IDENTIFICATION OF CELLULOLYTIC HOT SPRING ORGANISMS THROUGH BIOORTHOGONAL LABELING

by

Nicholas John Reichart

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2021

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Roland Hatzenpichler and my committee members Drs. Valérie Copié, Matthew Fields, Brent Peyton, and James Wilking. Without their insight, support, and guidance, none of this work would have been possible.

I would like to acknowledge all my colleagues and collaborators at Montana State University both in the Hatzenpichler lab and across campus. Countless rounds of experiments were discussed and developed with their help. And the many miles logged together hiking throughout Yellowstone, truly enriched my studies. Drs. Viola Krukenberg, Rachel Spietz, and Zack Jay were instrumental in helping me along with my experiments and I owe them a great debt of gratitude. A big thank you to Adam Thuen for motivating me to sticking with graduate school and succeeding together. Thank you to Dr. Tanja Woyke at the Joint Genome Institute for giving me an opportunity to conduct a year of my research at a National Lab, and Dr. Robert Bowers for teaching me all I know about computational biology.

Most importantly, I need to thank my family and friends back in Maryland. They believed in me more than I could ever do myself. My parents have always moved mountains to help me succeed and graduate school was no different. The best thing to come out of my graduate studies was my connection with my fiancé Juliana Beauchene. Her support and encouragement was truly significant in me being able to complete my studies.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

Tables Page

LIST OF FIGURES

Figures Page

GLOSSARY OF ABBREVIATIONS

AA: Auxiliary activity

ABPP: Activity based protein profiling

ASV: Amplicon sequence variant

BONCAT: Bioorthogonal non-canonical amino acid tagging

BSC-A: Back scatter – area

CAZyme: Carbohydrate-active enzyme

CBM: Carbohydrate-binding module

CBP: Consolidated bioprocessing

CE: Carbohydrate esterase

CPR: Candidate phyla radiation

DAPI: 4,6-Diamidino-2-phenylindole

EC: Enzyme classification

FACS: Fluorescence-activated cell sorting

FISH: Fluorescence *in situ* hybridization

FSC-A: Forward scatter – area

FSC-H: Forward scatter – height

FSC-W: Forward scatter – weight

GH: Glycosyl hydrolase

GT: Glycosyl transferase

HPG: Homopropargylglycine

LC-MS: Liquid chromatography – mass spectrometry

GLOSSARY OF ABBREVIATIONS CONTINUED

LFC: Log fold change

MAG: Metagenome assembled genome

MANOVA: Multivariate analysis of variance

nanoSIMS: Nanoscale secondary ion mass spectrometry

OTU: Operational taxonomic unit

PBS: Phosphate buffered saline

PCR: Polymerase chain reaction

PERMANOVA: Permutational multivariate analysis of variance

PL: Polysaccharide lyase

RFU: Relative fluorescence unit

SIP: Stable isotope probing

THPTA: Tris-hydroxypropyltriazolylmethylamine

YNP: Yellowstone National Park

ABSTRACT

Microbial physiology is the study of the metabolism and function of microorganisms. The recent expansion of genomic diversity has outpaced the description of physiology. To better understand microbial metabolisms and environmental processes, more detailed research is needed for both novel and undescribed microbes. While many new methods are being developed to describe *in situ* microbial activity, this dissertation implements bioorthogonal non-canonical amino acid tagging as a proxy to track metabolic activity of microbes under close to environment conditions. Using differential analyses on hot spring microbial communities, we were able to show that certain microbial taxa had preferential activity towards specific incubation amendments. Previous activity-based studies had shown that hot springs were a unique environment for discovering cellulolytic microbes that could be used in industrial processing of plant biomass. Herein, we used computational analysis to screen publicly available metagenomic datasets to identify the enzymatic potential of hot springs worldwide. The wide diversity of taxa and biomass degrading enzymes were investigated and hot springs were further highlighted as a system that could be used to find improvement for the industry of plant biomass degradation and processing. To build upon the cellulolytic potential found in hot spring metagenomic datasets, bioorthogonal non-canonical amino acid tagging coupled with fluorescence-activated cell sorting was applied to the biotechnological relevant field of plant biomass degradation to identify microbes involved in the cellulolytic process. Examination of the active microbes revealed difference in the community when supplemented with cellulose. Taken together, the work in this dissertation served to expand and apply the recent development of activity-based studies used to describe environmental microbial populations, with a focus on plant biomass degradation.

CHAPTER ONE

INTRODUCTION

This introduction serves to highlight the struggles of describing the function of microbes thus far using studies in the realm of multi-omic research. Recent studies of lignocellulose degradation from hot springs are reviewed to highlight the importance of this environment. Following the development and application of this bioorthogonal approach, a functional aspect directed towards plant biomass conversion into biofuels is discussed. The main objectives and brief overview of the following chapters of this thesis are provided as a conclusion to the introduction.

Microbial diversity through gene sequencing and metagenomics

The advent and subsequent development of metagenomic sequencing has greatly expanded the known diversity of environmental microorganisms. Efforts to describe the physiology and *in situ* activity of these microorganisms have not kept pace with their discovery. Traditionally, physiology was described with lab-based techniques using cultures of microbes that could be manipulated and visualized. However, genomic studies have uncovered a vast number of microbial lineages that have resisted cultivation efforts thus far (Lloyd et al., 2018). Early gene sequencing studies focused on sequencing short regions of a functional marker gene or ribosomal RNA gene to determine the identification of taxa in an environmental sample, but these studies can be biased by primer-based amplification (Parada et al., 2016; Karst et al., 2018). Considerable advancements have been made, including primer-less methods (Karst et al., 2018), but most amplicon sequencing studies still rely on primers to match to a wide diversity of

targets and have to manage with the inability to work with potential introns or missing regions within genomic material. Amplicon studies have greatly increased the known diversity of taxa, however, large branches of the tree of life were still missed, later to be discovered by metagenomics. Development of shotgun metagenomics allowed sequencing of environmental DNA without the biases of amplicon sequencing and helped to identify additional taxonomic diversity that had eluded detection (Eloe-Fadrosh et al., 2016a). Shotgun metagenomics and single-cell genomics were able to detect novel taxa in environments previously well-studied with amplicon sequencing, but these metagenomic studies lacked functional data and could only provide genomic potential (Rinke et al., 2013; Hedlund et al., 2014). Genomic studies can reveal metabolic pathways or nutrient requirements of uncultured microorganisms but most predictions have not yet been experimentally tested (Beam et al., 2016; Saxena et al., 2017; Reichart et al., 2021). The expansion of taxa from genomic sequencing generated large, novel branches of the tree of life such as the Candidate Phyla Radiation (CPR; (Hug et al., 2016)) but these microorganisms could only be described with inferred function as most members have yet eluded cultivation. Better understanding of *in situ* activity and function beyond genomic potential would be possible with cultivation but it is currently thought that approximately one quarter of the total microbial diversity do not have a cultured representative (Lloyd et al., 2018). Culture medium deficiencies can prevent cultivation as unknown symbiotic connections or cross feeding processes in the environment could be supplementing certain taxa's metabolic pathways that are no longer present when in the lab setting (Eloe-Fadrosh et al., 2016b). Additionally, yet uncultured microbes could be present in low abundances or have long generation times, reducing their amenability in cultivation efforts. Recently, cutting edge methods and approaches have

been developed concertedly with sequencing information to better study the *in situ* function of environmental microbes without the need for *a priori* cultivation.

Omic methods for microbial description of physiology

A multitude of techniques have been used to better study environmental samples in efforts to describe functional activity. In addition to metagenomics, other omic methods, including metatranscriptomics, metaproteomics, and metametabolomics, allow for investigation of a community's mRNA, proteins, and metabolites, respectively. Integrated multi-omics can provide a thorough scene of the predicted and functional potential of a microbial community. However, omic studies all have drawbacks, including the identification of the results is database dependent and the ability to link physiology to taxonomy is often muddled in community context. Metatranscriptomics fills an information gap left by metagenomics by providing data of mRNA to study genes that are being expressed in a system. This allows studies to analyze genes differentially under various conditions (Bang-Andreasen et al., 2020) or to reconstruct metabolic pathways that are being expressed (Beam et al., 2016). Metatranscriptomic studies have provided important information on the expression of enzymes that help define an organism's environmental niche. Increased expression levels of particular genes can suggest possible respiration and necessary environmental conditions such as oxygen requirements (Beam et al., 2016) or potential electron acceptors (Jay et al., 2018). Proteomics is used to analyze proteins that are in the sample either intra- or extracellularly, depending on sample preparation. With proteomics, functional interactions can be determined under changing conditions (Heyer et al., 2013). Alternatively, chemical metabolites can be examined with metabolomics in concert with metagenomics to connect genomically identified microbes and gene potential to environmental

metabolic products in the process of multi-omics. When used together, multi-omic studies can track a community's activity from potential function through expression and enzymatic action of the microbial members. A recent study into a synthetic mixed-species biofilm of *Streptococcus mutans* and *Candida albicans,* that represents dental caries, utilized transcriptomics and proteomics to better understand the synergistic networks therein. Carbohydrate metabolism genes were upregulated compared to monocultures and extracellular proteins from *S. mutans* improved sugar metabolism for *C. albicans* increasing overall growth of the mixed-species biofilm (Ellepola et al., 2019). An additional study used the combination of metagenomics and metabolomics to describe the functional differences of ruminants in dairy cows with either low or high milk production. The microbial populations and genomic potential were different between the two cow types and increases in certain small metabolites and end products were correlated with increased milk production (Xue et al., 2020). While these studies reflect multi-omic approaches, detailed data to link function to specific microbes can be improved upon with more directed single-cell activity studies.

Single-cell activity studies

Single-cell activity methods build upon previously discussed multi-omic studies by being able to separate individual cells to describe their function detached from the community. Multiple studies have been published over the past decade reviewing single-cell techniques conceptually (Wagner, 2009; Singer et al., 2017; Hatzenpichler et al., 2020) but so far, no studies have directly compared microbial activity of the different methods with a standardized mock community sample. Each approach has unique advantages and disadvantages and when used in combination, a thorough description of microbial function can be achieved.

Nanoscale secondary ion mass spectroscopy

Nanoscale secondary ion mass spectroscopy (nanoSIMS) is a destructive method in which an ion beam, typically cesium, is used to ablate particles from a microbial cell that are then identified by mass spectroscopy. Using stable isotopes during incubation, heavy isotopes can be identified with nanoSIMS and functional data of the cell, through incorporation of the heavy isotope, can be deduced. NanoSIMS has been used to investigate nitrogen fixation occurring in anaerobic methanotrophic archaea group 2 (ANME-2) and *Desulfosarcina/Desulfococcus* bacterial (DSS) consortium, attributing the function of fixation to ANME-2 cells through assimilation of ^{15}N from various nitrogen sources (Dekas et al., 2009).

Raman spectroscopy

Stable isotopes can also be used in conjunction with Raman spectroscopy, a nondestructive technique. Raman spectroscopy generates unique spectral fingerprints of the chemical composition of a cell by measuring different wavelengths of scattered light upon excitation by a laser. These Raman spectra can be used to distinguish different chemical compounds and among various microbial species (Liu et al., 2020). Garcia-Timermans and colleagues (García-Timermans et al., 2020) were able to analyze the Raman spectral shifts between normal and stressed cells and revealed a difference in metabolic expression, specifically, cells synthesized more proteins and less nucleic acids when stressed. When stable isotopes were incubated with a sample, the Raman wavelength chemical fingerprint of the cell shifts, indicating assimilation of the heavy isotope. Increasing the concentration of ^{13}C glucose in an incubation of *Pseudomonas fluorescens*, progressively decreased the wavenumber of the Raman spectra of several peaks compared to incubations with only ^{12}C glucose (Huang et al.,

2004). The nonspecific cellular incorporation of deuterium oxide (D_2O) can be used without the need for substrate specific heavy isotopes, such as ^{13}C glucose or ^{15}N ammonia, to analyze the overall activity of a cell (Berry et al., 2015). Fluorescence *in situ* hybridization (FISH) can be used to taxonomically identify cells when used with both nanoSIMS and Raman, but the cells analyzed by Raman are not destroyed and can be subsequently sorted with optical tweezers for downstream genomics. While nanoSIMS and Raman spectroscopy can provide detailed information on isotope incorporation at a fine resolution, the processing throughput and expensive instrumentation and reagents restrict wider implementation of these approaches.

Hook and bait capture

Additional methods to determine microbial metabolic activity have recently been developed that are more targeted to specific microbial functions. Identifying cellulose degrading microbes was the focus of a recent study that utilized fluorescent cellulose particles and fluorescence-activated cell sorting (FACS) to separate microbes that were bound to the cellulose particles (Doud et al., 2019). This “hook and bait” method is extremely specific to cellulosome containing microbes that can stay bound to cellulose particles. For example, after sorting 157 cellulose particles, 100 particles contained only one cell as determined by single-cell sequencing. The sorted population was significantly less diverse than the original hot spring sample diversity, selecting for thermophilic cellulose degraders (Doud et al., 2019). Overall, rare, low abundant environmental taxa were enriched through the hook and bait method and high-quality genomes sequenced from the sorted single-cells could be constructed to determine novel genes used for cellulose degradation (Doud et al., 2019).

Activity based protein profiling

Activity based protein profiling (ABPP) is a directed, functional approach to link probes onto proteins of interest before attaching a fluorescent dye via click chemistry and subsequently sorting fluorescent cells apart from nonfluorescent cells or directly analyzing extracellular proteins. ABPP can be accomplished using three types of probes that either target catalytic enzymes, protein interactions, or regulatory modification (Sadler and Wright, 2015). An advantage to using ABPP is the ability to target a specific enzyme class with a desired function. Chitin degradation was targeted and the expression of chitinases belonging to several glycosyl hydrolase families were tracked over time in combination with proteomic analysis after extracting the secretome of *Cellvibrio japonicus* cultures (Zegeye et al., 2020). ABPP is limited by the need to synthesize the unique probes for different enzymes or functions and in the case of chitinases, binding site promiscuity of glycosyl hydrolases caused unexpected labeling of non-chitinase enzymes (Zegeye et al., 2020).

Bioorthogonal labeling

Bioorthogonal non-canonical amino acid tagging (BONCAT) is one additional method that has gained traction in microbial functional studies that is possible without expensive or hard to synthesize compounds. BONCAT uses the incorporation of a synthetic amino acid, an analog of methionine, into newly synthesized proteins through the binding promiscuity of methionine tRNA synthetase. Incorporated synthetic amino acids can later be identified ("tagged") through azide-alkyne copper-catalyzed cycloaddition click chemistry (Kiick et al., 2002) to bind a fluorescent dye onto the synthetic amino acid. An alkyne group and an azide group are required on either the synthetic amino acid or fluorescent dye for bioconjugation using click chemistry to

create a triazole structure. Labeled proteins, extracted from cells, can be analyzed using liquid chromatography mass spectrometry (LC-MS) to identify functional changes within cells (Dieterich et al., 2006). BONCAT incubations have also been shown to not influence the metabolome of microbial cells (Steward et al., 2020) but had significant impact on the proteome of *Synechococcus* cultures (Michels et al., 2021). Active, BONCAT labeled cells can be sorted from inactive, nonfluorescent cells using FACS, similar to ABPP and laser tweezer usage with Raman, and then identified using 16S rRNA gene amplicon sequencing (Hatzenpichler et al., 2016; Couradeau et al., 2019; Reichart et al., 2020). This workflow allows for high throughput microbial taxonomic identification compared to FISH identification which requires *a priori* knowledge of the microbial community. BONCAT has been employed to analyze the fraction of active microbes in various environments coupled to FACS (Hatzenpichler et al., 2016; Couradeau et al., 2019) or over time series when analyzing fluorescence intensity (Leizeaga et al., 2017). The amount of cells and the fluorescence intensity of translationally active cells in a culture peaked at the end of the exponential growth phase and gradually decreased during stationary phase (Leizeaga et al., 2017). This shows that BONCAT can be used to track the physiological levels of microbes in a sample. Incorporation of the synthetic amino acid into proteins occurs rapidly, allowing slow growing microbes, such as anammox bacteria within batch reactor incubations, to be visualized via fluorescence microscopy within 3 % of their generation time (Chen et al., 2021). The time needed for anammox bacteria to become BONCAT labeled was in support with previously established labeling rates in *Escherichia coli* (2 % of generation time) (Hatzenpichler et al., 2014). Recently, studies using BONCAT have begun to focus more on comparing activity under different conditions, such as soil depths (Couradeau et

al., 2019) or cystic fibrosis lung sputum (Valentini et al., 2020). BONCAT active cells were shown to be a subset of the overall soil community, constituting 20-75 % of total cells in a sample depending on the soil depth and conditions selected (Couradeau et al., 2019). Indeed, BONCAT cells represent a subset of the total microbial community as was also determined in cystic fibrosis sputum where, depending on the subject, 5-56% of cells were BONCAT labeled (Valentini et al., 2020). Heterogeneity of activity in a sample is seen both across multiple microbial lineages but also more specifically within microbial genera when analyzing BONCAT-FACS positive, labeled populations and negative, unlabeled populations.

Differential analysis can be used to compare the BONCAT active fractions of microbial populations under various conditions. Comparing amended incubations to close to *in situ* conditions can allow for detection of microbial activity that was in response to substrate or condition amendment, as was seen in two recent studies using hot spring sediment and human gut microbiota (Reichart et al., 2020; Riva et al., 2020, respectively). Expanding on Riva and colleagues' (Riva et al., 2020) work with BONCAT to detect flavonoid utilization, BONCAT has also been used to investigate specific metabolisms or functions that carry biotechnological importance, such as cellulose degradation (Reichart et al., 2020), without the need for expensive or hard to synthesize compounds.

Cellulose degradation

Cellulose is the most abundant renewable polymer on Earth and its efficient, cost-effective conversion into bioproducts would be beneficial for many areas of biotechnology. Cellulose degradation is a complex process that requires a multitude of enzymes. Cellulose can be degraded into glucose which can then be fermented for biofuels and various other products of

degradation are used in the paper milling and textile industries. Plant cells walls are made of lignocellulose, which is comprised of three main components: lignin, hemicellulose, and cellulose. Lignin is the outermost polymer and contains branched alcohols linked by ether and carbon-carbon bonds. Ratios of coniferyl, syringyl, and hydroxyphenyl alcohol components in lignin vary among plant species (Bugg et al., 2011). Hemicellulose intertwines with lignin and connects to cellulose. Hemicellulose contains branches of pentose and hexose sugars from a xylose backbone. There are multiple sugars in hemicellulose, however, xylose and mannose are most common. At the center of lignocellulose is cellulose. Cellulose is repeating units of cellobiose, arranged as *D*-Glucose-β-1,4-*D*-Glucose, that are bound tightly together through hydrogen bonds to form crystalline fibrils. Cellulose fibrils can also form amorphous regions, where the structure lacks crystallinity and is more irregular in nature and thus provide more binding opportunities for cellulases. Each component of lignocellulose is unique in structure and requires specific enzymes for degradation.

Prior to enzymatic degradation of cellulose for industrial processing, plant biomass must be pretreated to remove unwanted components such as lignin and to increase the accessibility to cellulose. Lignin degradation produces a variety of toxic products, necessitating its removal prior to subsequent degradation steps. Pretreatment provides more access to the recalcitrant fibrils of cellulose and decreases the degree of polymerization and crystallinity of cellulose, improving hydrolysis. Mechanical, chemical, and biological are the three main categories of lignocellulose pretreatment. Mechanical pretreatment is often grinding or milling of the plant biomass to increase available surface area for later enzymatic processing. Chemical pretreatment utilizes various solutions and pH ranges. Acidic conditions can be used to target portions of

hemicellulose, whereas alkaline pretreatment is more effective at removing lignin components (He et al., 2018). Biological pretreatment mostly targets removing lignin using oxidases and laccases from fungi (Rai et al., 2019), however this process is viewed as time consuming on an industrial scale (Agbor et al., 2011). Combinations of multiple pretreatment methods have been shown to be advantageous and are of great interest in finding the most efficient and cost-effective method of cellulose degradation for industrial processing.

