from 1981 to 2014," *J. Nat. Prod.*, vol. 79, no. 3, pp. 629–661, 2016.

[164] N. H. Youssef *et al.*, "The genome of the anaerobic fungus orpinomyces sp. strain c1a reveals the unique evolutionary history of a remarkable plant biomass degrader," *Appl. Environ. Microbiol.*, vol. 79, no. 15, pp. 4620–4634, 2013.

[165] N. H. Youssef *et al.*, "The genome of the anaerobic fungus orpinomyces sp. strain c1a reveals the unique evolutionary history of a remarkable plant biomass degrader," *Appl. Environ. Microbiol.*, vol. 79, no. 15, pp. 4620–4634, 2013.

[166] R. Durand, C. Rascle, M. Fischer, and M. Fèvre, "Transient expression of the β-glucuronidase gene after biolistic transformation of the anaerobic fungus Neocalimastix frontalis," *Curr. Genet.*, vol. 31, no. 2, pp. 158–161, 1997.

[167] S. S. Calkins *et al.*, "Development of an RNA interference (RNAi) gene knockdown protocol in the anaerobic gut fungus *Pecoramyces ruminantium* strain C1A," *PeerJ*, vol. 6, p. e4276, 2018.

[168] S. P. Gilmore *et al.*, "Top-Down Enrichment Guides in Formation of Synthetic Microbial Consortia for Biomass Degradation," *ACS Synth. Biol.*, vol. 8, no. 9, pp. 2174–2185, 2019.

[169] K. V. Solomon, J. K. Henske, M. K. Theodorou, and M. A. O'Malley, "Robust and effective methodologies for cryopreservation and DNA extraction from anaerobic gut fungi," *Anaerobe*, vol. 38, pp. 39–46, 2016.

[170] A. Rhoads and K. F. Au, "PacBio Sequencing and Its Applications," *Genomics, Proteomics Bioinforma.*, vol. 13, no. 5, pp. 278–289, 2015.

[171] N. Varoquaux *et al.*, "Accurate identification of centromere locations in yeast genomes using Hi-C," *Nucleic Acids Res.*, vol. 43, no. 11, pp. 5331–5339, 2015.

[172] M. W. Phillips and G. L. Gordon, "Growth characteristics on cellobiose of three different anaerobic fungi isolated from the ovine rumen.," *Appl. Environ. Microbiol.*, vol. 55, no. 7, pp. 1695–1702, 1989.

[173] A. Healey, A. Furtado, T. Cooper, and R. J. Henry, "Protocol : a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species," *Plant Methods*, vol. 10, no. 1, pp. 1–8, 2014.

[174] J. J. Doyle and J. L. Doyle, "A rapid DNA isolation procedure for small quantities of leaf tissue.pdf," *Phytochemical Bulletin*, vol. 19, no. 1. pp. 11–15, 1987.

[175] L. H. Hagen *et al.*, "Proteome specialization of anaerobic fungi during ruminal degradation of recalcitrant plant fiber," *ISME J.*, 2020.

[176] M. D. Mungan, M. Alanjary, K. Blin, T. Weber, M. H. Medema, and N. Ziemert, "ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining," *Nucleic Acids Res.*, vol. 48, no. W1, pp. W546–W552, 2020.

[177] K. M. Nielsen, T. Bøhn, and J. P. Townsend, "Detecting rare gene transfer events in bacterial populations," *Front. Microbiol.*, vol. 4, no. JAN, pp. 1–12, 2013.

[178] S. R. Ahrendt *et al.*, "Leveraging single-cell genomics to expand the fungal tree of life," *Nat. Microbiol.*, vol. 3, no. 12, pp. 1417–1428, 2018.

[179] R. J. Gruninger *et al.*, "Anaerobic fungi (phylum Neocallimastigomycota): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential," *FEMS Microbiol Ecol*, vol. 90, 2014.

[180] R. T. Todd, A. Forche, and A. Selmecki, "Ploidy Variation in Fungi: Polyploidy, Aneuploidy, and Genome Evolution," *The Fungal Kingdom*, pp. 599–618, 2017.

[181] S. C. Lee, M. Ni, W. Li, C. Shertz, and J. Heitman, "The Evolution of Sex: a Perspective from the Fungal Kingdom," *Microbiol. Mol. Biol. Rev.*, vol. 74, no. 2, pp. 298–340, 2010.