Enzymatic mechanism of degradation

Biological degradation of lignocellulose occurs through the catalytic mechanisms of multiple enzyme types. The Carbohydrate Active Enzyme (CAZy) database provides sequence-based classification of enzyme activity involved in lignocellulose degradation as well as for other carbohydrate associated proteins (Lombard et al., 2014). Lignin is degraded using laccases (Enzyme Classification, EC 1.10.3.2) and various peroxidases (EC 1.11.1.14) in the auxiliary activity (AA) class of CAZymes. Many AA enzymes require oxygen and thus are not often seen in typical microbial-based degrading environments such as compost piles or rumen. Fungal peroxidases from white and brown rot fungi are well-studied and characterized AA enzymes. While, most often, the lignin portion of lignocellulose is incinerated for energy generation, in other processing scenarios, the alcohol groups composing lignin can be extracted for various biotechnologically relevant products. Increasing the amount of usable products from lignin degradation can also aid in improving the cost yield of biomass conversion. Products of lignin degradation can be toxic to cells so large-scale processing must remove these products in an efficient manner to maintain biological production rates. Lignin can be used to generate phenolic compounds as well as benzene, toluene, and xylene which have a large market cap potential.

Because of the difficulties of lignin removal and toxicity and the current low biotech production output, most lignin is removed using chemical or mechanical pretreatments in industrial degradation.

Once lignin is removed or partially degraded, hemicellulose is the next target for degradation. Several families of glycosyl hydrolases (GH) and carboxyl esterases (CE) are involved in hemicellulose degradation. Because of the wide variety of sugar moieties, many different enzymes are required to break the specific varieties of bonds between each sugar. Enzymes can be specific to mannose, xylan, or the various other sugar types. The variety of GHs must also be diverse to accommodate specific enzymatic functions for the hydrolysis reactions needed for the compositional diversity of hemicellulose. Endo and exo activity dictate the location of catalysis on a branch or chain the enzyme acts upon. Xylan, being a major component of hemicellulose, is cleaved by endoxylanses (EC 3.2.1.8) and exoxylanases (EC 3.2.1.156). Endo-acting GHs bind along a chain and break 1,4-β-*D*-xylosidic linkages producing two new ends of the xylose chain. Exo-acting GHs can catalyze the hydrolysis reaction from the reducing end of a xylose chain. Exo activity can occur one sugar at a time or in a processive manner, where the enzyme continues along the chain releasing monomers or dimers of sugars. The resulting monomers or dimers of the various sugar components can then be fermented for biofuels.

Cellulases target the β-1,4 linkage between glucose moieties. Due to the crystalline nature of cellulose fibers, endocellulases (EC 3.2.1.4) are necessary to expose chain ends to allow for exocellulase activity. Functional diversity of exocellulases also exists, with specificity towards non-reducing (EC 3.2.1.91) or reducing ends (EC 3.2.1.176) of the cellulose fibril. Amorphous

regions of cellulose contain less hydrogen bonding providing irregularity and more access for endocellulase binding. Monomers and dimers of glucose can be cleaved from chains of cellulose by exocellulases and β-glucosidases (EC 3.2.1.21). The excess cellobiose released by these enzymes inhibit endocellulases, resulting in a negative feedback loop to prevent further degradation of cellulose. Several studies have analyzed the cellulolytic performance with and without supplemental β-glucosidases in attempts to lessen the inhibition of cellobiose upon cellulases (Brunecky et al., 2013; Brethauer and Studer, 2014).

The stability and activity of lignocellulolytic enzymes are paramount for efficient use in industrial processing. Many environments have been examined for novel CAZymes and cellulose degrading organisms, including rumen (Henderson et al., 2015), soil (López-Mondéjar et al., 2020), compost (Wang et al., 2016), and hot springs (Strazzulli et al., 2020). However, hot springs are a promising environment for finding effective enzymes that are naturally adapted to extreme conditions used in industrial pretreatment and processing. Hot springs have been a focal point for biotechnology discovery in the past. Continual exploration has occurred in all varieties of hot springs and geothermal features. The extreme conditions in hot springs provide a chance to discover enzymes with naturally selected stability and thermo-activity. Enzymes from thermal environments benefit industrial use from their natural adaptions to extreme conditions commonly used in processing. Methods of pretreatment are also a cause for the high operating costs of biomass conversion so discovering or developing enzymes that maintain activity either in conjunction with or without the need for pretreatment, is seen as a possible cost saving mechanism.

Hot spring cellulose degradation

The extreme conditions of hot springs create an environment that is distinct from other places on Earth. Novel microbes and metabolisms have frequently been discovered in hot springs (Pohlschroeder et al., 1994; Mladenovska et al., 1995; Jay et al., 2018; McKay et al., 2019). Degradation of complex carbons and other compounds have been studied with regards to activity from hot spring microbes. Hot spring microbial research in the 20th century was mainly limited to organisms that could be cultured and grown in the laboratory and these studies generated many isolated strains. Most probable number counts were often used to quantify cellulose and xylan degradation from springs which may have limited the true enumeration of cellulolytic organisms (Mathrani et al., 1993). Multiple taxa of thermophilic microbes were isolated and shown to degrade various carbohydrates to fermentation products (Saiki et al., 1985; Sonne-Hansen et al., 1993; Mladenovska et al., 1995). *Dictyoglomus thermophilum* was isolated from an alkaline (pH 8.7) hot spring with an optimal growth temperature of 78 °C and was able to grow on fermentable sugars such as cellobiose and xylose but not insoluble carbohydrates such as cellulose (Saiki et al., 1985). A large amount of hot spring cellulolytic microbes, mainly from the phyla *Dictyoglomi* and *Firmicutes,* were isolated in the 1990's. Members of the *Firmicutes* genus', *Caldicellulosiruptor*, which can degrade and ferment cellulose, were recognized as the highest temperature (90 °C), high activity cellulolytic microbe (Yang et al., 2010). A specific species of *Caldicellulosiruptor*, *Caldicellulosiruptor bescii*, can degrade cellulose, xylose, and even untreated plant biomass that is high in lignin content (Yang et al., 2009). These properties position *C. bescii* as an important model organism for designing improvements for industrial conversion of plant biomass. The multidomain enzyme complex, CelA, from *C. bescii* is unique from other cellulases as it is excreted from the cell and contains several CAZyme domains. Two

of the domains are hydrolases, GH9 and GH48, whereas the third domain is for binding to cellulose, CBM3 (Brunecky et al., 2013). CelA has been adopted by industrial processing and is included in several enzyme cocktail mixtures due to its high cellulolytic activity. As mentioned previously, certain organisms can degrade lignocellulose and ferment the monomeric sugars into products of interest. This process has been termed consolidated bioprocessing (CBP) as it occurs in a single organism. These specialty organisms can breakdown cellulose and hemicellulose as well as ferment the sugar products into ethanol. CBP is advantageous for industry as the growth conditions only need to be optimized to one organism as opposed to multiple organisms or for enzyme cocktails. However, rates of biomass conversion can still be optimized when co-cultures of CBP organisms are used (Singh et al., 2017). Fermentation products can include acetate, lactate, ethanol, hydrogen, and other compounds, often determined by culturing conditions. Additionally, highly cellulolytic organisms can be engineered to produce helpful enzymes such as alcohol dehydrogenases to increase production of specific products such as ethanol (Chung et al., 2014).

In efforts to improve the efficiency of degradation through enzyme cocktails and microbial communities, consortia of cooperating taxa have been used. Optimized blending of enzyme families has increased the overall activity of various cellulases. Synergistic effects among enzymes have been a way to use microbial communities to efficiently degrade plant biomass. As mentioned, negative feedback loop inhibition of cellulases by cellobiose has lowered potential enzyme activity. Introducing enzyme activity either through multiple community members (Zhao et al., 2014), plasmid engineering (Kim et al., 2018), or enzyme supplementation (Brethauer and Studer, 2014), to reduce the inhibition of cellobiose on

cellulases can increase the overall cellulolytic activity. Combination of a saccharolytic spirochete isolate with a cellulolytic *Clostridium* improved rates of cellulose degradation almost double of what the *Clostridium* could hydrolyze in monoculture (Pohlschroeder et al., 1994). While this study did not examine the specific enzymes utilized, spent medium from the *Clostridium* culture could support cell growth of the spirochete whereas no growth was observed in fresh medium with only the spirochete, suggesting the hydrolysis products of cellulose were used as growth substrate for the spirochete (Pohlschroeder et al., 1994). With more genetically tractable species currently in culture, a cellobiose phosphorylase (EC 2.1.4.20), which is a variety of glycosyl transferase, from *Caldicellulosiruptor bescii* led to improved cellulolytic activity when expressed in *Thermotoga maritima* extracellularly and improved cell growth when expressed intracellularly (Kim et al., 2018). The expressed enzyme was able to catalyze cellobiose into glucose and glucose 1-phosphate which do not inhibit cellulases, increasing the activity but also caused a suggested higher energetic requirement to transport the glucose into the cell through ATP dependent transporters, as opposed to the cellobiose.

Similar to the increased effectiveness of multidomain enzymes (e.g., CelA) or enzyme complexes (e.g., cellulosomes), having a variety of GH families from a consortium can be beneficial towards degradation, especially when working with untreated plant biomasses that vary in lignin, hemicellulose, and cellulose composition (Jensen et al., 2017). Consortium-based degradation of plant biomass has been a common focus as an alternative to isolates from hot springs as multiple taxa with unique cellulolytic activity can be present. Corn stover or switchgrass has been used as an inoculum in hot springs to retrieve enriched populations that have attached and grow in connection with the substrate. Peacock and colleagues (Peacock et al.,

2013) used sterile aspen shavings and pretreated corn stover in permeable bags incubated in two different temperate regions of a hot spring for two or three months to enrich for cellular biomass *in situ*. Retrieved samples were sequenced for taxonomic identification through sequencing the V7/V8 region of the 16S rRNA gene. The two substrates developed varying enrichment communities, presumably due to different biomass compositions containing cellulose and hemiceullose. Taxa identified on both substrates included *Thermotogae*, *Thermoprotei*, and *Archaeoglobi*, while *Dictyoglomi* and *Thermodesulfobacteria* sequences were more often observed in enrichments on aspen shavings. Even after several months of incubation, the community remained diverse with several main taxa, all with relatives implicated in lignocellulose degradation (Peacock et al., 2013). Outcomes from consortia enrichments have determined that specific taxa can be responsible for subsequent stages of lignocellulose degradation (Zhao et al., 2014). In addition to substrate affecting the resultant degradation community, length of incubation can also select for different populations. Over a 7-day incubation of a consortium on filter paper, denaturing gradient gel electrophoresis analysis showed a distinct community shift from the first 3 days compared to the final 4 days (Zhao et al., 2014). *Firmicutes* were the dominant phylum in this study but representatives across four families in this lineage varied throughout the incubations indicating different activities towards cellulose degradation were present throughout the enriched consortium's incubation.

The prospects of hot spring microbes in the advancement of industrial processing of lignocellulosic biomass is of great importance. Thermophilic enzyme stability and synergism will increase the efficiency of processing and help reduce the costs required to bring biofuels into a larger portion of the energy consumption market. Further research is needed to take full

advantage of this opportunity but developing molecular techniques and computational resources are providing necessary means to do so.

Overview of following chapters

To further describe the physiology and *in situ* activity of environmental microbes, we set out to develop new approaches to link function to taxonomy. In chapter 2, we design incubation schemes of hot spring sediment slurry to screen 23 substrate amendments in a high-throughput fashion, including variations in substrate concentration and changes to the incubation headspace composition. We hypothesized that utilizing BONCAT coupled with FACS (BONCAT-FACS), we can detect microbial taxa that preferentially increase their metabolic activity in response to favorable amendment conditions. Two incubation conditions, cellobiose addition and anoxic headspace, significantly changed the community composition of sorted cells compared to our baseline activity control. The microbial populations were then analyzed to determine which microbes responded to those condition. In chapter 3, we highlight the importance of hot springs as a trove of novel and diverse carbohydrate active enzymes. Chapter 3 focuses on the multitude of sequencing information that is publicly available and the computational tools used to reveal the CAZyme potential of hot springs, which as previously discussed are ripe for biotechnological and industrial process improvements. We addressed the wide range of taxonomy through worldwide metagenomic datasets and some of the CAZyme families contained within. While many metagenomes contained the potential for cellulose degradation, four metagenomes were analyzed in detail to describe the taxonomic lineages of glycosyl hydrolase families 3, 5, and 10 which correspond to oligosaccharide degrading enzymes, cellulases, and hemicellulases, repectively. In chapter 4, we brought together the main ideas of chapters 2 and 3 to apply

BONCAT-FACS to a Yellowstone National Park hot spring to identify populations involved in cellulose degradation using functional data and metagenomics. We hypothesized that in a hot spring rich with plant biomass, we would be able to stimulate cellulolytic microbes with additional cellulose substrate and fluorescently separate the active microbes for 16S rRNA gene sequencing for identification. While microbial populations from amplicon analysis did not directly match populations recovered from metagenomics, both datasets contain evidence of potential cellulose degradation. Chapter 5 is a resource announcement that includes a brief description of metagenomically sequenced data from incubations of hot spring sediment slurry used for BONCAT experiments from the work in chapter 2. Metagenome assembled genomes (MAGs) of high and medium quality were described with regards to their sequencing quality and taxonomy. The research summarized in this thesis is aimed to develop and implement new approaches to be able to describe the physiology of uncultured microorganisms within hot spring systems.

REFERENCES CITED

Agbor, V. B., Cicek, N., Sparling, R., Berlin, A., and Levin, D. B. (2011). Biomass pretreatment : Fundamentals toward application. 29, 675–685. doi:10.1016/j.biotechadv.2011.05.005.

Bang-Andreasen, T., Anwar, M. Z., Lanzén, A., Kjøller, R., Rønn, R., Ekelund, F., et al. (2020). Total RNA sequencing reveals multilevel microbial community changes and functional responses to wood ash application in agricultural and forest soil. *FEMS Microbiol. Ecol.* 96, 1–13. doi:10.1093/femsec/fiaa016.

Beam, J. P., Jay, Z. J., Schmid, M. C., Rusch, D. B., Romine, M. F., M Jennings, R. De, et al. (2016). Ecophysiology of an uncultivated lineage of Aigarchaeota from an oxic, hot spring filamentous "streamer" community. *ISME J.* 10, 210–224. doi:10.1038/ismej.2015.83.

Berry, D., Mader, E., Lee, T. K., Woebken, D., Wang, Y., Zhu, D., et al. (2015). Tracking heavy water (D_2O) incorporation for identifying and sorting active microbial cells. *Proc. Natl. Acad. Sci.* 112, E194–E203. doi:10.1073/pnas.1420406112.

Brethauer, S., and Studer, M. H. (2014). Consolidated bioprocessing of lignocellulose by a microbial consortium. *Energy Environ. Sci.* 7, 1446–1453. doi:10.1039/c3ee41753k.

Brunecky, R., Alahuhta, M., Xu, Q., Donohoe, B. S., Crowley, M. F., Kataeva, I. A., et al. (2013). Revealing Nature's Cellulase Diversity: The Digestion Mechanism of Caldicellulosiruptor bescii CelA. *Science* 342, 1513–1516.

Bugg, T. D. H., Ahmad, M., Hardiman, E. M., and Rahmanpour, R. (2011). Pathways for degradation of lignin in bacteria and fungi. *Nat. Prod. Rep.* 28, 1883–1896. doi:10.1039/c1np00042j.

Chen, L., Zhao, B., Li, X., Cheng, Z., Wu, R., and Xia, Y. (2021). Isolating and characterizing translationally active fraction of anammox microbiota using bioorthogonal non-canonical amino acid tagging. *Chem. Eng. J.* 418, 129411. doi:10.1016/j.cej.2021.129411.

Chung, D., Cha, M., Guss, A. M., and Westpheling, J. (2014). Direct conversion of plant biomass to ethanol by engineered Caldicellulosiruptor bescii. *Proc. Natl. Acad. Sci.* 111, 8931–8936. doi:10.1073/pnas.1402210111.

Couradeau, E., Sasse, J., Goudeau, D., Nath, N., Hazen, T. C., Bowen, B. P., et al. (2019). Probing the active fraction of soil microbiomes using BONCAT-FACS. *Nat. Commun.* 10. doi:10.1038/s41467-019-10542-0.

Dekas, A. E., Poretsky, R. S., and Orphan, V. J. (2009). Deep-Sea Archaea Fix and Share Nitrogen in Methane-Consuming Microbial Consortia. *Science (80-.).* 326. doi:doi:10.1126/science.1178223.

Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A., and Schuman, E. M. (2006). Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* 103, 9482–9487. doi:10.1073/pnas.0601637103.

Doud, D. F. R., Bowers, R. M., Schulz, F., Raad, M. De, Deng, K., Tarver, A., et al. (2019). Function-driven single-cell genomics uncovers cellulose-degrading bacteria from the rare biosphere. *ISME J.* doi:10.1038/s41396-019-0557-y.

Ellepola, K., Truong, T., Liu, Y., Lin, Q., Lim, T. K., Lee, Y. M., et al. (2019). Multi-omics Analyses Reveal Synergistic Carbohydrate Metabolism in Streptococcus mutans-Candida

albicans Mixed-Species Biofilms. *Infect. Immun.* 87, 1–20. doi:10.1128/IAI.00339-19.
Eloe-Fadrosh, E. A., Ivanova, N. N., Woyke, T., and Kyrpides, N. C. (2016a). Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nat. Microbiol.* 1, 15032. doi:10.1038/nmicrobiol.2015.32.
Eloe-Fadrosh, E. A., Paez-Espino, D., Jarett, J., Dunfield, P. F., Hedlund, B. P., Dekas, A. E., et al. (2016b). Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat. Commun.* 7, 1–10. doi:10.1038/ncomms10476.
García-Timermans, C., Props, R., Zacchetti, B., Sakarika, M., Delvigne, F., and Boon, N. (2020). Raman spectroscopy-based measurements of single-cell phenotypic diversity in microbial communities. *mSphere* 5, 1–14. doi:10.1101/2020.05.21.109934.
Hatzenpichler, R., Connon, S. A., Goudeau, D., Malmstrom, R. R., Woyke, T., and Orphan, V. J. (2016). Visualizing in situ translational activity for identifying and sorting slow-growing archaeal−bacterial consortia. *Proc. Natl. Acad. Sci.* 113, E4069–E4078. doi:10.1073/pnas.1603757113.
Hatzenpichler, R., Krukenberg, V., Spietz, R. L., and Jay, Z. J. (2020). Next-generation physiology approaches to study microbiome function at single cell level. *Nat. Rev. Microbiol.* 18. doi:10.1038/s41579-020-0323-1.
Hatzenpichler, R., Scheller, S., Tavormina, P. L., Babin, B. M., Tirrell, D. A., and Orphan, V. J. (2014). In situ visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. *Environ. Microbiol.* 16, 2568–2590. doi:10.1111/1462-2920.12436.
He, Y.-C., Ma, C.-L., and Yang, B. (2018). "Pretreatment Process and Its Synergistic Effects on Enzymatic Digestion of Lignocellulosic Material," in *Fungal Cellulolytic Enzymes: Microbial Production and Application*, 1–25. doi:10.1007/978-981-13-0749-2.
Hedlund, B. P., Dodsworth, J. A., Murugapiran, S. K., Rinke, C., and Woyke, T. (2014). Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter." *Extremophiles* 18, 865–875. doi:10.1007/s00792-014-0664-7.
Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P. H., et al. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* 5. doi:10.1038/srep14567.
Heyer, R., Kohrs, F., Benndorf, D., Rapp, E., Kausmann, R., Heiermann, M., et al. (2013). Metaproteome analysis of the microbial communities in agricultural biogas plants. *N. Biotechnol.* 30, 614–622. doi:10.1016/j.nbt.2013.01.002.
Huang, W. E., Griffiths, R. I., Thompson, I. P., Bailey, M. J., and Whiteley, A. S. (2004). Raman microscopic analysis of single microbial cells. *Anal. Chem.* 76, 4452–4458. doi:10.1021/ac049753k.
Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nat. Microbiol.*, 1–6. doi:10.1038/nmicrobiol.2016.48.
Jay, Z. J., Beam, J. P., Dlakić, M., Rusch, D. B., Kozubal, M. A., and Inskeep, W. P. (2018). Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. *Nat. Microbiol.* 3, 732–740. doi:10.1038/s41564-018-0163-1.
Jensen, C. U., Rodriguez Guerrero, J. K., Karatzos, S., Olofsson, G., and Iversen, S. B. (2017). Fundamentals of Hydrofaction^{TM}: Renewable crude oil from woody biomass. *Biomass*

Convers. Biorefinery 7, 495–509. doi:10.1007/s13399-017-0248-8.

Karst, S. M., Dueholm, M. S., McIlroy, S. J., Kirkegaard, R. H., Nielsen, P. H., and Albertsen, M. (2018). Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nat. Biotechnol.* 36, 190–195. doi:10.1038/nbt.4045.

Kiick, K. L., Saxon, E., Tirrell, D. A., and Bertozzi, C. R. (2002). Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U. S. A.* 99, 19–24. doi:10.1073/pnas.012583299.

Kim, S. K., Himmel, M. E., Bomble, Y. J., and Westpheling, J. (2018). Expression of a cellobiose phosphorylase from Thermotoga maritima in Caldicellulosiruptor bescii improves the phosphorolytic pathway and results in a dramatic increase in cellulolytic activity. *Appl. Environ. Microbiol.* 84, 1–9. doi:10.1128/AEM.02348-17.

Leizeaga, A., Estrany, M., Forn, I., and Sebastián, M. (2017). Using click-chemistry for visualizing in situ changes of translational activity in planktonic marine bacteria. *Front. Microbiol.* 8, 1–11. doi:10.3389/fmicb.2017.02360.

Liu, Y., Xu, J., Tao, Y., Fang, T., Du, W., and Ye, A. (2020). Rapid and accurate identification of marine microbes with single-cell Raman spectroscopy. *Analyst* 145, 3297–3305. doi:10.1039/c9an02069a.

Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J., and Crosby, L. (2018). Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. *mSystems* 3, 1–12. doi:10.1128/msystems.00055-18.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, 490–495. doi:10.1093/nar/gkt1178.

López-Mondéjar, R., Tláskal, V., Větrovský, T., Štursová, M., Toscan, R., Nunes da Rocha, U., et al. (2020). Metagenomics and stable isotope probing reveal the complementary contribution of fungal and bacterial communities in the recycling of dead biomass in forest soil. *Soil Biol. Biochem.* 148. doi:10.1016/j.soilbio.2020.107875.

Mathrani, I. M., Nielsen, P., Sonne-Hansen, J., Kristjansson, J. K., and Ahring, B. K. (1993). Influence of pH and temperature on enumeration of cellulose- and hemicellulose-degrading thermophilic anaerobes in neutral and alkaline Icelandic hot springs. *Appl. Environ. Microbiol.* 59, 1963–1965. doi:10.1128/aem.59.6.1963-1965.1993.

McKay, L. J., Dlakić, M., Fields, M. W., Delmont, T. O., Eren, A. M., Jay, Z. J., et al. (2019). Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. *Nat. Microbiol.* 4, 614–622. doi:10.1038/s41564-019-0362-4.

Michels, D. E., Lomenick, B., Chou, T.-F., Sweredoski, M. J., and Pasulka, A. (2021). Amino acid analog induces stress response in marine Synechococcus . *Appl. Environ. Microbiol.* doi:10.1128/aem.00200-21.

Mladenovska, Z., Mathrani, I. M., and Ahring, B. K. (1995). Isolation and characterization of Caldicellulosiruptor lactoaceticus sp. nov., an extremely thermophilic, cellulolytic, anaerobic bacterium. *Arch. Microbiol.* 163, 223–230. doi:10.1007/BF00305357.

Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. doi:10.1111/1462-2920.13023.

Peacock, J. P., Cole, J. K., Murugapiran, S. K., Dodsworth, J. A., Fisher, J. C., Moser, D. P., et al. (2013). Pyrosequencing Reveals High-Temperature Cellulolytic Microbial Consortia in Great Boiling Spring after In Situ Lignocellulose Enrichment. *PLoS One* 8, 1–12. doi:10.1371/journal.pone.0059927.

Pohlschroeder, M., Leschine, S. B., and Canale-Parola, E. (1994). Spirochaeta caldaria sp. nov., a thermophilic bacterium that enhances cellulose degradation by Clostridium thermocellum. *Arch. Microbiol.* 161, 17–24. doi:10.1007/BF00248889.

Rai, R., Agrawal, D., and Chadha, B. S. (2019). *New paradigm in degradation of lignocellulosic biomass and discovery of novel microbial strains*. doi:10.1007/978-981-13-8487-5_16.

Reichart, N. J., Bowers, R. M., Woyke, T., and Hatzenpichler, R. (2021). High Potential for Biomass-Degrading Enzymes Revealed by Hot Spring Metagenomics. *Front. Microbiol.* 12, 1–13. doi:10.3389/fmicb.2021.668238.

Reichart, N. J., Jay, Z. J., Krukenberg, V., Parker, A. E., Spietz, R. L., and Hatzenpichler, R. (2020). Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. *ISME J.* 14, 2851–2861. doi:10.1038/s41396-020-00749-1.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437. doi:10.1038/nature12352.

Riva, A., Kolimár, D., Spittler, A., Wisgrill, L., Herbold, C. W., Abrankó, L., et al. (2020). Conversion of Rutin, a Prevalent Dietary Flavonol, by the Human Gut Microbiota. *Front. Microbiol.* 11, 1–11. doi:10.3389/fmicb.2020.585428.

Sadler, N. C., and Wright, A. T. (2015). Activity-based protein profiling of microbes. *Curr. Opin. Chem. Biol.* 24, 139–144. doi:10.1016/j.cbpa.2014.10.022.

Saiki, T., Kobayashi, Y., Kawagoe, K., and Beppu, T. (1985). Dictyoglomus thermophilum gen. nov., sp. nov., a Chemoorganotrophic, Anaerobic, Thermophilic Bacterium. *Int. J. Syst. Bacteriol.* 35, 253–259.

Saxena, R., Dhakan, D. B., Mittal, P., Waiker, P., Chowdhury, A., Ghatak, A., et al. (2017). Metagenomic analysis of hot springs in central india reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Front. Microbiol.* 7. doi:10.3389/fmicb.2016.02123.

Singer, E., Wagner, M., and Woyke, T. (2017). Capturing the genetic makeup of the active microbiome in situ. *ISME J.* 11, 1949–1963. doi:10.1038/ismej.2017.59.

Singh, N., Mathur, A. S., Tuli, D. K., Gupta, R. P., Barrow, C. J., and Puri, M. (2017). Cellulosic ethanol production via consolidated bioprocessing by a novel thermophilic anaerobic bacterium isolated from a Himalayan hot spring. *Biotechnol. Biofuels* 10, 1–18. doi:10.1186/s13068-017-0756-6.

Sonne-Hansen, J., Mathrani, I. M., and Ahring, B. K. (1993). Xylanolytic anaerobic thermophiles from Icelandic hot-springs. *Appl. Microbiol. Biotechnol.* 38, 537–541. doi:10.1007/BF00242951.

Steward, K. F., Eilers, B., Tripet, B., Fuchs, A., Dorle, M., Rawle, R., et al. (2020). Metabolic Implications of Using BioOrthogonal Non-Canonical Amino Acid Tagging (BONCAT) for Tracking Protein Synthesis. *Front. Microbiol.* 11, 1–12. doi:10.3389/fmicb.2020.00197.

Strazzulli, A., Cobucci-Ponzano, B., Iacono, R., Giglio, R., Maurelli, L., Curci, N., et al. (2020). Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme

environments. *FEBS J.* 287, 1116–1137. doi:10.1111/febs.15080.

Valentini, T. D., Lucas, S. K., Binder, K. A., Cameron, L. C., Motl, J. A., Dunitz, J. M., et al. (2020). Bioorthogonal non-canonical amino acid tagging reveals translationally active subpopulations of the cystic fibrosis lung microbiota. *Nat. Commun.* 18, S36. doi:10.1016/s1569-1993(19)30233-4.

Wagner, M. (2009). Single-Cell Ecophysiology of Microbes as Revealed by Raman Microspectroscopy or Secondary Ion Mass Spectrometry Imaging. *Annu. Rev. Microbiol.* 63, 411–429. doi:10.1146/annurev.micro.091208.073233.

Wang, C., Dong, D., Wang, H., Müller, K., Qin, Y., Wang, H., et al. (2016). Metagenomic analysis of microbial consortia enriched from compost: New insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnol. Biofuels* 9, 1–17. doi:10.1186/s13068-016-0440-2.

Xue, M. Y., Sun, H. Z., Wu, X. H., Liu, J. X., and Guan, L. L. (2020). Multi-omics reveals that the rumen microbiome and its metabolome together with the host metabolome contribute to individualized dairy cow performance. *Microbiome* 8, 1–19. doi:10.1186/s40168-020-00819-8.

Yang, S. J., Kataeva, I., Hamilton-Brehm, S. D., Engle, N. L., Tschaplinski, T. J., Doeppke, C., et al. (2009). Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "Anaerocellum thermophilum" DSM 6725. *Appl. Environ. Microbiol.* 75, 4762–4769. doi:10.1128/AEM.00236-09.

Yang, S. J., Kataeva, I., Wiegel, J., Yin, Y., Dam, P., Xu, Y., et al. (2010). Classification of "Anaerocellum thermophilum" strain DSM 6725 as Caldicellulosiruptor bescii sp. nov. *Int. J. Syst. Evol. Microbiol.* 60, 2011–2015. doi:10.1099/ijs.0.017731-0.

Zegeye, E. K., Sadler, N. C., Lomas, G. X., Attah, I. K., Jansson, J. K., Hofmockel, K. S., et al. (2020). Activity-Based Protein Profiling of Chitin Catabolism. *ChemBioChem* 21, 1–8. doi:10.1002/cbic.202000616.

Zhao, C., Deng, Y., Wang, X., Li, Q., Huang, Y., and Liu, B. (2014). Identification and characterization of an anaerobic ethanol-producing cellulolytic bacterial consortium from great basin hot springs with agricultural residues and energy crops. *J. Microbiol. Biotechnol.* 24, 1280–1290. doi:10.4014/jmb.1401.01022.

CHAPTER TWO

ACTIVITY-BASED CELL SORTING REVEALS RESPONSES OF UNCULTURED ARCHAEA AND BACTERIA TO SUBSTRATE AMENDMENT

Contribution of Authors and Co-Authors

Author: Nicholas J. Reichart

Contributions: Designed the study, performed the incubations, optimized and tested the sorter experiments, processed the samples, analyzed the data, and wrote the manuscript.

Author: Zackary J. Jay

Contributions: Provided bioinformatic support with regards to Qiime2 processing and helped edit the manuscript.

Author: Viola Krukenberg

Contributions: Performed the incubations, optimized and tested the sorter experiments, and helped edit the manuscript.

Author: Albert E. Parker

Contributions: Provided statistical support to analyze the data and helped edit the manuscript.

Author: Rachel L. Spietz

Contributions: Helped optimize and test the sorter experiments and edit the manuscript.

Author: Roland Hatzenpichler

Contributions: Designed the study and wrote the manuscript.

Manuscript Information Page

Nicholas J. Reichart, Zackary J. Jay, Viola Krukenberg, Albert E. Parker, Rachel L. Spietz, Roland Hatzenpichler

The ISME Journal

Status of Manuscript:

____ Prepared for submission to a peer-reviewed journal

____ Officially submitted to a peer-reviewed journal

____ Accepted by a peer-reviewed journal

_X__ Published in a peer-reviewed journal

Publisher: Nature Publishing Group

Reprinted with permission from Reichart, N.J., Jay, Z.J., Krukenberg, V. *et al.* Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. *ISME J* **14,** 2851-2861 (2020). https://doi.org/10.1038/s41396-020-00749-1

CHAPTER TWO

ACTIVITY-BASED CELL SORTING REVEALS RESPONSES OF UNCULTURED ARCHAEA AND BACTERIA TO SUBSTRATE AMENDMENT

Abstract

Metagenomic studies have revolutionized our understanding of the metabolic potential of uncultured microorganisms in various ecosystems[1]. However, many of these genomic predictions have yet to be experimentally tested, and the functional expression of genomic potential often remains unaddressed. In order to obtain a more thorough understanding of cell physiology, novel techniques capable of testing microbial metabolism under close to *in situ* conditions must be developed[2]. Here, we provide a benchmark study to demonstrate that bioorthogonal non-canonical amino acid tagging (BONCAT) in combination with fluorescence-activated cell sorting (FACS) and 16S rRNA gene sequencing can be used to identify anabolically active members of a microbial community incubated in the presence of various growth substrates or under changing physicochemical conditions. We applied this approach to a hot spring sediment microbiome from Yellowstone National Park (Wyoming, USA) and identified several microbes that changed their activity levels in response to substrate addition, including uncultured members of the phyla *Thaumarchaeota*, *Acidobacteria*, and *Fervidibacteria*. Because shifts in activity in response to substrate amendment or headspace changes are indicative of microbial preferences for particular growth conditions, results from this and future BONCAT-FACS studies could inform the development of cultivation media to specifically enrich uncultured microbes. Most

importantly, BONCAT-FACS is capable of providing information on the physiology of uncultured organisms at as close to *in situ* conditions as experimentally possible.

Introduction

Traditionally, metagenomic predictions about the ecophysiology of microbes have been tested using stable isotope probing (SIP), which can determine cellular activity through incorporation of growth substrates. SIP experiments have been limited by their need for isotope-labeled molecules that can be prohibitively expensive or unavailable because of compositional complexity. An alternative to SIP targeted at a specific assimilatory pathway is to label all anabolically active cells in a sample with deuterated water (^2H_2O) and visualize heavy isotope uptake into individual cells using Raman microspectroscopy[3,4]. Laser tweezers can then be used to separate functionally active cells from the microbial community[3,4]. However, even using automated platforms, Raman-activated cell sorting currently can only achieve sorting rates of up to 500 cells per hour[5]. Such low-throughput limits functional tests to the most abundant, active members of a community and a handful of incubation conditions of high interest to the researcher.

Bioorthogonal non-canonical amino acid tagging combined with fluorescence-activated cell sorting (BONCAT-FACS) is a recently developed high-throughput approach[2] that can be used to identify uncultured microorganisms that are translationally active in a high-throughput approach, analyzing thousands of cells per second[6–10]. BONCAT tracks translational activity through the incorporation of a synthetic amino acid, such as *L*-homopropargylglycine (HPG). HPG is a structural analog of *L*-methionine and is incorporated into newly synthesized proteins due to the substrate promiscuity of methionyl-tRNA synthetase[11]. HPG-containing proteins can

be detected via copper-catalyzed azide-alkyne click chemistry dye-staining and subsequently visualized with epifluorescence microscopy[6,7,12]. In combination with FACS, BONCAT fluorescence can be used to sort translationally active members of a microbial community and taxonomically identify them with subsequent 16S ribosomal RNA (rRNA) gene sequencing[7,8]. BONCAT-FACS was recently used to identify the anabolically active fraction of microbes in soil[8] and deep-sea sediments[7] under close to *in situ* conditions. A previous study demonstrated the capability of BONCAT to reveal the activity response of cultured bacteria to substrate amendment[6], suggesting that BONCAT could also be used to identify beneficial growth conditions for yet uncultured microorganisms. Here, we used a sediment slurry sample from a hot spring in Yellowstone National Park (Wyoming, USA) to demonstrate that BONCAT-FACS can be used to test genomic predictions of microbial physiology and screen community-wide responses in anabolic activity to numerous treatments in parallel, including amendments with potential growth substrates as well as varied headspace conditions.

Materials and methods

Incubation setup

All work in Yellowstone National Park was conducted under US National Park Service permit no. YELL-SCI-8010 (2017-2019). A sediment slurry sample (approximately 10 % sediment, 90 % water) was collected using a stainless-steel measuring cup fixed to the end of a 12-foot telescoping pole from a hot spring within the Five Sisters hot spring group in Yellowstone National Park, WY (44.532619, -110.797272) (Figure 1).

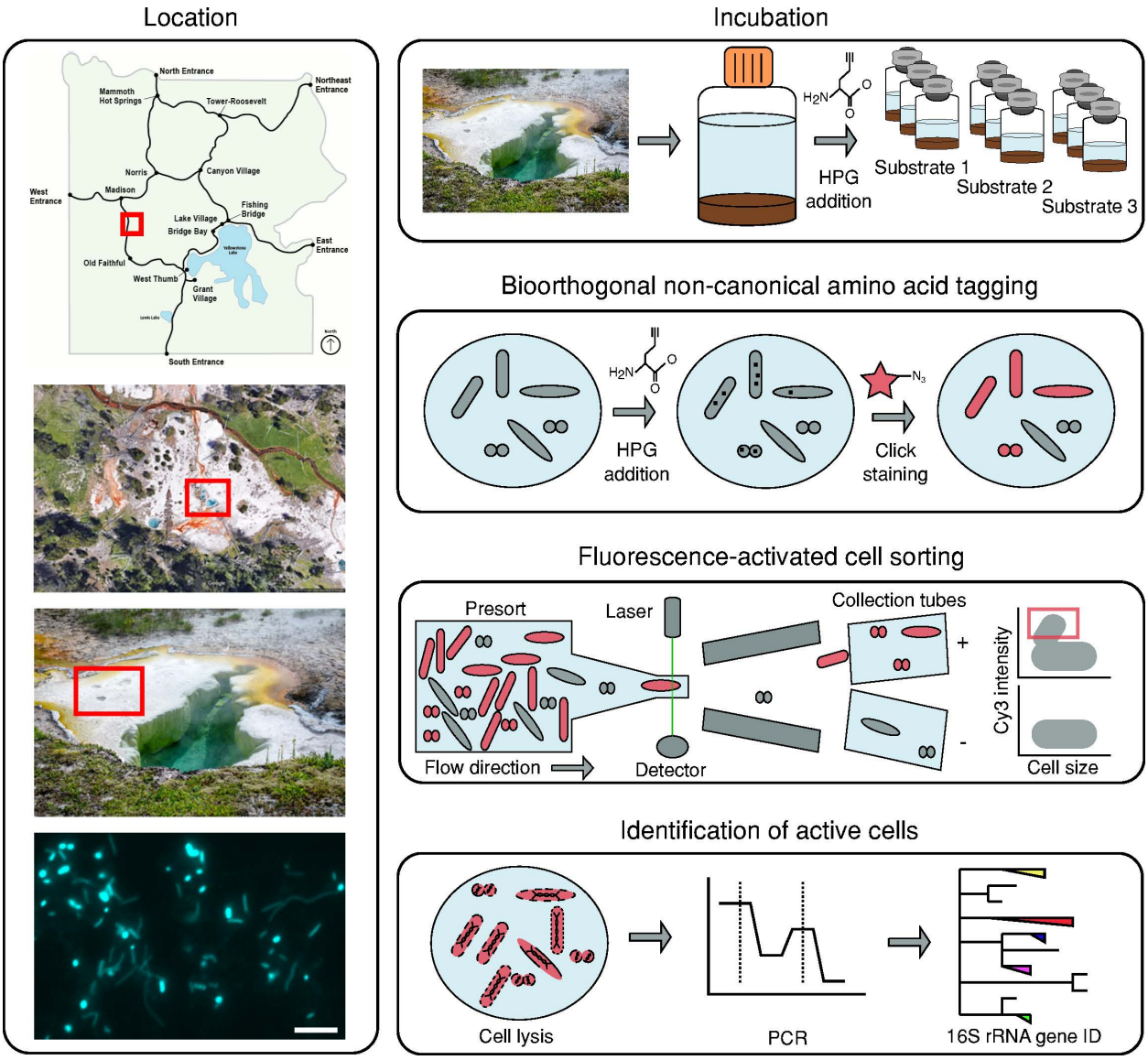


Figure 1. Site location and workflow for BONCAT-FACS identification of active cells from environmental samples. Location of hot spring FS5 in Yellowstone National Park. Microscopy image of unsorted hot spring cells visualized with general nucleic acid stain, 4',6-diamidino-2-phenylindole (DAPI). Scale bar represents 5 µm. Samples were incubated in the presence of *L*-homopropargylglycine (HPG) under various treatments (Supplementary Table 1) for 48 hours. Cells were detached from particle material and newly synthesized proteins were dye-stained via azide-alkyne click chemistry. Translationally active, fluorescent cells were sorted using fluorescence-activated cell sorting (FACS) from the presort, total extractable cell community. Cells were lysed and 16S rRNA genes were amplified with PCR. Black squares (■) in cells represent new proteins with incorporated HPG. Map of YNP provided by the US National Park Service. Satellite image obtained from Google Maps.