[182] M. Gao *et al.*, "Innovating a nonconventional yeast platform for producing shikimate as the building block of high-value aromatics," *ACS Synth. Biol.*, vol. 6, no. 1, pp. 29–38, 2017.

[183] S. E. Lowe, M. K. Theodorou, A. P. J. Trinci, and R. B. Hespell, "Growth of anaerobic rumen fungi on defined and semi-defined media lacking rumen fluid," *J. Gen. Microbiol.*, vol. 131, no. 9, pp. 2225–2229, 1985.

[184] E. Lieberman-Aiden *et al.*, "Comprehensive mapping of long-range interactions reveals folding principles of the human genome.," *Science*, vol. 326, no. 5950, pp. 289–293, Oct. 2009.

[185] H. Li and R. Durbin, "Fast and accurate long-read alignment with Burrows-Wheeler transform.," *Bioinformatics*, vol. 26, no. 5, pp. 589–595, Mar. 2010.

[186] G. G. Faust and I. M. Hall, "SAMBLASTER: Fast duplicate marking and structural variant read extraction," *Bioinformatics*, vol. 30, no. 17, pp. 2503–2505, 2014.

[187] H. Li *et al.*, "The Sequence Alignment/Map format and SAMtools.," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, Aug. 2009.

[188] Z. N. Kronenberg *et al.*, "Extended haplotype phasing of de novo genome assemblies with FALCON-Phase," *bioRxiv*, 2019.

[189] D. M. Bickhart *et al.*, "Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome.," *Nat. Genet.*, vol. 49, no. 4, pp. 643–650, Apr. 2017.

[190] J. N. Burton, A. Adey, R. P. Patwardhan, R. Qiu, J. O. Kitzman, and J. Shendure, "Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions.," *Nat. Biotechnol.*, vol. 31, no. 12, pp. 1119–1125, Dec. 2013.

[191] N. C. Durand *et al.*, "Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom.," *Cell Syst.*, vol. 3, no. 1, pp. 99–101, Jul. 2016.

[192] S. S. P. Rao *et al.*, "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping," *Cell*, vol. 159, no. 7, pp. 1665–1680, 2014.

[193] H. A. Crissman and G. T. Hirons, “Chapter 13 Staining of DNA in Live and Fixed Cells,” in *Flow Cytometry Second Edition, Part A*, vol. 41, Z. Darzynkiewicz, J. Paul Robinson, and H. A. Crissman, Eds. Academic Press, 1994, pp. 195–209.

[194] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, “KEGG as a reference resource for gene and protein annotation,” vol. 44, no. October 2015, pp. 457–462, 2016.

[195] M. J. Teunissen, E. P. W. Kets, H. J. M. Op den Camp, J. H. J. Huis in’t Veld, and G. D. Vogels, “Effect of coculture of anaerobic fungi isolated from ruminants and non-ruminants with methanogenic bacteria on cellulolytic and xylanolytic enzyme activities,” *Arch. Microbiol.*, vol. 157, no. 2, pp. 176–182, 1992.

[196] C. A. Hooker, K. Z. Lee, and K. V. Solomon, “Leveraging anaerobic fungi for biotechnology,” *Curr. Opin. Biotechnol.*, vol. 59, pp. 103–110, 2019.

[197] S. E. Wilken, M. Saxena, L. R. Petzold, and M. A. O’Malley, “In silico identification of microbial partners to form consortia with anaerobic fungi,” *Processes*, vol. 6, no. 1, pp. 1–14, 2018.

[198] A. K. Löbs, R. Engel, C. Schwartz, A. Flores, and I. Wheeldon, “CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in Kluyveromyces marxianus,” *Biotechnol. Biofuels*, vol. 10, no. 1, pp. 1–14, 2017.

[199] G. G. Fonseca, A. K. Gombert, E. Heinzle, and C. Wittmann, “Physiology of the yeast Kluyveromyces marxianus during batch and chemostat cultures with glucose as the sole carbon source,” *FEMS Yeast Res.*, vol. 7, no. 3, pp. 422–435, 2007.