At the time of sampling the temperature, pH, and dissolved oxygen of the spring were 72.2 °C, 8.45, and 0.909 mg/L (Hach DR900 dissolved oxygen colorimetric assay), respectively. The slurry sample was transferred to a 1 L acid-washed and autoclaved glass bottle, sealed with a screw cap, and transported to the laboratory in an insulated heated container within 4 hours of sampling, maintaining an ambient temperature of 55 °C. Incubation experiments were prepared the day following sample collection after allowing the sample to adjust to 74 °C in an incubator overnight. Each incubation consisted of 6 mL homogenized slurry aliquoted into 30 mL acid-washed and autoclaved glass vials. The slurry was kept homogenous by stirring with a magnetic stir bar during setup. A T_0 sample was taken from the homogenous mixture and frozen in 10 % glycerol-TE[13] buffer (glycerol, Tris-EDTA buffer, milliQ water) for later DNA extraction. A set of three vials to which no substrate or bioorthogonal amino acid was added served as a negative control (*i.e.* no-HPG control). After the sample was transferred to the no-HPG control, *L*-homopropargylglycine (HPG; Click Chemistry Tools) was added to the remaining bulk slurry to yield a final concentration of 50 μM. Afterwards, 6 mL of the HPG-containing slurry was transferred to all incubation vials that had already been amended with a substrate (Supplementary Table 1). Substrates were confirmed to be stable at 74 °C for 48 hours by Liquid Chromatography Mass Spec (LC-MS) analysis (not shown). A set of three vials, which contained no substrate amendment but HPG, served as a control (*i.e.* HPG-only control) for close to *in situ* activity of the microbial community. All vials were sealed with sterile butyl rubber stoppers with ambient lab air as a headspace (21 % O_2). Two sets of triplicate incubations tested headspace composition on microbial activity, 2 % O_2 (microoxic) and 100 % N_2 (anoxic) were prepared by flushing the headspace with 0.2 μm filtered N_2 gas for three minutes. The vials for 2 % O_2

incubations had approximately 10 % of the N_2 headspace volume removed and replaced with 0.2 µm filtered ambient lab air. All incubations were performed in triplicate and were incubated at 74 °C in the dark for 48 hours without shaking. This incubation time was chosen after preliminary experiments had been performed with 6, 24, and 48 hours incubation times. An incubation time of 48 hours was selected to increase cell labeling rates to allow for the reliable sorting of BONCAT-positive cells. At the completion of the incubation, vials were uncapped, and the content of each vial was immediately transferred to a 15 mL conical tube with 10 % glycerol-TE[13] buffer (glycerol, Tris-EDTA buffer, milliQ water) to preserve cells. Afterwards, these tubes were stored at -80 °C until further processing.

Sample processing

For cell extractions, samples preserved in glycerol-TE were thawed at 4 °C, vortexed briefly at maximum speed, and 1 mL of slurry was transferred to a 15 mL conical tube containing 4 mL of 1x phosphate buffered saline (PBS) with Tween 20 (Promega) at a final concentration of 0.01%. The conical tubes were vortexed for 5 minutes at maximum speed to detach cells from particles before being centrifuged for 5 minutes at 500 g to separate particles from the biomass. Following centrifugation, the cell-containing supernatant was passed through a 35 µm pore size filter and the filtrate was divided evenly across 1.5 mL tubes and centrifuged at 14,000 g for 5 minutes to pellet the cells. The supernatant was discarded by careful pipetting, and the pellet of each sample was resuspended and combined in a final volume of 300 µL 1x PBS. A bulk click reaction solution was prepared[12] and 200 µL of this mix was aliquoted to each sample. Succinctly, the final reaction mix was comprised of 5 mM amino guanidine hydrochloride (Sigma Aldrich), 5 mM sodium *L*-ascorbate (Sigma Aldrich), 100 µM copper

sulfate pentahydrate (Sigma Aldrich), 500 µM THPTA (Click Chemistry Tools), and 4 µM Cy3 picolyl-azide dye (Click Chemistry Tools) in 1x PBS. The reaction mixtures were vortexed briefly to mix and then rotated in the dark at room temperature for 1 hour. After this incubation step, all samples were washed three times by a series of centrifugation steps at 14,000 *g* for 5 minutes and resuspended in 1 mL of 1xPBS. Following the final wash, cells were resuspended in 500 µL of 1x PBS and stored at 4 °C in the dark before being sorted on a fluorescence-activated cell sorter (FACS) the subsequent day. One replicate of each incubation condition was extracted per day and all samples were checked for successful completion of the click reaction by epifluorescence microscopy (Leica DM4B microscope with DAPI and Cy3 filter sets).

Fluorescence-activated cell sorting

On the day following cell extraction and click staining, cells were sorted based on Cy3 fluorescence intensity using a Sony SH800S FACS. Prior to sorting, a 50 µL aliquot of each sample ("Presort" sample) was transferred to a new tube and stored at 4 °C for later processing. For sorting, a first gate was drawn based on Forward Scatter Area (FSC-A) and Back Scatter Area (BSC-A) to exclude any large particles that remained after filtration (Supplementary Figure 2). A second gate was then drawn based on Forward Scatter Width (FSC-W) and Forward Scatter Height (FSC-H) to further constrain the size of particles examined. The third and final gate was determined using FSC-A and Cy3 fluorescence intensity to capture fluorescent, translationally active ("BONCAT gate") cells only. The gate for Cy3 fluorescence was drawn based on the no-HPG control which contained only background fluorescence of the sample and no positive BONCAT signal. The fluorescence of the no-HPG control was recorded to define the lower limit of the "BONCAT gate" to capture <0.1 % of all events in the no-HPG control

(Supplementary Figure 2). The HPG-only control was also recorded to define the upper limits of the gate that captured fluorescence cells. Cells within the "BONCAT gate" were sorted into 1.5 mL tubes containing 200 µL of 1x PBS to prevent cells from sticking to the tube walls and lysing. The cell sorter contained sheath fluid composed of UV-sterilized and 0.2 µm filtered 1x PBS. Each sample was sorted for a maximum of 250,000 fluorescent events or until the volume was exhausted prior to meeting this total (see Supplementary Table 2 for a list of events sorted per replicate). Sorted cells were stored at 4 °C. At the end of each day of FACS sorting, all samples, including "Presort" samples, from that day were centrifuged at 14,000 g for 5 minutes and the supernatants were discarded by careful aspiration before the cell pellets were resuspended in 20 µL of nuclease-free water. Epifluorescence microscopy was used to check the purity of sorted cells, confirming the presence of only fluorescent cells in the "BONCAT gate" (not shown). Cells were stored at 4 °C until all samples were sorted. SYBR Green I nucleic acid stain (Invitrogen) was tested as a counterstain to fluorescently identify all events containing biomass prior to gating on a BONCAT fluorescent signal. However, after testing several different combinations of SYBR Green I with clickable dyes (*i.e.* Cy3, 594, and Cy5) it was determined that too much bleed-over across fluorescence channels would prohibit the reliable sorting of SYBR and BONCAT positive events with the current laser and detector setup.

DNA extraction, PCR, and sequencing

For DNA extraction, three cycles of freeze-thaw were used to lyse cells. Cells stored in 20 µL nuclease-free water at 4 °C were transferred into 96 well microtiter plates that were sealed with sterile adhesive foil sheets. Freezing was performed by placing the plate into a -80 °C freezer for 20 minutes. Thawing was done at 99 °C for 10 minutes in a thermal cycler

(Eppendorf Mastercycler nexus GSX1). Prior to each subsequent freezing step, the plates were pulse centrifuged at 500 g. The T_0 sample taken at the beginning of the incubation was extracted using the FAST DNA spin kit for soil (MP Biomedicals). Amplification of bacterial and archaeal 16S rRNA genes was performed by PCR using the Earth Microbiome protocol[14] with updated 515F[15] (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R[16] (5'-GGACTACNVGGGTWTCTAAT-3') primers directly into the microtiter plates containing the extracted DNA. The final PCR volume was 50 µL and consisted of 20 µL Invitrogen Platinum Taq II 2X Master Mix, 1 µL 515F primer (10 µM; final: 0.2 µM), 1 µL 806R primer (10 µM; final: 0.2 µM), and 8 µL nuclease-free water added directly into the microtiter plates containing the 20 µL of lysate. The thermocycler conditions were: 94 °C for 3 minutes followed by 28 cycles of 94 °C for 45 sec, 50 °C for 60 sec, and 72 °C for 90 sec before a final elongation step at 72 °C for 10 minutes. PCR products were purified with Ampure XP beads (Beckman Coulter) to remove excess primers following the manufacturer's protocol with a final elution step in 40 µL nuclease-free water. A second PCR reaction was performed to attach dual barcode indices and sequencing adapters to the previously amplified and purified products. This PCR was performed in a 25 µL final volume with 5 µL purified, amplified DNA, 12 µL Invitrogen Platinum Taq II 2X Master Mix, 2.5 µL i5 primer (final: 0.25 µM), 2.5 µL i7 primer (final: 0.25 µM), and 2.5 µL water. The PCR conditions were as follows: 95 °C for 3 minutes followed by 8 cycles of 95 °C for 30 sec, 55 °C for 30 sec, and 72 °C for 30 sec, followed by a final elongation step at 72 °C for 5 minutes. Following completion of the barcoding PCR, samples were randomly selected and checked for presence of a PCR band on a 1.5 % agarose gel. From each PCR plate, a no template control and a positive control that used E. *coli* DNA were included. After confirming correct

band length, each PCR reaction was again purified with Ampure XP beads to remove excess primers. Triplicate Quant-iT Picogreen dsDNA Assay (Invitrogen) reactions were performed for all samples following manufacturer's guidelines and quantified with a Biotek Synergy H1 Hybrid microplate reader. For each sample, excluding *E. coli* positive controls, 10 ng of barcoded and cleaned DNA per sample was pooled. Additional samples that were amplified and included in the sample pool were reagent extraction blanks and sheath fluid aliquots from the sorter sheath fluid tank. Half of the pooled volume was concentrated with the QIAquick PCR purification spin column kit (Qiagen) following manufacturer's guidelines. Sequencing was performed by Laragen Inc. (Culver City, CA) using Illumina 2x250 paired end read MiSeq sequencing. A total of 12,071,436 raw reads were generated from 209 samples.

Testing of FACS sorting and DNA extraction efficiency

The minimum number of sorted events required for sufficient biomass recovery of an organism was determined by sorting events from an *Aliivibrio fischeri* pure culture, and sorting purity was confirmed by 16S rRNA gene sequencing. Events ranging from $5x10^2$ to $5x10^4$ were sorted and sequenced. It was determined that *A. fischeri* accounted for 98.91% of the reads when $5x10^4$ events were sorted. The sequencing reads assigned to *A. fischeri* decreased with decreasing number of events sorted, consistent with literature reports on FACS[17]. Therefore, $5x10^4$ events sorted was used as a minimum threshold of events for sample processing.

Validation of successful DNA extraction through consecutive freeze-thaw cycles was confirmed with four pure culture isolates and one environmental sample. The following cultures were purchased from the German Collection of Microorganisms (DSMZ) and grown in the recommended media: *Caldicellulosiruptor bescii* (DSMZ 6725), *Desulfobacterium*

autotrophicum (DSMZ 3382), *Rhodopirellula baltica* (DSMZ 10527), and *Aliivibrio fischeri* (DSMZ 507). The environmental sample tested was sediment-water slurry from hot spring FS5. DNA was quantified by Qubit high sensitivity dsDNA assay (Invitrogen) on varying amounts of cell numbers (achieved by consecutive 1:10 dilutions) for each test sample. The lowest quantity of cells required for detection with this Qubit assay was determined to be 100,000 cells for *R. baltica* while all other cultures needed a minimum of 250,000 cells. DNA was extracted from FS5 with the FAST DNA spin kit for soil (MP Biomedicals) and 16S rRNA genes were amplified and sequenced as described above. The community composition of the environmental sample based on 16S rRNA gene sequence analysis was comparable to the sequence composition of the "presort" samples that were lysed with repeated freeze thaw cycles. Sequence composition from the DNA extraction kit samples were similar to the composition of samples extracted with freeze thaw cycles.

BONCAT substrate specificity tests with *Escherichia coli*

E. coli was grown overnight in an Erlenmeyer flask in M9 minimal media with 0.04 % glucose. The culture was then split into equal volumes and received new M9 media as before but containing different substrate amendments. Five conditions were tested in triplicate with the following addition to the base medium (0.04 % glucose) and a final concentration of 50 µM HPG; 0.36 % sucrose, 0.36 % sorbitol, 0.36 % glucose, 0.36 % glucose and 10 µg/mL chloramphenicol, or no substrate addition. A sixth condition was tested without additional substrate or HPG. After 2 hours of incubation, *E. coli* cells were fixed with 2 % paraformaldehyde for 1 hour at room temperature, washed three times with 1x PBS, and click stained with Cy3 picolyl-azide dye. Click stained cells were visualized with an epifluorescence

microscope and imaged with Leica software. Images were analyzed using daime software package[18] for relative fluorescence units (RFU). RFU were averaged for each condition (Supplementary Figure 1).

Amplicon data processing

Sequence reads were processed using QIIME 2 (version 2018.6)[19]. Briefly, barcode adapter sequences were first removed using cutadapt prior to truncation of the forward (130 bases) and reverse (150 bases) reads based on review of the sequencing quality. Reads were then denoised, merged, and chimera-checked with DADA2, resulting in 4,627 amplicon sequence variants (ASVs). Taxonomy was assigned using Silva SSU database release 128 with the classify-sklearn method. R[20] package *decontam* (version 1.1.2)[21] was run to remove contaminants using the "Prevalence" model with a threshold of 0.5. *decontam* identified 438 contaminating ASVs which represented 399,454 reads (3.31 % of total reads). Samples represented by at least 5,000 sequence reads were used for further analysis. ASVs were collapsed at the genus level, relative sequence abundance was calculated for each sample, and only taxa above 0.01 % across all samples were analyzed. From here on, relative abundance is referring to the relative abundance of sequencing reads. Outlying replicates were determined using the *DESeq2* package[22] by clustering rlog transformed counts by Euclidean distance and clustering raw counts by Poisson distance. One presort sample was removed as an outlier (Glycine replicate 3) and four sorted samples were removed (Biotin replicate 3, Leucine replicate 3, Glycine replicate 1, and Isoleucine replicate 2). Removed samples were confirmed by manual inspection of sequence data and all samples contained a majority of reads attributed to known contaminating organisms that were not filtered with *decontam*. Microbial community analysis was performed in R[20] using

PERMANOVA with substrate and replicate as factors in the vegan package[23] with Bray-Curtis dissimilatory. Shannon's diversity index was calculated for all treatments and separately compared to the HPG-only controls using linear mixed effects models in the *lme4* package[24] with substrate treatment as a fixed effect and replicate as a random effect with p-values adjusted for multiple comparisons[25]. Log2 fold change of the taxa expression in each treatment was assessed using a negative binomial model with treatment and replicate as factors in *DESeq2*. All sequence data associated with this study have been deposited in *Genbank* under BioProject number PRJNA601738. All files generated during data processing are available in the supplementary online information.

Results and discussion

To expand on previous BONCAT findings, we initially performed an experiment with *Escherichia coli*, which demonstrated that translational activity as measured by BONCAT can only be detected when cultures were grown with glucose or sorbitol as the sole carbon and energy source but not with sucrose (Supplementary Figure 1). This is consistent with the notion that *E. coli* is genetically incapable of sucrose utilization. Addition of chloramphenicol, an antibiotic targeting ribosome function, to the culture decreased the BONCAT fluorescence intensity to background levels, *i.e.* fluorescence values observed when cells were grown with HPG but without growth substrate. These results demonstrated that BONCAT can be used to study cell activity responses to substrate amendment and suggested that it could be used to study complex microbiomes.

For this benchmark study, we selected a high temperature (74 °C), alkaline (pH 8.2) hot spring (Five Sisters 5, FS5) in the Lower Geyser Basin of Yellowstone National Park, Wyoming,

USA (Figure 1). Hot springs harbor lower complexity microbial communities compared to other environments, such as soils and marine sediments, making them an ideal system for the development of novel approaches. As indicated by 16S rRNA gene amplicon sequencing, hot spring FS5 was dominated by the archaeal candidate phylum *Aigarchaeota* (41.9 %), with the next most abundant community members belonging to bacterial candidate phylum *Fervidibacteria* (9.57 %) and phylum *Deinococcus-Thermus* (7.98 %) (). Members of the *Aigarchaeota* have been detected in many high temperature (65-88 °C) hot springs over a wide pH range (2.9-9.3)[26]. Despite multiple studies describing their versatile metabolic potential, *Aigarchaeota* remain elusive to cultivation and no experimental data on their functional activity is currently available. Reconstructed genomes suggest the capacity for aerobic respiration or anaerobic respiration with nitrate as an electron acceptor[26]. In addition, a recent study found multiple endoglucanases and β-glucosidases that might be involved with degradation of cellulose and cellobiose in an *Aigarchaeota* metagenome bin[27]. One of the goals of our study was to test these functional predictions using BONCAT-FACS.

We applied BONCAT-FACS combined with 16S rRNA gene sequencing to identify the active microbial community members in hot spring material when incubated in the presence of 23 different substrates or growth conditions (Figure 1; see list of treatments in Supplementary Table 1). Because BONCAT relies on the incorporation of a synthetic amino acid as an indirect activity tracer, we were also able to test activity responses towards changes of headspace gas composition. Specifically, we tested the activity of microbes in the presence of atmospheric (21 % O_2), microoxic (2 % O_2) and anoxic (100 % N_2) headspace conditions. Laboratory incubations were established with a slurry containing FS5 water and sediment to which HPG and substrates

were added. After 48 hours of incubation, biomass was resuspended in glycerol Tris-EDTA buffer[13] and samples were frozen at -80 °C for later processing. After samples were thawed, cells were extracted from the slurry, fluorescently stained using click chemistry, and sorted via FACS based on fluorescence signal. The sorted cells were lysed with repeated freeze-thaw cycles prior to PCR amplification of their 16S rRNA genes. An HPG-only incubation was used to identify cells active under close to *in situ* conditions (with low perturbation) and represented the baseline for comparing cellular activity under varying incubation conditions.

The microbial community of FS5 was highly active as represented by the high proportion of BONCAT labeled events in the HPG-only control (Supplementary Table 1). Furthermore, all abundant taxa (>1 % relative abundance of the total extractable community, ASVs collapsed to the genus level) were active in at least one condition, and most taxa were active under several conditions (Figure 2). This demonstrates that BONCAT can be applied to a wide variety of phylogenetic groups, which is consistent with previous reports[6–10,28]. The sorted, active fraction from incubations with different treatments contained no statistically significant difference when compared to the HPG-only control based on Bray-Curtis dissimilarity (MANOVA, $p = 1$) (Figure 2). This result indicates that the microbial activity response to any treatment, as captured by BONCAT, was not large enough or consistent enough among replicates to significantly change the overall active community when compared to the HPG-only control.

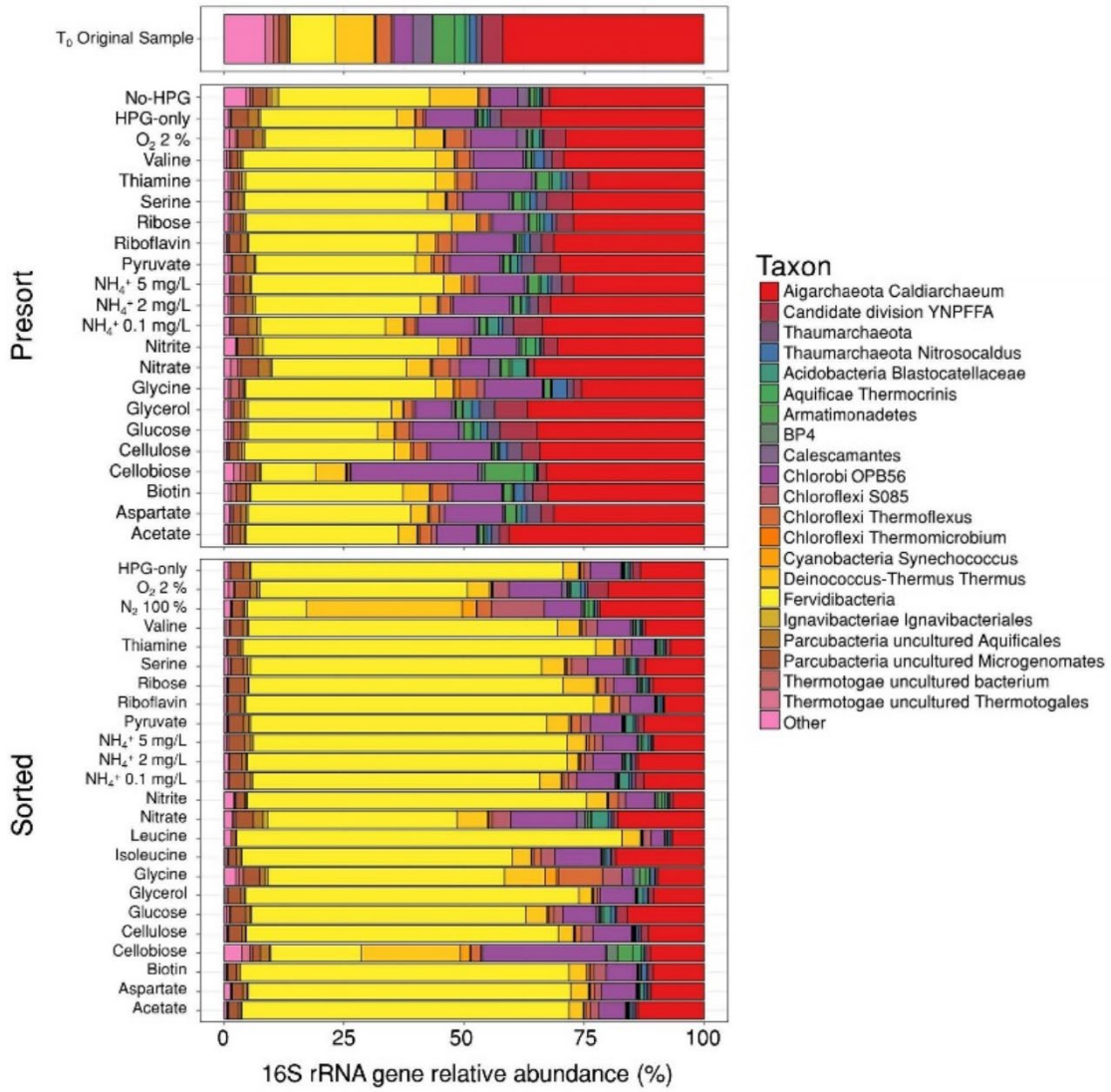


Figure 2. 16S rRNA gene relative abundances in averaged incubations. Top panel, unamended FS5 community at the time of sampling (T_0). Middle panel, community composition of presort samples, representing the extractable microbial community, after the incubation experiments. Bottom panel, composition of the sorted active cell fraction after incubation with substrate amendment. Only taxa that were represented above 1 % in at least one sample are shown. Additional taxa were combined into "Other" category. Taxon level represents highest taxonomic resolution for each taxa (Silva 128). Some replicates did not pass the quality control steps of our bioinformatics pipeline (<5,000 reads), which precluded including them in this figure.

Although the overall composition of the active communities did not vary significantly in response to incubation conditions, changes within the richness and evenness of the communities were detected. To further describe the community composition, the Shannon's diversity index of each incubation was calculated (Figure 3). For each treatment, the bulk, presort fraction (representing the total, extractable cell community), and the sorted, active cell fraction were analyzed and compared to the respective HPG-only control. Overall, the variation in the presort fraction of Shannon's diversity indices was less variable than the sorted fraction from the same treatment incubation (Figure 3). This demonstrated that the BONCAT-FACS approach could detect changes in the diversity of the active community due to individual cell responses prior to shifts in cell abundance occurring in presort communities. We expected little variation in the presort community because the limited time of incubation should not have allowed for an overall shift in community composition on a bulk level. However, *Fervidibacteria* were observed in higher proportion in most presort populations as compared to the original T_0 bulk sample. This could be attributed to either favorable growth of this yet uncultured lineage or be a result of preferential cell extraction or cell lysis during freeze-thaw cycles as compared to bulk sample DNA extraction. Alternatively, their increase in relative abundance could have resulted from sample cooling (72º to 55 ºC) during transit from the field to the laboratory (four hours).

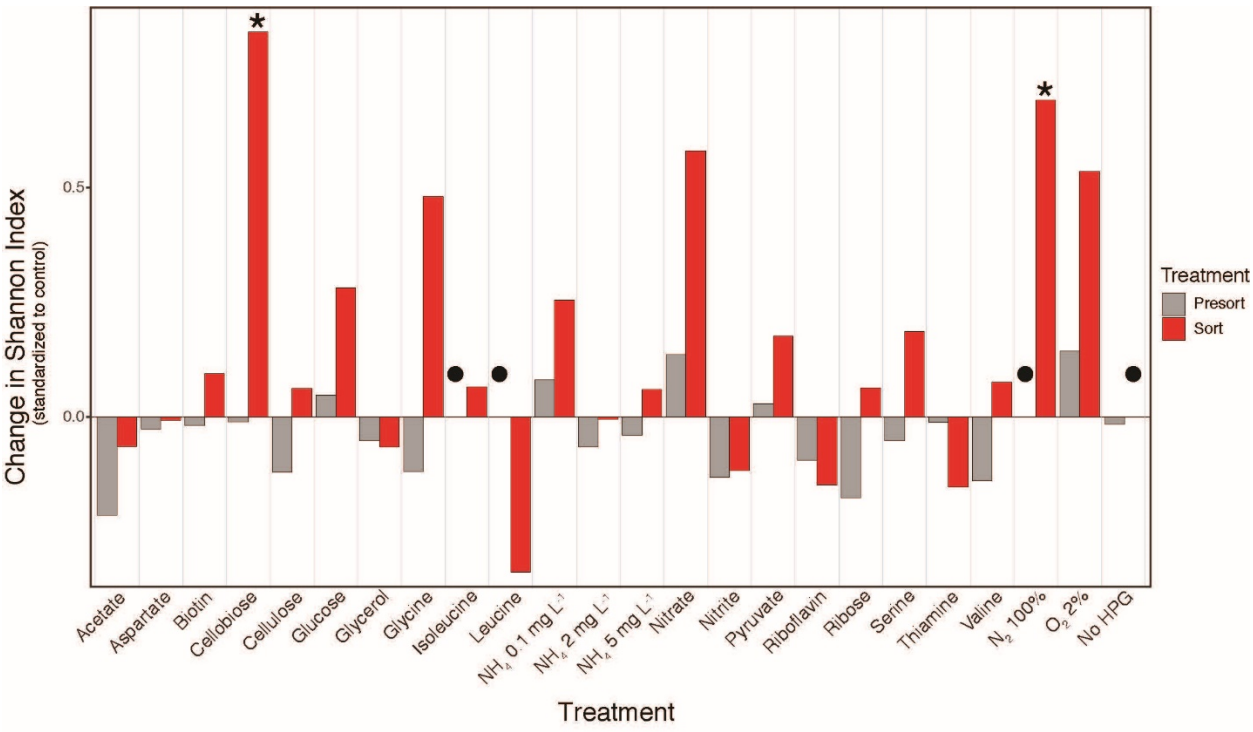


Figure 3. Change in Shannon's diversity indices of each treatment standardized to HPG-only controls. The Shannon's diversity indices for presort and sorted samples were compared relative to their respective HPG-only controls. The value for HPG-only was set to 0 and the difference for each sample is plotted. Overall, presort samples exhibited less variability than sorted samples. Sorted samples for cellobiose and anoxic conditions (100% N_2 headspace) were significantly different from the sorted HPG-only (atmospheric air) control ($p < 0.05$) and marked with an asterisk (*). Samples marked with a circle (●) indicate no data available.

The Shannon's diversity indices of the active, sorted fractions were more varied when standardized to the HPG-only control than the presort fractions. This indicates that the activity response among different taxa varies between treatments. For example, incubations with cellobiose or under anoxic conditions led to statistically significantly different diversity indices on average than the HPG-only control incubations (linear mixed effects model, $p = 0.00349$ and $p = 0.0334$, respectively). Cellobiose is a disaccharide of repeating glucose monomers that compose cellulose polymers. Cellobiose can be cleaved by β-1,4 hydrolases into glucose monomers which serve as initial substrate for most central metabolic pathways. The Shannon's diversity index of cells active with cellobiose-amendment was the highest we observed alluding to wide use of cellobiose increasing the evenness of active microbes in these incubations ($p = 0.00349$).

The perimeter of FS5 has a photosynthetic zone (<70 °C; Figure 1), away from the sampled area, which, together with constant exchange with the atmosphere, serves as source of O_2 in the spring (dissolved oxygen levels were 0.909 mg/L at the time of sampling but can vary 1-3 mg/L). Headspace exchange from aerobic (atmospheric air, 21 % O_2) to anoxic conditions (100 % N_2) in the incubations caused the active community to have a significant increase in Shannon's diversity ($p = 0.0334$) compared to the HPG-only control. This is consistent with the idea that facultative anaerobes persist in the hot spring and their activity varies with oxygen availability.

The relative abundance of six taxa increased and two decreased statistically significantly in the active, sorted community in response to cellobiose amendment based on $\log_2$ fold change (LFC) when compared to the HPG-only sorted fraction ($p < 0.10$). This suggests that cellobiose

has an impact on the activity of several taxa despite the hot spring lacking noticeable large carbon inputs or organic matter that would be degraded regularly *in situ*. Uncultured phylum *BP4* had an LFC of 3.62 in response to cellobiose amendment (Figure 4; for a list of all ASVs see Supplementary Table 3). While not much is known about *BP4*, these results could suggest a substrate preference for *BP4* towards cellobiose or its degradation products. In contrast, candidate phylum *Fervidibacteria* decreased in relative abundance in the active fraction of cellobiose amendment with an LFC of -2.88 (Figure 4). This decrease in *Fervidibacteria* abundance in the active fraction was surprising because this lineage had previously been proposed to have the genomic potential for complete lignocellulose degradation[29]; further, we did not detect a significant response to the addition of cellulose ($p = 1$). These results demonstrate the importance to functionally test genomic predictions under environmentally relevant conditions.

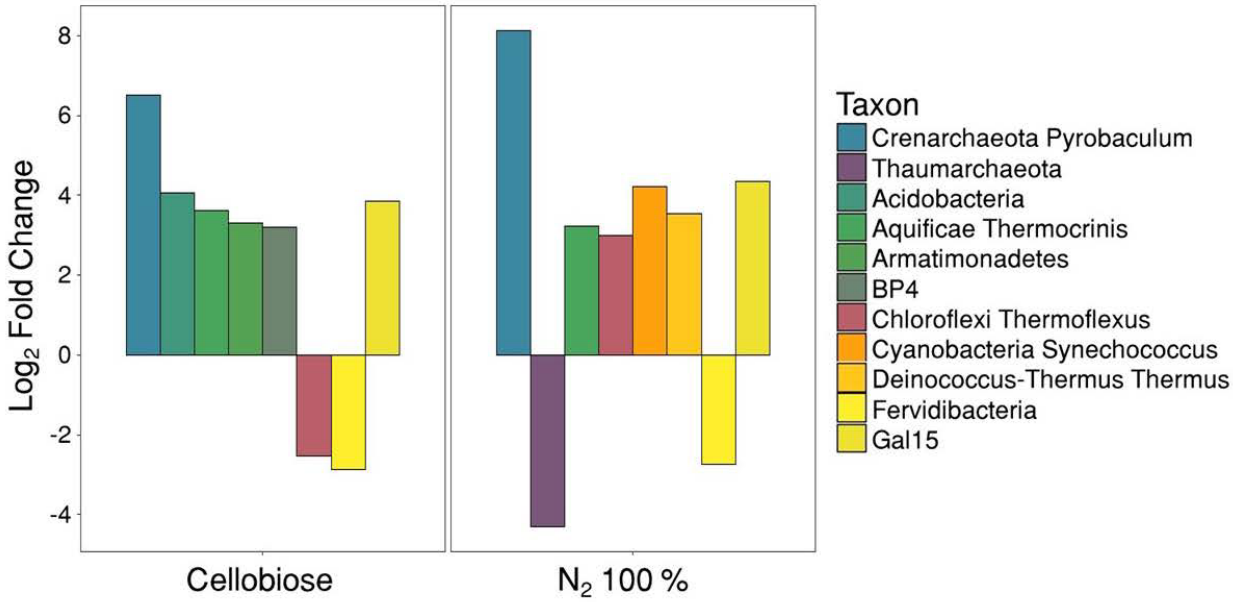


Figure 4. Log_2 fold change (LFC) of active taxa in cellobiose substrate amendment and N_2 100 % headspace condition. Each treatment was compared to the HPG-only control for sorted samples and LFC was calculated. The value for HPG-only was set to 0 and the LFC for each taxon is plotted. Only taxa significantly different from the control are shown ($p < 0.1$). Taxon level represents highest taxonomic resolution for each taxa (Silva 128).

In anoxic incubations, *Thermus* increased (LFC = 3.53, Figure 4) and represented a large portion of the sorted, translationally active community (Figure 2). In contrast, a rare member of the *Thaumarchaeota* (averaged relative abundance in the HPG-only sample was 0.4 %), whose taxonomic affiliation could not be resolved beyond phylum level, decreased (LFC = -4.31, Figure 4) in activity in response to anoxic conditions. This is consistent with the observation that all cultured thaumarchaeotes, including thermophilic representatives, have a strictly aerobic lifestyle[30–33].

Sequences related to the genus *Pyrobaculum* (phylum *Crenarchaeota*) exhibited the largest LFC increase in abundance of any taxon detected in the presence of cellobiose (LFC = 6.52) or under anoxic conditions (LFC = 8.12) (Figure 4). *Pyrobaculum* isolates demonstrate widely varying phenotypic capabilities, including both autotrophic and heterotrophic lifestyles, which can be coupled to H_2, O_2, AsO_4^{3-}, S^0, $S_2O_3^{2-}$, SeO_4^{2-}, Fe^{3+}, and NO_3^- respiration[34–38]. This metabolic versatility permits *Pyrobaculum* species to occupy geochemically diverse, mildly-acidic to basic (pH > ~4) geothermal environments, including hot springs throughout YNP[39] and likely explains their significant LFC increases in the two treatments.

Cellobiose and anoxic incubations were the only treatments to inform us of community-wide substrate and headspace selectivity with statistical significance. However, organisms belonging to *Thermocrinis* and uncultured phylum *Gal15* increased in relative abundance in both the cellobiose ($p = 0.0219$ and 0.0488, respectively) and anoxic ($p = 0.0229$ and 0.0209, respectively) incubations' sorted fractions alluding to possible favorable cultivation conditions for these organisms. Other organisms were also detected with large LFCs, though they were not statistically significant when averaged across all three biological replicates. For some samples

high variability was observed, emphasizing the need for replication in order to draw biologically meaningful conclusions.

An uncultured member of the *Aigarchaeota* represented a smaller proportion of presort reads (31.4 %) after the two-day incubation than in the original sample (41.9 %) (Figure 2). While *Aigarchaeota* sequences were dominant in the presort fractions, they were not highly abundant in any active fraction (averaged 12.4 %). In contrast, the BONCAT-labeled, active fraction of cells were dominated by the uncultured candidate phylum *Fervidibacteria*[40]. *Fervidibacteria* also comprised 48.0 % of the total reads across all presorted, bulk communities and the active, sorted cell fraction (Figure 2). These results could indicate a potential for slow growth or a lack of necessary growth factors for *Aigarchaeota* in our incubations while *Fervidibacteria* maintained high activity regardless of the amendment. Previously published genomic predictions[27] informed us of possible favorable conditions for *Aigarchaeota* , including acetate and biotin, but none of the tested substrates stimulated a strong activity response. This suggests the importance of unknown community interactions or growth conditions of *Aigarchaeota* that so far have not been replicated in laboratory experiments and highlights the limits of genomic functional predictions and interpretations. *Aigarchaeota* have been proposed to have the genomic potential for both aerobic and anaerobic growth[26] and our experimental results support this hypothesis. Members of *Aigarchaeota* were most abundant in the active fractions of microoxic (20.0 %) and anoxic (21.7 %) treatments despite not being statistically significant ($p = 1$) (HPG-only = 13.3 %). The ability to analyze microbial activity under varying headspace compositions demonstrates that BONCAT-FACS is an approach with great potential for studying cellular responses involving non-assimilatory pathways.

Preliminary experiments determined that 48 hours was to be the optimal time for incubation of this sample type to allow for enough HPG incorporation to be detected efficiently on our cell sorter (data not shown). However, this incubation time showed high baseline activity in the HPG-only control making activity variation due to substrate amendment difficult to determine. Extended incubation times with HPG have been shown to not be detrimental to the survival or activity of microorganisms in other sample types[6–8,10,12]. However, we posit that proteins related to environmental changes or defensive or stress mechanisms could be responsible for increased protein synthesis. Longer incubation times could select for specific populations and act as pseudo-enrichment allowing "weed" organisms to obscure activity response from hard-to-culture or rare taxa. Despite our limited incubation time, we consistently observed a higher abundance of *Fervidibacteria* populations in presort samples, possibly attributing to higher proportions in sorted, active samples. Alternatively, sampling too early might not provide enough opportunity for organisms to respond to the conditions provided, attributing the BONCAT signal to activity independent of sample treatment, or a general lack of signal. This is a problem with any methodology surveying activity in response to substrate amendment including BONCAT and heavy water SIP[4].

Conclusion

In this study, the microbial community activity response to substrate amendments and varying oxygen conditions in the headspace was compared to HPG-only incubations that represented close to *in situ* conditions. This allowed us to determine substrate or oxygen preference of microorganisms without the need for prior cultivation or genomic analysis. The overall composition of the active microbial community did not vary significantly from the HPG-

only control (according to Bray Curtis dissimilarity), but both the Shannon's diversity indices and the relative abundances of specific microbial populations (LFC) in the active communities were affected by some treatments.