[200] Y. Sakihama, R. Hidese, T. Hasunuma, and A. Kondo, “Increased flux in acetyl-CoA synthetic pathway and TCA cycle of Kluyveromyces marxianus under respiratory conditions,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–7, 2019.

[201] W. Mo, M. Wang, R. Zhan, Y. Yu, Y. He, and H. Lu, “Kluyveromyces marxianus developing ethanol tolerance during adaptive evolution with significant improvements of multiple pathways,” *Biotechnol. Biofuels*, vol. 12, no. 1, pp. 1–15, 2019.

[202] C. Löser, T. Urit, E. Gruner, and T. Bley, “Efficient growth of Kluyveromyces marxianus biomass used as a biocatalyst in the sustainable production of ethyl acetate,” *Energy. Sustain. Soc.*, vol. 5, no. 1, pp. 1–15, 2015.

[203] A. K. Löbs, J. L. Lin, M. Cook, and I. Wheeldon, “High throughput, colorimetric screening of microbial ester biosynthesis reveals high ethyl acetate production from Kluyveromyces marxianus on C5, C6, and C12 carbon sources,” *Biotechnol. J.*, vol. 11, no. 10, pp. 1274–1281, 2016.

[204] M. Li *et al.*, “CRISPR-mediated multigene integration enables Shikimate pathway refactoring for enhanced 2-phenylethanol biosynthesis in Kluyveromyces marxianus,” *Biotechnol. Biofuels*, vol. 14, no. 1, pp. 1–15, 2021.

[205] J. D. Keasling and H. Chou, “Metabolic engineering delivers next-generation biofuels Jay,” *Nat. Biotechnol.*, vol. 26, no. 3, pp. 298–299, 2008.

[206] M. İşleten Hoşoğlu, “Study of increasing the production of volatile flavor compounds by the yeasr Kluyveromyces marxianus through optimization of carbon and nitrogen sources,” *Food Heal.*, vol. 4, no. 2, pp. 112–123, 2018.

[207] J. Yuan, P. Mishra, and C. B. Ching, “Engineering the leucine biosynthetic pathway for isoamyl alcohol overproduction in Saccharomyces cerevisiae,” *J. Ind. Microbiol. Biotechnol.*, vol. 44, no. 1, pp. 107–117, 2017.

[208] A. S. Rajkumar, J. A. Varela, H. Juergens, J. M. G. Daran, and J. P. Morrissey, “Biological parts for Kluyveromyces marxianus synthetic biology,” *Front. Bioeng. Biotechnol.*, vol. 7, no. MAY, pp. 1–15, 2019.

[209] X. Lang, P. B. Besada-Lombana, M. Li, N. A. Da Silva, and I. Wheeldon, “Developing a broad-range promoter set for metabolic engineering in the thermotolerant yeast Kluyveromyces marxianus,” *Metab. Eng. Commun.*, vol. 11, no. April, p. e00145, 2020.

[210] G. G. Fonseca, E. Heinzle, C. Wittmann, and A. K. Gombert, “The yeast Kluyveromyces marxianus and its biotechnological potential,” *Appl. Microbiol. Biotechnol.*, vol. 79, no. 3, pp. 339–354, 2008.

[211] J. V. Vermaas, L. Petridis, X. Qi, R. Schulz, B. Lindner, and J. C. Smith, “Mechanism of lignin inhibition of enzymatic biomass deconstruction,” *Biotechnol. Biofuels*, vol. 8, no. 1, pp. 1–16, 2015.

[212] J. M. A. Geertman, J. P. Van Dijken, and J. T. Pronk, “Engineering NADH metabolism in Saccharomyces cerevisiae: Formate as an electron donor for glycerol production by anaerobic, glucose-limited chemostat cultures,” *FEMS Yeast Res.*, vol. 6, no. 8, pp. 1193–1203, 2006.

[213] T. R. Zuroff, S. B. Xiques, and W. R. Curtis, “Consortia-mediated bioprocessing of cellulose to ethanol with a symbiotic Clostridium phytofermentans/yeast co-culture,” *Biotechnol. Biofuels*, vol. 6, no. 1, pp. 1–12, 2013.

[214] M. Salmela, T. Lehtinen, E. Efimova, S. Santala, and V. Santala, “Towards bioproduction of poly-α-olefins from lignocellulose,” *Green Chem.*, vol. 22, no. 15, pp. 5067–5076, 2020.