Acknowledgements

We thank Rebecca Mueller, Brent Peyton, and Dana Skorupa for their assistance with traveling to and sampling the FS5 hot spring, Katherine Steward for LC-MS analyses, and Heidi Smith for helpful discussions on BONCAT-FACS during the early stages of this project. This work was supported through a NSF RII Track-2 FEC award (DBI-1736255), a NASA Exobiology program award (80NSSC19K1633), a NASA Early Career Fellowship to R.H. (80NSSC19K0449), and a grant by the W. M. Keck Foundation. The acquisition of a FACS by the Hatzenpichler lab was made possible through a grant by the Gordon and Betty Moore Foundation (GBMF5999). N.J.R. was partially supported by a Molecular Biosciences Program Fellowship from Montana State University, the Thermal Biology Institute, Montana State University's Vice President's Office of Research, Economic Development and Graduate Education, and a Department of Energy Office of Science Graduate Student Research award. The authors appreciate the US National Park Service, in particular Annie Carlson at the Yellowstone Center for Resources, for permitting work in YNP under permit number YELL-SCI-8010, 2017-2019.

Author contributions

N.J.R. and R.H. designed the study. N.J.R. and V.K. performed the incubations. N.J.R., V.K. and R.L.S. optimized and tested the sorter experiments. N.J.R. processed the samples. Z.J.J.

provided bioinformatics support. N.J.R. and A.E.P. analyzed the data. N.J.R. and R.H. wrote the manuscript. All authors reviewed and edited the manuscript.

Conflict of interest statement

The authors declare no conflict of interest.

REFERENCES CITED

1. Castelle, C. J. & Banfield, J. F. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. *Cell* **172**, 1181–1197 (2018).
2. Hatzenpichler, R., Krukenberg, V., Spietz, R. L. & Jay, Z. J. Next-generation physiology approaches to study microbiome function at the single cell level. *Nat. Rev. Microbiol.* **18**, 241-256 (2020) 10.1038/s41579-020-0323-1.
3. Huang, W. E., Ward, A. D. & Whiteley, A. S. Raman tweezers sorting of single microbial cells. *Environ. Microbiol. Rep.* **1**, 44–49 (2009).
4. Berry, D. *et al.* Tracking heavy water (D_2O) incorporation for identifying and sorting active microbial cells. *Proc. Natl. Acad. Sci.* **112**, E194–E203 (2015).
5. Lee, K. S. *et al.* An automated Raman-based platform for the sorting of live cells by functional properties. *Nat. Microbiol.* **4**, 1035–1048 (2019).
6. Hatzenpichler, R. *et al.* In situ visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. *Environ. Microbiol.* **16**, 2568–2590 (2014).
7. Hatzenpichler, R. *et al.* Visualizing in situ translational activity for identifying and sorting slow-growing archaeal−bacterial consortia. *Proc. Natl. Acad. Sci.* **113**, E4069–E4078 (2016).
8. Couradeau, E. *et al.* Probing the active fraction of soil microbiomes using BONCAT-FACS. *Nat. Commun.* **10**, (2019).
9. Leizeaga, A., Estrany, M., Forn, I. & Sebastián, M. Using click-chemistry for visualizing in situ changes of translational activity in planktonic marine bacteria. *Front. Microbiol.* **8**, 1–11 (2017).
10. Samo, T. J., Smriga, S., Malfatti, F., Sherwood, B. P. & Azam, F. Broad distribution and high proportion of protein synthesis active marine bacteria revealed by click chemistry at the single cell level. *Front. Mar. Sci.* **1**, (2014).
11. Kiick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 19–24 (2002).
12. Hatzenpichler, R. & Orphan, V.J. Detection of protein-synthesizing microorganisms in the environment via bioorthogonal noncanonical amino acid tagging (BONCAT). *Hydrocarb. Lipid Microbiol. Protoc. - Springer Protoc. Handbooks* 1–29 (2015). doi:10.1007/8623
13. Rinke, C. *et al.* Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* **9**, 1038–1048 (2014).
14. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
15. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).
16. Apprill, A., Mcnally, S., Parsons, R. & Weber, L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* **75**, 129–137 (2015).
17. Minich, J. J. *et al.* KatharoSeq Enables High-Throughput Microbiome Analysis from

Low-Biomass Samples. *mSystems* **3**, 1–16 (2018).
18. Daims, H., Lücker, S. & Wagner, M. daime , a novel image analysis program for microbial ecology and biofilm research. *Environ. Microbiol.* **8**, 200–213 (2006).
19. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
20. R Core Team. A langauge and environment for statistical computing. *R Foundation for Statistical Computing* Vienna, Austria. http://R-project.org/ (2019).
21. Davis, N. M., Proctor, Di. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 1–14 (2018).
22. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 1–21 (2014). doi:10.1186/s13059-014-0550-8
23. Oksanen, J. *et al*. vegan: community ecology package. *R package version 2.5-6*. https://CRAN.R-project.org/package=vergan (2019).
24. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67**, (2015).
25. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous Inference in General Parametric Models. *Biometrical J.* **50**, 346–363 (2008).
26. Hedlund, B. P. *et al.* Uncultivated thermophiles: Current status and spotlight on '*Aigarchaeota'*. *Curr. Opin. Microbiol.* **25**, 136–145 (2015).
27. Hua, Z. S. *et al.* Genomic inference of the metabolism and evolution of the archaeal phylum *Aigarchaeota*. *Nat. Commun.* **9**, (2018).
28. Pasulka, A. L. *et al.* Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. *Environ. Microbiol.* **20**, 671–692 (2018).
29. Peacock, J. P. *et al.* Pyrosequencing Reveals High-Temperature Cellulolytic Microbial Consortia in Great Boiling Spring after In Situ Lignocellulose Enrichment. *PLoS One* **8**, 1–12 (2013).
30. Hatzenpichler, R. *et al.* A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc. Natl. Acad. Sci.* **105**, (2008).
31. De Torre, J. R., Walker, C. B., Ingalls, A. E., Könneke, M. & Stahl, D. A. Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ. Microbiol.* **10**, 810–818 (2008).
32. Abby, S. S. *et al.* Candidatus *Nitrosocaldus cavascurensis*, an Ammonia Oxidizing , Extremely Thermophilic Archaeon with a Highly Mobile Genome. *Front. Microbiol.* **9**, 1–19 (2018).
33. Daebeler, A. *et al.* Cultivation and Genomic Analysis of " Candidatus *Nitrosocaldus islandicus*," an Obligately Thaumarchaeon from a Hot Spring. *Front. Microbiol.* **9**, 1–16 (2018).
34. Cozen, A. E. *et al.* Transcriptional Map of Respiratory Versatility in the Hyperthermophilic Crenarchaeon *Pyrobaculum aerophilum*. *J. Bacteriol.* **191**, 782–794 (2009).
35. Jay, Z. J. *et al. Pyrobaculum yellowstonensis* strain WP30 respires on elemental sulfur and/or arsenate in circumneutral sulfidic geothermal sediments of Yellowstone National

Park. *Appl. Environ. Microbiol.* **81**, 5907–5916 (2015).
36. Lee, J. Y. *et al. Pyrobaculum igneiluti* sp . nov ., a novel anaerobic hyperthermophilic archaeon that reduces thiosulfate and ferric iron. *Int. J. Sys. Evol. Microbiol.* **67**, 1714–1719 (2017). doi:10.1099/ijsem.0.001850
37. Nunoura, T., Sako, Y., Wakagi, T. & Uchida, A. Regulation of the aerobic respiratory chain in the facultatively aerobic and hyperthermophilic archaeon *Pyrobaculum oguniense*. *Microbiology* **149**, 673–688 (2003). doi:10.1099/mic.0.26000-0
38. Slobodkina, G. B., Lebedinsky, A. V, Chernyh, N. A. & Slobodkin, A. I. *Pyrobaculum ferrireducens* sp. nov., a hyperthermophilic Fe (III) - , selenate- and arsenate- reducing crenarchaeon isolated from a hot spring. *Int. J. Sys. Evol. Microbiol.* **65**, 851–856 (2015). doi:10.1099/ijs.0.000027
39. Jay, Z. J. *et al.* The distribution, diversity and function of predominant Thermoproteales in high-temperature environments of Yellowstone National Park. *Environ. Microbiol.* **18**, 4755–4769 (2016).
40. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).

CHAPTER THREE

HIGH POTENTIAL FOR BIOMASS-DEGRADING ENZYMES REVEALED BY HOT SPRING METAGENOMICS

Contribution of Authors and Co-Authors

Author: Nicholas J. Reichart

Contributions: Designed the study, performed the data analyses, and wrote the manuscript.

Author: Robert M. Bowers

Contributions: Provided bioinformatic support and helped edit the manuscript.

Author: Tanja Woyke

Contributions: Designed the study and edited the manuscript.

Author: Roland Hatzenpichler

Contributions: Designed the study and wrote the manuscript.

Manuscript Information Page

Nicholas J. Reichart, Robert M. Bowers, Tanja Woyke, Roland Hatzenpichler

Frontiers in Microbiology

Status of Manuscript:

____ Prepared for submission to a peer-reviewed journal

____ Officially submitted to a peer-reviewed journal

____ Accepted by a peer-reviewed journal

_X__ Published in a peer-reviewed journal

Publisher: Frontiers Media

Reprinted with permission from Reichart NJ, Bowers RM, Woyke T, and Hatzenpichler R (2021) High Potential for Biomass-Degrading Enzymes Revealed by Hot Spring Metagenomics. *Front. Microbiol*. 12:668238. doi: 10.3389/fmicb.2021.668238.

CHAPTER THREE

HIGH POTENTIAL FOR BIOMASS-DEGRADING ENZYMES REVEALED BY HOT SPRING METAGENOMICS

Abstract

Enzyme stability and activity at elevated temperatures are important aspects in biotechnological industries such as the conversion of plant biomass into biofuels. In order to reduce the costs and increase the efficiency of biomass conversion, better enzymatic processing must be developed. Hot springs represent a treasure trove of underexplored microbiological and protein chemistry diversity. Herein, we conduct an exploratory study into the diversity of hot spring biomass-degrading potential. We describe the taxonomic diversity and carbohydrate active enzyme (CAZyme) coding potential in 71 publicly available metagenomic datasets from 58 globally distributed terrestrial geothermal features. Through taxonomic profiling, we detected a wide diversity of microbes unique to varying temperature and pH ranges. Biomass-degrading enzyme potential included all five classes of CAZymes and we described the presence or absence of genes encoding 19 glycosyl hydrolases hypothesized to be involved with cellulose, hemicellulose, and oligosaccharide degradation. Our results highlight hot springs as a promising system for the further discovery and development of thermo-stable biomass-degrading enzymes that can be applied toward generation of renewable biofuels. This study lays a foundation for future research to further investigate the functional diversity of hot spring biomass-degrading enzymes and their potential utility in biotechnological processing.

Introduction

Lignocellulose is one of the most abundant polymers on Earth and is used in many industrial settings including textile processing, paper milling, and biofuel conversion (Cabrera and Blamey, 2018). Non-food crops and agricultural waste present a target for the conversion of lignocellulose into ethanol as a second-generation biofuel, without the need for competition with food sources such as corn and wheat (Gupta and Verma, 2015). However, lignocellulose is extremely recalcitrant and difficult to break down in a cost-efficient manner (Cheah et al., 2020). In 2019, the United States of America consumed approximately 541 billion liters of gasoline while only producing 60 billion liters of bioethanol (Renewable Fuels Association website for production, https://ethanolrfa.org). Reducing the costs of biofuel production while increasing yield would allow renewable energy to become a greater portion of today's energy market.

Lignocellulose is composed of lignin, hemicellulose, and cellulose in varying compositions depending on the plant source (Gupta and Verma, 2015; He et al., 2018). The lignin component contains aromatic compounds that encompass branched polysaccharides of hemicellulose and crystalline chains of cellulose. Currently, industrial processing of lignocellulose occurs with costly mechanical or chemical pretreatment to depolymerize the lignin component and make the saccharides more accessible (Cheah et al., 2020). Pretreatment using various methods increases the amount of saccharification by an order of magnitude over untreated biomass (Bala and Singh, 2019). Removal of lignin through pretreatment exposes hemicellulose and cellulose and increases surface area available for saccharification by continued pretreatment or enzymatic degradation. The released sugar products can then be fermented to produce biofuels.

Prior to fermentation, a multitude of enzymes are required for the complete degradation of lignocellulose, including laccases, hemicellulases, and cellulases. The Carbohydrate Active Enzyme (CAZy) database (Lombard et al., 2014) provides sequence-based classification families of the many types of enzymes involved. Lignin degradation occurs from laccases and peroxidases most commonly in the Auxiliary Activity (AA) families 1 and 2, respectively (Levasseur et al., 2014). These enzymes catalyze oxidation reactions of phenolic compounds and have been widely studied in fungi and some aerobic bacteria (Bugg et al., 2011). Hemicellulases, including xylanases, mannases, and other debranching enzymes, catalyze the hydrolysis of β-1,4 glycosidic bonds of the corresponding sugars and these enzymes are classified in a multitude of Glycosyl Hydrolase (GH) families. Cellulases act exclusively upon the Glucose-β-1,4-Glucose bonds of long chain fibrils that are encompassed by hemicellulose and lignin. As with hemicellulases, cellulases are classified as GHs and are either endo- or exo-acting, depending on the structure of their active site (Davies and Henrissat, 1995).

Lignocellulolytic enzymes and their activities have been studied in detail (Zarafeta et al., 2016; Zayulina et al., 2020), but the discovery of new enzymes has mainly been limited to isolated microorganisms or enrichment cultures. Most cellulolytic isolates come from only three classes of bacteria, *Actinobacteria*, *Bacilli*, and *Clostridia* (Bayer et al., 2013). Much of the early work on bacterial cellulases was performed with members of the genus *Caldicellulosiruptor* (originally *Anaerocellum*) and its multi-domain, high activity CelA (Brunecky et al., 2013; Young et al., 2014). Traditionally, enrichment cultures have been established using crystalline cellulose or plant biomass (*e.g.*, corn stover) and have captured consortia of cellulolytic capabilities (Graham et al., 2011; Zhao et al., 2014). However, rare cellulolytic community

members or members of yet uncultured lineages are often missed in these traditional assays and require novel techniques to be captured (Doud et al., 2019). *In silico* discovery of CAZymes has also been used to identify novel cellulolytic genes from metagenomes and the subsequent expression of the enzymes in recombinant hosts allowed for characterization studies to be conducted (Wohlgemuth et al., 2018; Strazzulli et al., 2020) and the activities could be tested under pretreatment conditions such as ionic liquids (Heins et al., 2014).

Many studies have tested environmental samples and cellulolytic enrichments for microbial biomass degradation activity, but these studies have generally been limited to taxa that grow during the enrichment or are naturally abundant. Rumen (Hess et al., 2011; Henderson et al., 2015), termite guts (Marynowska et al., 2020), compost (Ma et al., 2020), forest soils (Wilhelm et al., 2018), and hot springs (Peacock et al., 2013) are sample types that have been the focal points of discovering yet uncultivated cellulolytic organisms. Hot springs are a particularly favorable system for further research into lignocellulolytic enzymes because of the enzymatic adaptations to the extreme conditions in these ecosystems. Industrial chemical pretreatment can range from pH 1 to 13 (Pedersen et al., 2011) with variable activity of the pretreatment processes towards the different components of lignocellulose depending on the pH of the system. Such extreme conditions are typically absent from most natural systems but are normal in geothermal systems where thermostable enzymes have been selected to be active across a broad range of temperature and pH. For example, sites in Yellowstone National Park (YNP) exhibit a wide range of pH (0.52-10.6) and temperature (10-99°C) (see http://rcn.montana.edu/Default.aspx for a database of ~7,700 features in the park). Hot springs in New Zealand exhibit similar extreme conditions (pH <1-9.7, 14-101°C; (Power et al., 2018)). Thermostability in lignocellulose

processing is desired because it allows for increased substrate solubility, favorable enzyme kinetics, and lowers the risk of contamination by unwanted organisms. Exploring environmental metagenomes of microbial communities from thermal sites across a wide spectrum of temperature and pH provides a unique opportunity to expand the repertoire of known extremozymes with lignocellulolytic functions.

In this study we utilized 71 publicly available metagenomes from 58 unique geothermal features to analyze the microbial diversity and CAZyme composition in hot spring systems. To our knowledge, this is the largest survey of hot spring metagenomes for biomass degradation potential. We connected microbial populations to high lignocellulolytic potential based on temperature and pH of the hot springs to identify target springs for future biotechnology applications. This work provides an initial exploratory approach to mining metagenomic datasets of a wide range of physicochemical conditions in hot springs. The diversity of biomass-degrading potential described here will provide new targets for future *in vitro* functional assays of thermophilic CAZymes that can be tested using methods such as synthetic biology in conjunction with functional assays.

Methods

Assembled metagenomic samples were retrieved from the JGI's Integrated Microbial Genomes and Microbiomes (IMG/M; Chen et al., 2021) database and followed the standard assembly and annotation protocol used for JGI-sequenced metagenome datasets at the time of submission. Corresponding assembly and annotation methods are listed in Supplementary Table 1 as provided by IMG/M. We acknowledge variation in these methods as a result of comparing datasets generated over a long range of time. Criteria for selection were as follows; Genomes

OnLine Database (GOLD) (Mukherjee et al., 2021) analysis project type as metagenomic analysis and GOLD ecosystem type as thermal springs. Repeated metagenome hot springs sites were selected based on sequencing size or presence of a linked publication in GOLD. Several sites were included as unique samples corresponding to different transects from the source water (*i.e.* Dewar Creek) or due to temporal sampling (*i.e.* Jinze). Metadata for spring location, temperature, and pH were retrieved from IMG and GOLD (Supplementary Table 1). For missing data, the corresponding publications were searched, or the PIs were contacted for unpublished samples. PIs of all published and unpublished metagenome datasets were contacted to acquire permission of data usage; all PIs gave permission to use their data.

To assess the taxonomic profile, Kraken2 (Wood et al., 2019) and Bracken (Lu et al., 2017) were used to assign the assembled contigs taxonomy and calculate the percentage of each taxa per metagenome, respectively. Kraken2's standard database was built using the --download-library switch with options bacteria and archaea. Each metagenome was run through Kraken2 using GNU parallel (Tange, 2018) and 8 threads per job using default kmer length of 35. Bracken was used for estimation of taxa abundance from Kraken2 reports using options -r 150, -l P, and -t 10. Diversity metrics of the taxonomic profile were calculated using R (R Core Team, 2014) package Phyloseq and statistical analyses with lme4 (Bates et al., 2015) and vegan (Oksanen et al., 2019) packages.

Lignocellulolytic enzymes were identified from the Carbohydrate Active Enzyme Database (CAZy.org, (Lombard et al., 2014)) using the dbCAN2 package (Zhang et al., 2018). The HMMER, HotPep, and DIAMOND tools were applied using default parameters to annotate CAZymes in the assembled metagenomic samples. Only CAZy gene hits that were present in at

least two of the three tools were considered for later analysis. All computing was performed on the NERSC Cori cluster. The CAZyme heatmap was generated using R package Pheatmap (Kolde, 2015) and Kolde 2015) with presence-absence normalized data from the decostand function from the vegan package and the heatmap was annotated with data available from IMG. An additional heatmap was generated for CAZyme relative abundance (Supplementary Figure 1). Four metagenomes were selected for deeper analysis. TaxonOID's 3300029977, 3300006865, and 3300007072 were selected because of their high abundance of glycosyl hydrolases (GH). While 330009455 also had a similar number of GHs, we did not include this sample due to it having the lowest sample collection temperature (28.5 °C). TaxonOID 3300029625 was included as the fourth sample because it contained a majority of the GHs but was also the closest clustered sample with acidic pH. The most abundant GH for each enzyme type was further described by selecting the gene identification number and searching the metagenome on IMG for the corresponding contig. Out of the three GHs (GH3, GH5 and GH10), the main catalytic Pfam domain was selected and those sequences were queried in IMG for taxonomic affiliation. The data containing enzyme class, taxonomic distribution, and metagenome location was illustrated in a Sankey diagram using R package networkD3 (Allaire et al., 2017).