[215] A. K. Löbs, C. Schwartz, S. Thorwall, and I. Wheeldon, “Highly Multiplexed CRISPRi Repression of Respiratory Functions Enhances Mitochondrial Localized Ethyl Acetate Biosynthesis in Kluyveromyces marxianus,” *ACS Synth. Biol.*, vol. 7, no. 11, pp. 2647–2655, 2018.

[216] S. E. Lowe, M. K. Theodorou, and A. P. J. Trinci, “Cellulases and xylanase of an anaerobic rumen fungus grown on wheat straw, wheat straw holocellulose, cellulose, and xylan,” *Appl. Environ. Microbiol.*, vol. 53, no. 6, pp. 1216–1223, 1987.

[217] T. Perli *et al.*, *Identification of oxygen-independent pathways for pyridine-nucleotide and Coenzyme-A synthesis in anaerobic fungi by expression of candidate genes in yeast*. 2020.

[218] E. Russo, "Special Report: The birth of biotechnology," *Nature*, vol. 421, no. January, pp. 456–457, 2003.

[219] H. Chi *et al.*, "Engineering and modification of microbial chassis for systems and synthetic biology," *Synth. Syst. Biotechnol.*, vol. 4, no. 1, pp. 25–33, 2019.

[220] J. Athey *et al.*, "A new and updated resource for codon usage tables," *BMC Bioinformatics*, vol. 18, no. 1, pp. 1–10, 2017.

[221] F. Bernaudat *et al.*, "Heterologous expression of membrane proteins: Choosing the appropriate host," *PLoS One*, vol. 6, no. 12, 2011.

[222] F. J. Fernández and M. C. Vega, "Choose a Suitable Expression Host: A Survey of Available Protein Production Platforms," in *Advanced Technologies for Protein Complex Production and Characterization*, M. C. Vega, Ed. Cham: Springer International Publishing, 2016, pp. 15–24.

[223] S. R. Maloy, *Genetic analysis of pathogenic bacteria : a laboratory manual*. Plainview, N.Y.: Cold Spring Harbor Laboratory Press, 1996.

[224] V. Wood *et al.*, "The genome sequence of Schizosaccharomyces pombe," *Nature*, vol. 421, no. 94, 2003.

[225] J. C. Guimaraes *et al.*, "A rare codon-based translational program of cell proliferation," *Genome Biol.*, vol. 21, no. 1, pp. 1–20, 2020.

[226] R. Jansen, H. J. Bussemaker, and M. Gerstein, "Revisiting the codon adaptation index from a whole-genome perspective: Analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models," *Nucleic Acids Res.*, vol. 31, no. 8, pp. 2242–2251, 2003.

[227] B. D. Lee, "Python Implementation of Codon Adaptation Index," *J. Open Source Softaware*, vol. 30, no. 3, pp. 356–362, 2018.

[228] B. B. Khomtchouk and W. Nonner, "Gaussian-distributed codon frequencies of genomes," *G3 Genes, Genomes, Genet.*, vol. 9, no. 5, pp. 1449–1456, 2019.

[229] T. E. F. Quax, N. J. Claassens, D. Söll, and J. van der Oost, "Codon Bias as a Means to Fine-Tune Gene Expression," *Mol. Cell*, vol. 59, no. 2, pp. 149–161, 2015.

[230] S. E. Wilken *et al.*, "Genomic and proteomic biases inform metabolic engineering strategies for anaerobic fungi," *Metab. Eng. Commun.*, vol. 10, no. October 2019, 2020.

[231] J. Peden, "Analysis of codon usage," 2000.

[232] J. L. Chaney *et al.*, "Widespread position-specific conservation of synonymous rare codons within coding sequences," *PLoS Comput. Biol.*, vol. 13, no. 5, pp. 1–19, 2017.

[233] P. Mittal, J. Brindle, J. Stephen, J. B. Plotkin, and G. Kudla, "Codon usage influences fitness through RNA toxicity," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 34, pp. 8639–8644, 2018.

[234] G. Boël *et al.*, "Codon influence on protein expression in E. coli correlates with mRNA levels," *Nature*, vol. 529, no. 7586, pp. 358–363, 2016.