Results and discussion

Metagenomic dataset

Metadata filtering to retrieve hot spring metagenomes from the Integrated Microbial Genomes and Microbiomes system (IMG/M) (Chen et al., 2021) yielded 71 assembled metagenomic samples from 58 unique hot springs (Supplementary Table 1). Several sites had multiple samples included if distinct areas of the hot spring were sampled (*e.g.*, Dewar Creek

transects) or when sampling occurred in consecutive years (*e.g.*, Jinze). Geothermal features were distributed across six countries, with the majority located in the USA (Figure 1); this is because of the long tradition of hot spring research in Yellowstone National Park (YNP), which hosts approximately 55 % of the world's geothermal features. The temperature and pH ranges of the samples were 28.5 - 90 °C and 2.1 - 9.7, respectively (Figure 1). Metagenomic sequencing size varied considerably among samples with a median assembly size of 49.4 Mb. The wide range of assembly size (3.2 - 1,666.8 Mb) is attributed to the different sequencing technologies and facilities used. As IMG/M is a repository for metagenomic data, not all samples have been processed the same way prior to annotation. All metagenomic sequencing metadata of samples used in this study has been collected in Supplementary Table 1.

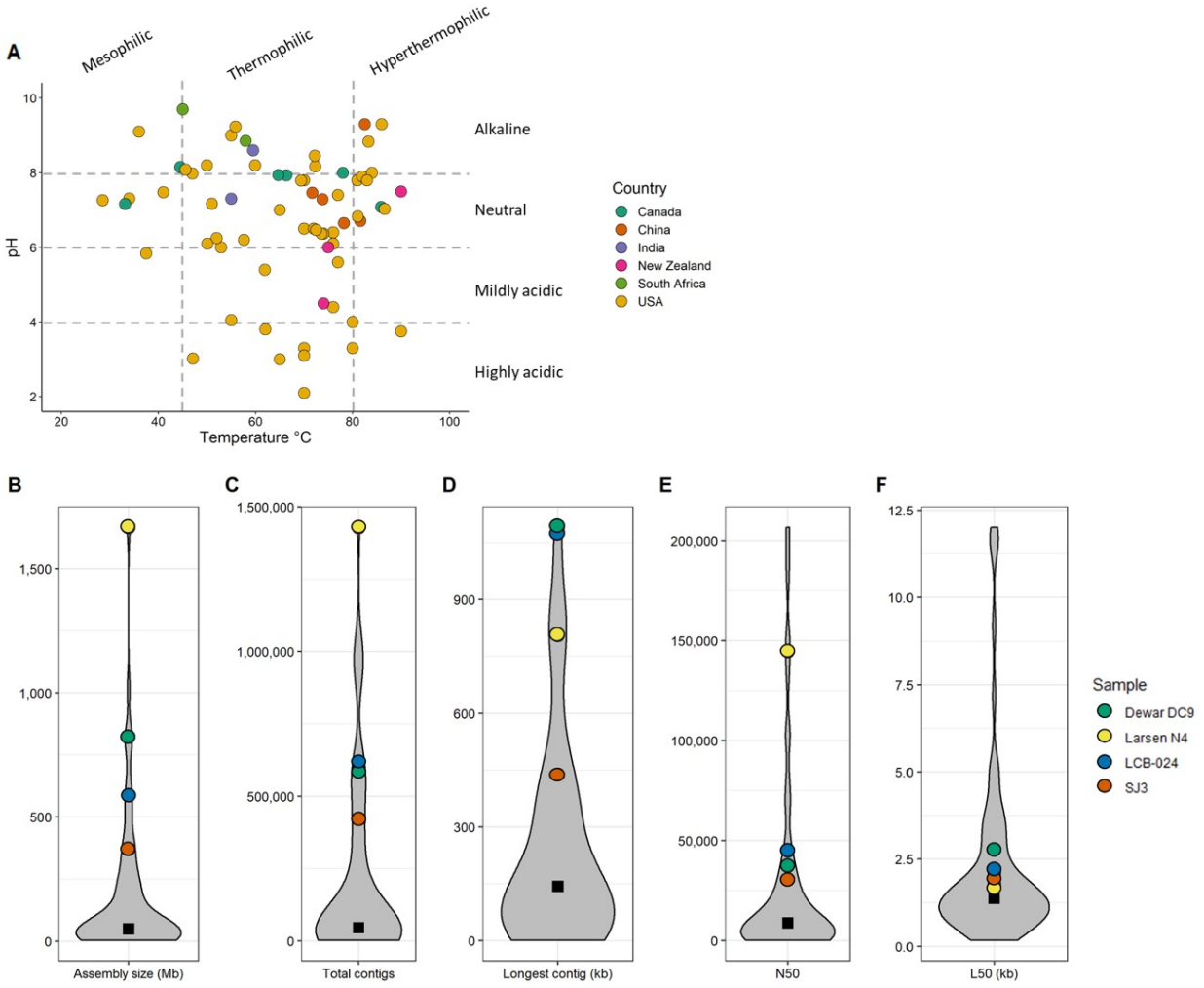


Figure 1. (A) Temperature and pH scatter plot of all 71 metagenomes used in this study. Points are colored by country of sample origin. Thick grey lines denote the temperature and pH categories used. Temperature categories: mesophilic (< 45.0 °C), thermophilic (45.0 - 79.9 °C), and hyperthermophilic (> 80.0 °C). pH categories: highly acidic (< 4.0), mildly acidic (4.0 – 5.9), neutral (6.0 – 7.9), and alkaline (> 8.0). (B-F) Violin plots of sequencing statistics for all 71 metagenomes. The four metagenomes of particular interest to this study are highlighted. Black square denotes median value.

Taxonomic profile

Kraken2 (Wood et al., 2019) was used to assign taxonomy to the assembled contigs followed by processing with Bracken (Lu et al., 2017) to calculate the relative abundances of taxa in each metagenomic sample. We first categorized the metagenomic datasets based on temperature and pH to identify taxonomic trends due to these parameters. The temperature categories were mesophilic (< 45.0 °C), thermophilic (45.0 - 79.9 °C), and hyperthermophilic (> 80.0 °C). The pH categories were highly acidic (< 4.0), mildly acidic (4.0 – 5.9), neutral (6.0 – 7.9), and alkaline (> 8.0). The description of sites within each category is summarized in Supplementary Table 1. Overall, at the phylum level, contigs were identified most frequently as being derived from members of *Proteobacteria*, followed by *Firmicutes*, *Actinobacteria*, *Aquificae*, and *Bacteroidetes*. Out of all contigs that could be classified using the Kraken2 standard database, 45 archaeal and bacterial phyla were identified in total across all metagenomes (Supplementary Table 2). *Proteobacteria* dominated mesophilic and thermophilic samples while *Aquificae* was the most abundant phylum in hyperthermophilic samples; the latter has been well documented for high temperature streamer communities in YNP (Inskeep et al., 2013a, 2013b). The phylum *Crenarchaeota* was also most abundant in highly acidic and mildly acidic samples, while *Proteobacteria* were the dominant phylum in neutral and alkaline samples. Microbes belonging to *Aquificae* and *Crenarchaeota* have been shown to be widespread in geothermal environments with genus-specific adaptations to narrow temperature or pH conditions (Hou et al., 2013). Phylum level taxonomy can highlight overall trends but to understand how microbial populations are distributed across temperature and pH conditions, finer taxonomic resolution is required.

Across all 71 metagenomes, the assembled contigs were assigned to 1,474 unique genera. In the mesophilic category, *Streptomyces* was the most abundant genus, closely followed by *Roseiflexus* and *Pseudomonas*. In thermophilic samples, there was a taxonomic shift towards *Thermus*, *Sulfurihydrogenibium*, *Thermocrinis*, and *Bacillus* assigned contigs. While contigs assigned to *Thermocrinis* were still in moderately low abundance in thermophilic samples, they were the top contig assignment for hyperthermophilic samples followed by *Pyrobaculum* and *Thermus*. Highly acidic metagenomes contained contigs assigned to *Hydrogenobaculum*, *Sulfolobus,* and *Acidolobus,* among others. Consistent with reports that *Hydrogenobaculum* is the dominant genus of *Aquificae* in thermophilic, acidic hot springs (55.1- 64.5 °C, pH 2.5 - 2.6; (Hou et al., 2013; Takacs-Vesbach et al., 2013)), members of this genus were abundant in metagenomes of the highly acidic category. Members of the crenarchaeotal genus *Sulfolobus* are common and abundant in acidic hot springs in both YNP and Chinese sites (Inskeep et al., 2010; Song et al., 2013).

Our results are largely consistent for most taxa with the findings of Inskeep et al., who previously analyzed 20 unique features in YNP (Inskeep et al., 2013a, 2013b). However, our expanded dataset, which consisted of 58 unique features, including features outside YNP, included a higher abundance of *Sulfurihydrogenibium* in acidic sites than previously observed. In the mildly acidic range, *Acidilobus* was dominant, followed by *Streptomyces* which was more abundant than it was in highly acidic samples and remained abundant, although not dominant, in the metagenomes from pH-neutral sites. The pH-neutral samples were dominated by genera *Thermocrinis* and *Thermus*. A study of 16 Chinese hot springs detected high amounts of *Thermocrinis* and *Thermus* sequences in high temperature, non-acidic samples (Song et al.,

2013); we observed similar trends for both taxa in similar temperature and pH categories. Finally, in alkaline features, metagenomes were taxonomically dominated by *Thermocrinis*-, *Bacillus*-, and *Thermus*-affiliated contigs.

Overall, mesophilic samples had the highest α-diversity followed by thermophilic, then hyperthermophilic samples. This result was supported by mesophilic samples being significantly higher than either thermophilic and hyperthermophilic samples for both observed taxa and Shannon Diversity ($p < 0.001$, Figure 2 A-B, Supplementary Table 3 lists diversity metrics for each metagenome while Supplementary Table 4 contains statistical comparisons and p values for each category tested). Several studies have found similar results of decreased α-diversity with increased temperature using 16S rRNA gene amplicon sequencing in hot springs (De León et al., 2013; Sharp et al., 2014; Lavrentyeva et al., 2018).

The α-diversity was lowest in the highly acidic samples but only for one of the diversity metrics tested. When pH of the samples increased, so did the α-diversity. However, the number of observed taxa for highly acidic metagenomes was not significant when compared to neutral or alkaline samples ($p = 0.0726$ and $p = 0.137$, respectively) and similarly, none of the pH categories were significantly different in terms of their Shannon Diversity ($p > 0.05$) with a p-value adjusted using the Holm method (Figure 2 A-B, Supplementary Table 4). The α-diversity measured in our dataset aligns with the results of a 16S rRNA genes amplicon study of 925 New Zealand hot springs in which the highest diversity was found at 21.5 °C and pH 6.4 with diversity decreasing toward higher temperatures and lower pH values (Power et al., 2018). The same study confirmed pH as the major driver of α-diversity in springs < 70 °C and that the pH of the system may affect the resultant nutrient and trace metal composition and availability.

Consistently, a study on *Sulfolobus islandicus* populations in YNP hot springs attempted to elucidate effects of spring size, geographical location, and overall nutrients on microbial populations but pH was identified as the main driver of population diversity (Campbell et al., 2017).

Microbial community composition was affected by the temperature of the hot spring across all three temperature categories and for some of the pH categories. For β-diversity, Bray Curtis dissimilarity was plotted by canonical correspondence analysis (CCA) (Figure 2 C-D) and the significance calculated by PERMANOVA. The β-diversity of all temperature categories were significantly different from each other in pairwise comparisons ($p = 0.003$). In contrast, the only pH comparisons that were significantly different were highly acidic samples to neutral and alkaline samples ($p = 0.006$) (Supplementary Table 4). As temperature increases and pH shifts to either end of the spectrum, microbial diversity has been observed to decrease in thermal systems (De León et al., 2013; Sharp et al., 2014; Menzel et al., 2015; Lavrentyeva et al., 2018; Power et al., 2018; Podar et al., 2020), an observation that is consistent with soils (Lauber et al., 2009). The notion that photosynthetic microorganisms are typically absent at temperatures > 74 °C due to the instability of pigments at such high temperatures explains the significance between thermophilic and hyperthermophilic communities. The microbial community in highly acidic sites was dominated by known acidophiles, specifically *Hydrogenobaculum*, *Sulfurihydrogenibium*, and *Sulfolobus,* which were the source of diversity compared to samples with pH > 6.0.

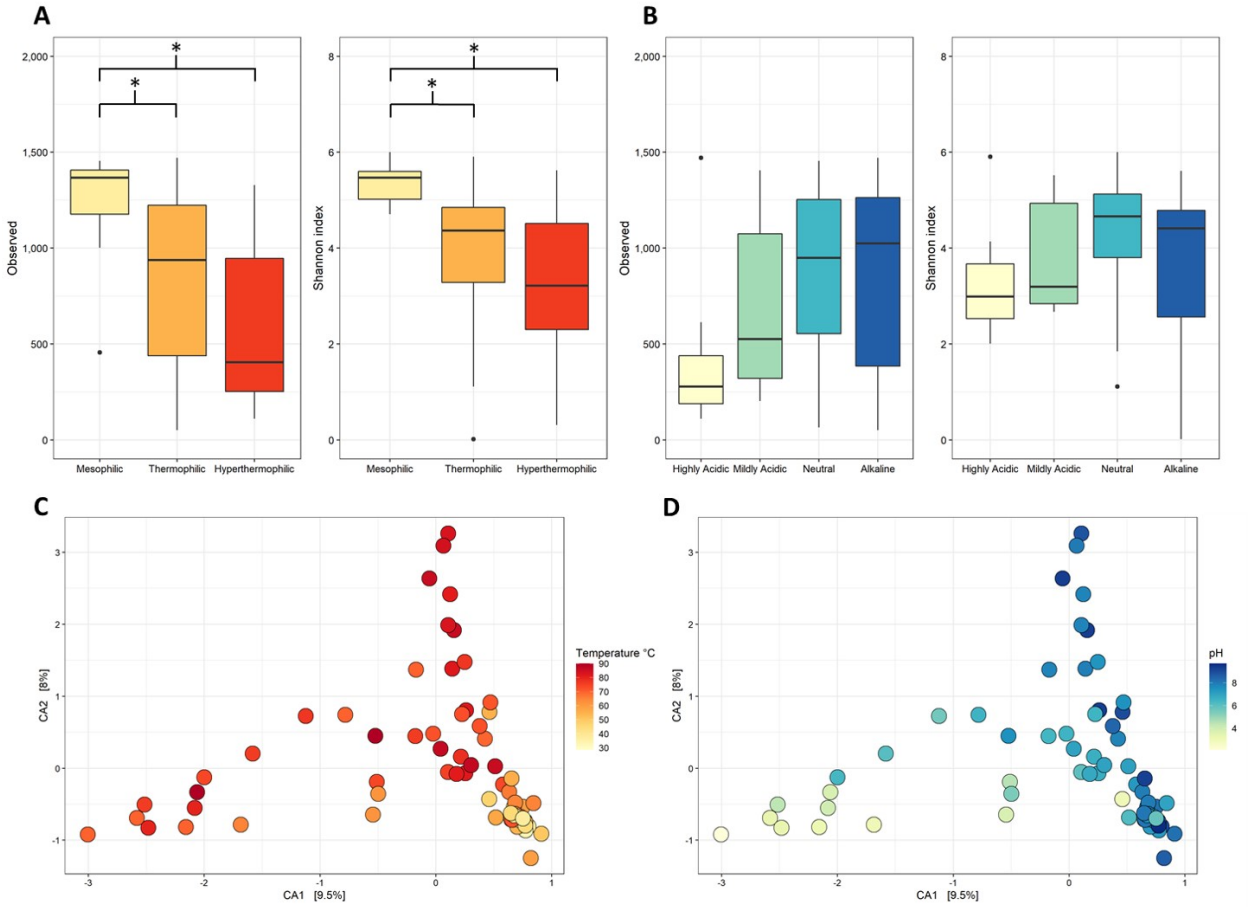


Figure 2. Observed taxa and Shannon Diversity metrics for the taxonomic profile assignment of metagenomic contigs at genus level classification. Taxonomy was assigned to metagenomic contigs using the Kraken2 standard database and contig abundances were calculated using Bracken. (A) α-diversity split by temperature categories, which were: mesophilic (< 45.0 °C), thermophilic (45.0 - 79.9 °C), and hyperthermophilic (> 80.0 °C). (B) α-diversity split by pH categories, which were: highly acidic (< 4.0), mildly acidic (4.0 – 5.9), neutral (6.0 – 7.9), and alkaline (> 8.0). * denotes significance $p<0.05$. Statistics calculated with pairwise comparisons and p values adjusted by the Holm method. (C-D) Canonical correspondence analysis of β-diversity calculated by Bray Curtis dissimilarity with points colored by sample temperature or pH, respectively.

Many studies have analyzed hot spring microbial communities but were either limited by the availability of metagenomic datasets or focused on 16S rRNA gene amplicon short reads, which are limited in their taxonomic resolution and potential for functional profiling. Here, we have described the richness, evenness, and composition of 71 metagenomic assemblies. Temperature and pH ranges influenced the microbial communities and significant differences were observed for both α- and β-diversity. The α-diversity was not significant between thermophilic (45.0 - 79.9 °C) and hyperthermophilic (> 80.0 °C) categories, however the microbial community analyzed by β-diversity revealed the composition of these temperatures to be significantly different ($p = 0.003$). In a previous 16S rRNA gene amplicon based study, pH was seen to be the main variable in community diversity below 70 °C, with temperature becoming the main driver of diversity above this threshold (Power et al., 2018). Similarly, a nonlinear trend of decreasing α-diversity was observed for several hot springs with a sharp drop in diversity above 80 °C (Podar et al., 2020) and no significance in a temperature range of 50 - 80 °C suggests the range of temperature or categories used for analysis is important to consider.

These results demonstrate the existence of unique microbial communities in physicochemically diverse conditions geothermal sites, suggesting that hot springs could provide a unique opportunity for novel CAZyme discovery.

CAZyme profile

To better understand the biomass degradation potential in each metagenome, CAZyme families related to cellulases, hemicellulases, and oligosaccharide-degrading enzymes were analyzed. The 71 metagenomes were examined with dbCAN2 for potential CAZymes, selecting only gene hits detected by at least two of the three methods utilized by dbCAN2 (Diamond,

Hotpep, and HMMER). Two additional metagenomes, obtained from the P3 gut compartment of a *Nasutitermes corniger* termite ((He et al., 2013), 2030936001) and a compost sample from the São Paulo Zoo ((Martins et al., 2013), 2209111003), were included to compare metagenomes from geothermal features to environmental systems known for their high rates of plant biomass degradation. The *N. corniger* gut metagenome, an anaerobic sample, contained high relative abundances of cellulases and hemicellulases that degrade components of the termite's wood-based diet (He et al., 2013). The aerobic compost metagenome was sampled from the São Paulo Zoo and contained enzymes for cellulose degradation along with additional CAZymes for lignin and pectin degradation (Martins et al., 2013).