[235] P. Liao, A. Hemmerlin, T. J. Bach, and M. L. Chye, "The potential of the mevalonate pathway for enhanced isoprenoid production," *Biotechnol. Adv.*, vol. 34, no. 5, pp. 697–713, 2016.

[236] V. J. J. Martin, D. J. Piteral, S. T. Withers, J. D. Newman, and J. D. Keasling, "Engineering a mevalonate pathway in Escherichia coli for production of terpenoids," *Nat. Biotechnol.*, vol. 21, no. 7, pp. 796–802, 2003.

[237] I. Frumkin, M. J. Lajoie, C. J. Gregg, G. Hornung, G. M. Church, and Y. Pilpel, "Codon usage of highly expressed genes affects proteome-wide translation efficiency," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 21, pp. E4940–E4949, 2018.

[238] P. Xu, A. Vansiri, N. Bhan, and M. A. G. Koffas, "EPathBrick: A synthetic biology platform for engineering metabolic pathways in E. coli," *ACS Synth. Biol.*, vol. 1, no. 7, pp. 256–266, 2012.

[239] M. J. McDonald, C. H. Chou, K. B. S. Swamy, H. Da Huang, and J. Y. Leu, "The evolutionary dynamics of tRNA-gene copy number and codon-use in E. coli.," *BMC Evol. Biol.*, vol. 15, no. 1, pp. 1–10, 2015.

[240] E. Angov, C. J. Hillier, R. L. Kincaid, and J. A. Lyon, "Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host," *PLoS One*, vol. 3, no. 5, pp. 1–10, 2008.

[241] L. Kizer, D. J. Pitera, B. F. Pfleger, and J. D. Keasling, "Application of functional genomics to pathway optimization for increased isoprenoid production," *Appl. Environ. Microbiol.*, vol. 74, no. 10, pp. 3229–3241, 2008.

[242] A. Krogh, B. Larsson, G. Von Heijne, and E. L. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes," *J. Mol. Biol.*, vol. 305, no. 3, pp. 567–580, 2001.

[243] S. M. Ma *et al.*, "Optimization of a heterologous mevalonate pathway through the use of variant HMG-CoA reductases," *Metab. Eng.*, vol. 13, no. 5, pp. 588–597, 2011.

[244] J. A. Jones, T. D. Toparlak, and M. A. G. Koffas, "Metabolic pathway balancing and its role in the production of biofuels and chemicals," *Curr. Opin. Biotechnol.*, vol. 33, pp. 52–59, 2015.

[245] J. A. Jones *et al.*, "EPathOptimize: A combinatorial approach for transcriptional balancing of metabolic pathways," *Sci. Rep.*, vol. 5, pp. 1–10, 2015.

[246] J. Yang *et al.*, "Enhancing production of bio-isoprene using hybrid MVA pathway and isoprene synthase in E. coli," *PLoS One*, vol. 7, no. 4, pp. 1–7, 2012.

[247] H. Nordberg *et al.*, "The genome portal of the Department of Energy Joint Genome Institute: 2014 updates," *Nucleic Acids Res.*, vol. 42, no. D1, pp. 26–31, 2014.

[248] C. Camacho *et al.*, "BLAST+: architecture and applications," *BMC Bioinformatics*, vol. 10, no. 1, p. 421, 2009.

[249] P. J. A. Cock *et al.*, "Biopython: Freely available Python tools for computational molecular biology and bioinformatics," *Bioinformatics*, vol. 25, no. 11, pp. 1422–1423, 2009.

[250] S. Li, C. B. Jendresen, and A. T. Nielsen, "Increasing production yield of tyrosine and mevalonate through inhibition of biomass formation," *Process Biochem.*, vol. 51, no. 12, pp. 1992–2000, 2016.

[251] J. K. Henske *et al.*, "Transcriptomic characterization of Caecomyces churrovis: A novel, non-rhizoid-forming lignocellulolytic anaerobic fungus," *Biotechnol. Biofuels*, vol. 10, no. 1, pp. 1–12, 2017.

[252] I. A. Podolsky, S. Seppälä, T. S. Lankiewicz, J. L. Brown, C. L. Swift, and M. A. O'Malley, "Harnessing Nature's Anaerobes for Biotechnology and Bioprocessing," *Annu. Rev. Chem. Biomol. Eng.*, vol. 10, no. 1, pp. 105–128, 2019.

[253] I. Liachko, R. A. Youngblood, U. Keich, and M. J. Dunham, "High-resolution mapping, characterization, and optimization of autonomously replicating sequences in yeast," *Genome Res.*, vol. 23, no. 4, pp. 698–704, 2013.

[254] M. Cao, A. S. Seetharam, A. J. Severin, and Z. Shao, "Rapid Isolation of Centromeres from Scheffersomyces stipitis," *ACS Synth. Biol.*, vol. 6, no. 11, pp. 2028–2034, 2017.

[255] A. D. Garst *et al.*, "Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering," *Nat. Biotechnol.*, vol. 35, no. 1, pp. 48–55, 2017.

[256] M. S. E. Shelby Calkins, Nicole C. Elledge, Radwa A. Hanafy and N. Youssef, "A fast and reliable procedure for spore collection from anaerobic fungi - Application for RNA uptake and long-term storage of isolates.pdf," *J. Microbiol. Methods*, vol. 127, no. 2016, pp. 206–213, 2016.

[257] H. E. Chia, T. Zuo, N. M. Koropatkin, E. N. G. Marsh, and J. S. Biteen, "Imaging living obligate anaerobic bacteria with bilin-binding fluorescent proteins," *Curr. Res. Microb. Sci.*, vol. 1, no. April, pp. 1–6, 2020.

[258] A. Mukherjee, J. Walker, K. B. Weyant, and C. M. Schroeder, "Characterization of Flavin-Based Fluorescent Proteins: An Emerging Class of Fluorescent Reporters," *PLoS One*, vol. 8, no. 5, 2013.

[259] D. M. Shcherbakova and V. V. Verkhusha, "Near-infrared fluorescent proteins for multicolor in vivo imaging," *Nat. Methods*, vol. 10, no. 8, pp. 751–754, 2013.

[260] I. Liachko, A. Bhaskar, C. Lee, S. C. C. Chung, B. K. Tye, and U. Keich, "A comprehensive genome-wide map of autonomously replicating sequences in a naive genome," *PLoS Genet.*, vol. 6, no. 5, p. 22, 2010.

[261] M. K. Chaveroche, J. M. Ghigo, and C. d'Enfert, "A rapid method for efficient gene replacement in the filamentous fungus Aspergillus nidulans.," *Nucleic Acids Res.*, vol. 28, no. 22, pp. 1–6, 2000.

[262] Y. Gao and Y. Zhao, "Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing," *J. Integr. Plant Biol.*, vol. 56, no. 4, pp. 343–349, 2014.

[263] C. Zhang, X. H. Xing, and K. Lou, "Rapid detection of a gfp-marked Enterobacter aerogenes under anaerobic conditions by aerobic fluorescence recovery," *FEMS Microbiol. Lett.*, vol. 249, no. 2, pp. 211–218, 2005.

[264] K. P. Scott, D. K. Mercer, L. A. Glover, and H. J. Flint, "The green fluorescent protein as a visible marker for lactic acid bacteria in complex ecosystems," *FEMS Microbiol. Ecol.*, vol. 26, no. 3, pp. 219–230, 1998.

[265] I. Liachko and M. J. Dunham, "An autonomously replicating sequence for use in a wide range of budding yeasts," *FEMS Yeast Res.*, vol. 14, no. 2, pp. 364–367, 2014.

[266] S. H. Kung, A. C. Retchless, J. Y. Kwan, and R. P. P. Almeida, "Effects of DNA size on transformation and recombination efficiencies in xylella fastidiosa," *Appl. Environ. Microbiol.*, vol. 79, no. 5, pp. 1712–1717, 2013.

[267] M. Fischer, R. Durand, and M. Fèvre, "Characterization of the 'promoter region' of the enolase-encoding gene enol from the anaerobic fungus Neocallimastix frontalis: Sequence and promoter analysis," *Curr. Genet.*, vol. 28, no. 1, pp. 80–86, 1995.

[268] Q. Wang, P. A. Cobine, and J. J. Coleman, "Efficient genome editing in Fusarium oxysporum based on CRISPR/Cas9 ribonucleoprotein complexes," *Fungal Genet. Biol.*, vol. 117, no. May, pp. 21–29, 2018.

[269] R. A. Hanafy, B. Johnson, N. H. Youssef, and M. S. Elshahed, "Assessing anaerobic gut fungal diversity in herbivores using D1/D2 large ribosomal subunit sequencing and multi-year isolation," *Environ. Microbiol.*, vol. 22, no. 9, pp. 3883–3908, 2020.

[270] R. A. Hanafy *et al.*, “Seven new Neocallimastigomycota genera from wild, zoo-housed, and domesticated herbivores greatly expand the taxonomic diversity of the phylum,” *Mycologia*, vol. 112, no. 6, pp. 1212–1239, 2020.

[271] C. J. B. Harvey *et al.*, “HEx: A heterologous expression platform for the discovery of fungal natural products,” *bioRxiv*, no. April, 2018.

[272] E. T. Hillman, L. R. Readnour, and K. V. Solomon, “Exploiting the natural product potential of fungi with integrated -omics and synthetic biology approaches,” *Curr. Opin. Syst. Biol.*, vol. 5, no. Figure 1, pp. 50–56, 2017.

[273] K. V Solomon, T. M. Sanders, and K. L. J. Prather, “A dynamic metabolite valve for the control of central carbon metabolism,” *Metab. Eng.*, vol. 14, no. 6, pp. 661–671, 2012.

[274] J. M. Callura, D. J. Dwyer, F. J. Isaacs, C. R. Cantor, and J. J. Collins, “Physiology Using Synthetic Riboregulators,” *Proc. Natl. Acad. Sci.*, vol. 107, no. 36, pp. 15898–15903, 2010.

PUBLICATIONS

E. T. Hillman, M. Li, C. A. Hooker, J. A. Englaender, I. Wheeldon, and K. V. Solomon "Hydrolysis of lignocellulose by anaerobic fungi produces free sugars and organic acids for two-stage fine chemical production with *Kluyveromyces marxianus*", *Journal of Biotechnology Progress*. (Submitted)

E. T. Hillman, C. A. Hooker, J. Munoz, J. Holmes, R. Slaughter, I. Liachko, K. Berry, S. Mondo, and K. V. Solomon, "Complete genomes of lignocellulolytic anaerobic fungi pave the way for genetic tool development," *Biotechnology for Biofuels.* (In preparation)

E. T. Hillman, E. Frazier, E. Shank, A. N. Ortiz-Velez, J. A. Englaender, and K. V. Solomon, "Codon-optimization enabled heterologous expression of anaerobic fungal mevalonate pathway in E. coli," *Metabolic Engineering Communications*. (In preparation)

E. T. Hillman *et al.*, "Comparative genomics of the genus Roseburia reveals divergent biosynthetic pathways that may influence colonic competition among species," *Microb. genomics*, vol. 6, no. 7, 2020.

E. D. Lee, E. R. Aurand, D. C. Friedman, and Engineering Biology Research Consortium Microbiomes Roadmapping Working Group*, "Engineering Microbiomes—Looking Ahead," *ACS Synth. Biol*., vol. 9, no. 12, pp. 3181–3183, 2020

R. M. RedCorn, E. T. Hillman, K. V Solomon, and A. S. Engelberth, "Xanthobacter-dominated biofilm as a novel source for high-value rhamnose," *Appl. Microbiol. Biotechnol.*, vol. 103, no. 11, pp. 4525–4538, 2019.

C. A. Hooker, E. T. Hillman, *et al.*, "Hydrolysis of untreated lignocellulosic feedstock is independent of S-lignin composition in newly classified anaerobic fungal isolate, *Piromyces sp. UH3-1*," *Biotechnol. Biofuels*, vol. 11, no. 1, pp. 1–14, 2018.

E. T. Hillman, L. R. Readnour, and K. V. Solomon, "Exploiting the natural product potential of fungi with integrated -omics and synthetic biology approaches," *Curr. Opin. Syst. Biol.*, vol. 5, no. Figure 1, pp. 50–56, 2017.

E. T. Hillman, H. Lu, T. Yao, and C. H. Nakatsu, "Microbial ecology along the gastrointestinal tract," *Microbes Environ.*, vol. 32, no. 4, 2017.