The chemical nature and availability of carbohydrates pressures microbiome to adapt their CAZyme composition. For example, varying the feed or diet of ruminants resulted in differences in their microbiota and CAZyme families (He et al., 2013; Wong et al., 2017; Bohra et al., 2019). In addition to ruminants, soil (López-Mondéjar et al., 2020) and oil reservoir (Lewin et al., 2017) ecosystems have been well studied in regards to CAZymes potential but studies on hot spring CAZymes are lacking. The termite gut and compost metagenomes were included in our analysis as controls of high potential CAZyme systems able to degrade plant biomass, encoding 17 and 14 CAZyme families, respectively; they were excluded from clustering. When examining CAZyme families by relative abundance, the two non-geothermal metagenomes have high relative abundance of GH3 and GH5 genes compared to all CAZymes (Supplementary Figure 1). These two glycosyl hydrolases are important CAZyme families for oligosaccharide and cellulose degrading activity, respectively. Figure 3 shows the presence or absence and clustering of CAZymes attributed to major enzyme families for cellulase,

hemicellulase, and oligosaccharide degradation for the 71 hot spring metagenomes (Figure 3). CAZyme hits and corresponding IMG/M identification for all metagenomes are included in Supplementary Table 5. Three hot spring metagenomes that were also among the most deeply sequenced samples included in this study contained nearly all 20 CAZyme families, including the major families for cellulases, hemicellulases, and oligosaccharide-degrading enzymes. Two of these metagenomes, LCB-024 (3300029977, 69.4 °C, pH 7.79; Lynes *et al*., in preparation) and Dewar Creek DC9 (3300007072, 64.7 °C, pH 7.94), had been obtained from thermophilic, pH-neutral hot springs and contained 18 and 19 of the selected 20 CAZyme families, respectively. Dewar Creek DC9 is one of several samples collected from an interface between geothermal source water and soil. The third metagenome, Larsen N4, had been obtained from a mesophilic, pH-neutral geothermally-affected sediment (3300006865, 33.1 °C, pH 7.16) and contained 19 of the 20 CAZyme families. One additional metagenome from the mildly acidic hot spring SJ3 (3300029625, 61.9 °C, pH 5.4, (Colman et al., 2019)) was included for further analysis to expand the range of spring conditions analyzed and contained 17 of the 20 CAZymes (Figure 4). We analyzed these metagenomes on a community-wide level to determine the overall biomass-degrading potential of geothermal sites with high CAZyme variability. All sequencing statistics are visualized in Figure 1 and are detailed in Supplementary Table 1.

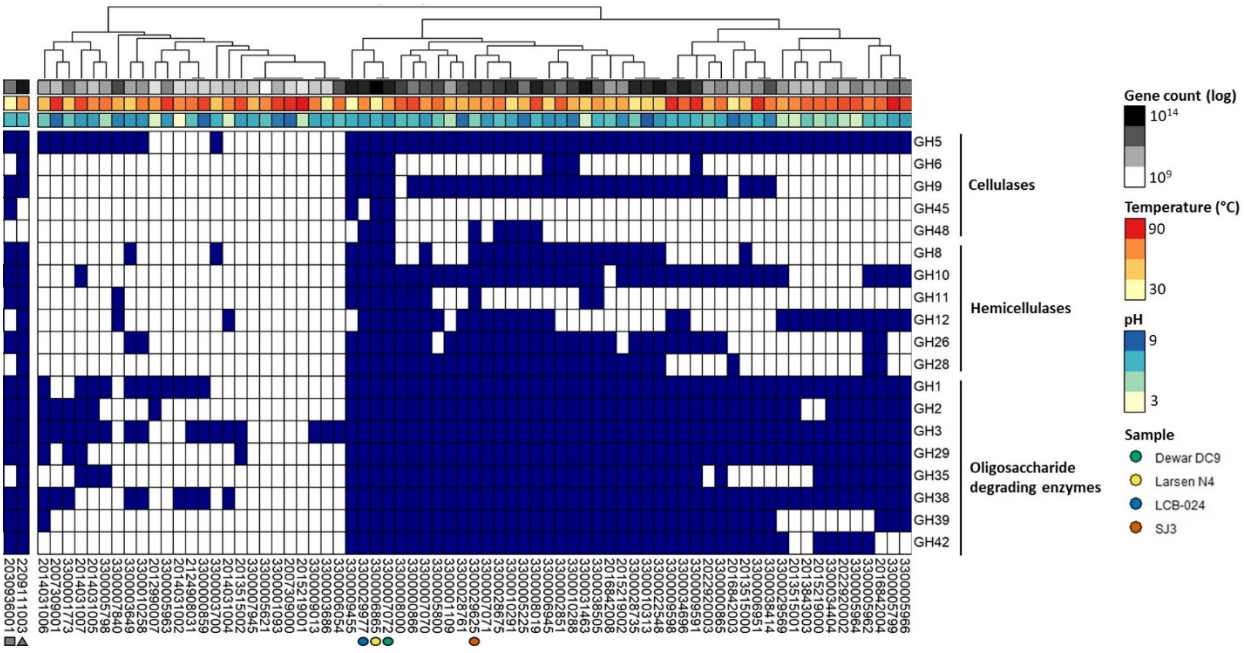


Figure 3. Heatmap showing presence-absence of specific CAZymes with predicted cellulase, hemicellulase, and oligosaccharide-degrading enzyme activities in the 71 metagenomes. Dark blue color represents the presence of a CAZyme gene sequence. Sample clustering was based on hierarchical cluster analysis of dissimilarity of the CAZyme presence-absence profile. Two metagenomes from non-hot spring lignocellulolytic environments are shown on the left and indicated with a grey square (2030936001 - *Nasutitermes corniger* P3 gut compartment) and triangle (2209111003 - São Paulo Zoo compost). Four hot spring metagenomes of particular interest to this study are indicated with colored circles. 3300029977 - LCB-024 - blue, 3300006865 - Larsen N4 – yellow, 3300007072 - Dewar Creek DC9 - green, and 3300029625 - SJ3 - orange.

These four metagenomes (Dewar Creek DC9, Larsen N4, LCB-024, and SJ3) contained a wide variety of biomass-degrading CAZymes and were further analyzed for the taxonomic affiliation of GH3, GH5, and GH10 containing contigs, which were the most abundant GH families for oligosaccharide-degrading enzymes, cellulases, and hemicellulases, respectively. Contig taxonomic identification was retrieved from IMG/M and plotted to show enzyme class and hot spring sample origin (Figure 4). Cellulase- and hemicellulase-containing contigs were assigned to fewer phyla (20 and 22 phyla respectively) than oligosaccharide-degrading enzyme contigs (39 phyla), but when phyla > 1 % relative abundance were plotted, all enzyme types were similar in the number of phyla represented (Figure 4, full table of contig taxonomy and relative percentages in Supplementary Table 6).

The relative abundance of unclassified CAZyme contigs varied among the four metagenomes. Dewar Creek DC9 contained the largest amount of such contigs at 50.1 % unable to be taxonomically classified in IMG/M, followed by SJ3 at 44.1 % and then Larsen N4 and LCB-024 at 38.0 % and 36.8 %, respectively. This highlights the novel taxonomic diversity that remains to be described and explored within hot spring systems. With more genomic screening, databases will be able to identify and assign taxonomy to contigs more readily, as well as recognize additional CAZyme genes encoded on the contigs.

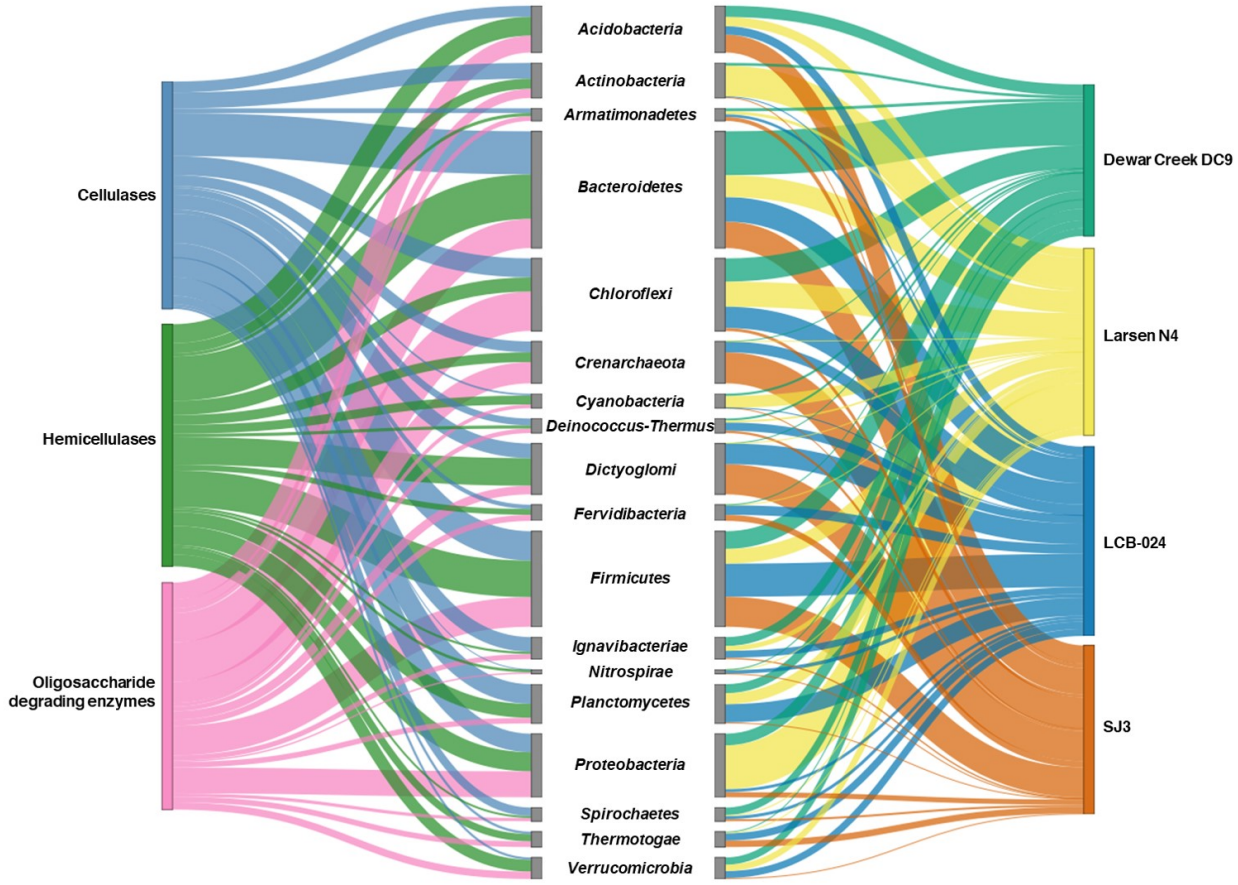


Figure 4. Sankey diagram illustrating taxonomic distribution of CAZyme-containing contigs and the hot spring metagenomes they originated from. Indicated on the left is the enzyme type determined by the most abundant Pfam domain for the most abundant glycosyl hydrolase family for contigs identified by dbCAN2 using the CAZy database (Pfam00150 - GH5 - Cellulase, Pfam00331 - GH10 - Hemicellulase, Pfam00933 - GH3 - Oligosaccharide-degrading enzymes). In the center is the taxonomic affiliation of the contigs determined by IMG. Taxonomically unclassified contigs were removed prior to plotting. Shown on the right is the hot spring metagenome location for the taxa. The height of the Sankey bars are assigned proportionately to the relative abundances of CAZyme-containing contigs and taxa.

Highly abundant taxa such as *Acidobacteria*, *Bacteroidetes*, *Chloroflexi*, and *Firmicutes* were common for all enzyme types while several low abundance species belonging to phyla *Aigarchaeota*, *Calescasmantes*, *Euryarchaeota*, and *Parcubacteria* were only linked to oligosaccharide degrading enzymes. In an earlier study of 5,123 sequenced bacterial genomes, 56 % of genomes were found to contain genes associated with β-glucosides while 24 % of the total genomes contained cellulases and β-glucosides (Berlemont and Martiny, 2013). Additionally, a comparative genomics study on 24 soil *Acidobacteria* isolate genomes found GH3 in all examined genomes while GH5, GH8, GH9, GH12, and GH44 were only found in select subdivisions of *Acidobacteria* (Eichorst et al., 2018).

The *Acidobacteria* CAZymes detected in the four metagenomes were identified to belong to cellulases, hemicellulases, and oligosaccharide-degrading enzymes. SJ3 had the largest relative abundance of *Acidobacteria* compared to the other hot springs, which could be attributed to the adaptation of *Acidobacteria* subfamily 1 to mildly acidic conditions (Jones et al., 2009; Campanharo et al., 2016). Additionally, various genera from *Acidobacteria* subfamily 1 have been shown to utilize cellulose and xylan as a growth substrate while other members of this family have not, demonstrating that further research on the biomass degradation potential of *Acidobacteria* is warranted (Kielak et al., 2016). *Actinobacteria* contigs contained a major portion of CAZymes for Larsen N4 with GH families in all three categories of enzymes. This abundance was greater in Larsen N4 compared to the other hot spring metagenomes which had higher temperatures. Larsen N4, with a temperature of 33.1 °C, was expected to have a different population of potential biomass degraders than the other springs that are above 60 °C as was also seen by the difference in taxonomic composition of its community. *Actinobacteria* enrichment

cultures from lower temperature systems like compost (Wang et al., 2016) and soil (Anderson et al., 2012) have been shown to have diverse CAZyme repertoire and cellulose degrading capabilities.

CAZymes encoded by members of the *Bacteroidetes* encompass all three enzyme families and were present across all four analyzed metagenomes. *Bacteroidetes* are a prolific lineage for biomass degradation and are found in a variety of environments including soil (Taillefer et al., 2018), compost (Wang et al., 2016), rumen (Opdahl et al., 2018), and hot springs (Eichorst et al., 2013). Recently, a novel member of the *Bacteroidetes* with the ability to hydrolyze xylan was isolated using crystalline cellulose (Liu et al., 2020). In contrast, other *Bacteroidetes* from similar environments did not grow on crystalline cellulose but were part of late-stage enrichments suggesting soluble products as their growth substrate (Eichorst et al., 2013). The variability of biomass degradation potential within the *Bacteroidetes* highlights this phylum-level lineage as a promising target for the identification of novel CAZymes in future research. In a recent study, proteomics was used to examine the enzymatic activity of CAZymes of two species grown on both cellulose and pectin. Expression of endoglucanases and β-glucosidases varied between the two and pending time of incubation (Taillefer et al., 2018).

Contigs attributed to *Chloroflexi* contained many oligosaccharide-degrading enzymes. Members of the *Chloroflexi* have been enriched and isolated from biomass-rich systems (*e.g.* wastewater and sludge samples) and had been seen in high abundances in late stage enrichments (Ishii et al., 2008), suggesting that some *Chloroflexi* are able to scavenge the breakdown products of cellulose using oligosaccharide-degrading enzymes. The decreased photosynthetic temperature limit under acidic conditions (Boyd et al., 2012; Inskeep et al., 2013b) could be one

explanation for the decreased abundance of *Chloroflexi* in SJ3 relative to the other springs as many taxa of *Chloroflexi* from hot springs have phototrophic metabolisms (Klatt et al., 2011, 2013). We observed *Dictyoglomus* in the LCB-024 and SJ3 metagenomes with a majority of their contigs being assigned to contain hemicellulases. *Dictyoglomus* isolates have been shown to produce multiple xylanases (Mathrani and Ahring, 1992) and have the ability to grow on a variety of carbon polymers (Saiki et al., 1985; Brumm et al., 2016). Interestingly, certain *Dictyoglomus* also encode glycosyl hydrolases with both cellulase and hemicellulase activity and are able to degrade carboxymethyl cellulose and mannans (Fusco et al., 2018). A broad pH growth range (pH 5-9) and high temperature optimum (72-80 °C) with broad cellulase and hemicellulase activity highlights *Dictyoglomus* as an attractive organism for industrial lignocellulose degradation. Out of the 18 phyla present at > 1 % abundance in the four metagenomes, only *Fervidibacteria* currently lacks an isolated representative. However, *Fervidibacteria* has conflicting reports of biomass degradation potential. On the one hand, members of the *Fervidibacteria* have been enriched from a switchgrass inoculum (Peacock et al., 2013); on the other hand, hot spring (FS5, 74 °C, pH 8.2) *Fervidibacteria* exhibited decreased translational activity in the presence of cellobiose or cellulose as compared to when these substrates were absent (Reichart et al., 2020).

Firmicutes are a high priority target for biomass conversion due to consolidated bioprocessing (CBP) of lignocellulose occurring in a single organism without the need for pretreatment and some of the most thermophilic cellulolytic organisms identified so far are within this phylum. *Firmicutes* contigs were abundant in the four metagenomes and contained genomic potential for all three enzyme types. The class *Clostridia* is anaerobic and can utilize a

multi-enzyme structure called a cellulosome to stay bound to cellulose during degradation. The sugars cleaved from cellulose can then be fermented by *Caldicellulosiruptor thermophilum* producing lactate, ethanol, and acetate as end-products (Yang et al., 2009). *Caldicellulosiruptor* has been a target of metabolic engineering to increase rates of conversion of plant biomass to ethanol due to their potential for CBP. When (Kim et al., 2018) expressed a cellobiose phosphorylase in *Caldicellulosiruptor bescii*, inhibition of cellulases by cellobiose was decreased, which in turn increased the cellulolytic activity. (Chung et al., 2014) transformed a strain of *C. bescii* lacking lactate dehydrogenase to contain a plasmid-prone acetaldehyde/alcohol dehydrogenase gene to produce over 14 mM of ethanol from a 1 % cellobiose solution compared to no ethanol production from the wild type strain. Another *Firmicutes* consortium enriched from a hot spring sample was able to degrade a variety of substrates and produced ethanol. This led to a shift in community composition over a 7-day incubation period, suggesting synergism and specialization of individual community members to specific steps during the process (Zhao et al., 2014). While a majority of the contigs remain taxonomically unclassified in IMG/M, many of the detected taxa are linked to species with biomass-degrading activity. Together, these results demonstrate the diversity of CAZyme-containing bacteria in hot springs and highlight the attractiveness of these extreme ecosystems for future biotechnological research.

Owing mainly to the lack in archaeal isolates (Baker et al., 2020; Sun et al., 2020), archaeal CAZymes have drastically lagged characterization of their bacterial counterparts. At the time of publication, the CAZy database contained 18,171 bacterial genomes and 392 archaeal genomes. The majority of archaea capable of cellulose degradation are affiliated to *Crenarchaeota* and *Euryarchaeota*, which also contain the majority of archaeal isolates, and

have optimal temperatures ranging from 80 to 115 °C when grown on cellulose (Suleiman et al., 2020). Isolates and enriched communities of *Crenarchaeota* have been recovered from several hyperthermophilic sources, including terrestrial hot springs (Perevalova et al., 2010; Graham et al., 2011; Zayulina et al., 2020). While we did not recover CAZyme containing contigs affiliated with *Euryarchaeota*, crenarchaeotal contigs were detected in LCB-024 and SJ3, with the latter hot spring containing more such contigs. SJ3 contigs were affiliated to genera *Caldivirga*, *Ignisphaera*, and *Thermofilum* and contained all three enzyme types. *Crenarchaeotal* contigs from LCB-024 were affiliated to genera *Ignisphaera*, *Thermofilum*, and *Thermosphaera* and encoded cellulases and oligosaccharide degrading enzymes. While most of these crenarchaeotal genera have a broad growth range of 60 - 98 °C and are adapted to circumneutral pH (Huber et al., 1998; Niederberger et al., 2006; Toshchakov et al., 2015), *Caldivirga* isolates grow optimally at pH 4 (Itoh et al., 1999). These observations are consistent with a higher abundance of *Caldivirga*-associated contigs in the metagenome of the mildly acidic SJ3 site.

While this study focused on glycosyl hydrolases for biomass degradation, the output from dbCAN2 also included multiple hits corresponding to pectinases, peroxidases, and laccases that should also be examined for potential novel targets of lignin, and other recalcitrant biomass degradation of biotechnological impact. Carbohydrate-binding modules (CBM) for binding domains also provide information about the proximity of organisms to biomass for degradation as they would be bound to carbohydrate particles. The catalytic domains for lignin degradation are classified into several auxiliary activity (AA) families that have not been analyzed here. In addition, carboxyl esterase (CE) families are an alternative way to cleave bonds composing lignocellulosic biomass. One limitation of our study is that it did not examine the presence of

polysaccharide utilization loci (*e.g.*, PUL and cellulosome loci via dbCAN-PUL; (Ausland et al., 2021)); the analyses of these enzyme families were outside the scope of this work. However, the broad diversity in geochemistry and microbiology of hot springs and results reported herein suggest that geothermal sites are a promising source for future biotechnological applications. Another limitation is that the metagenomes investigated herein were generated over 13 years (2007-2020). Unfortunately, this made certain comparisons difficult because datasets were generated using dramatically different sequencing technologies and pipelines; this variation sometimes prohibited us from making as clear recommendation for future research into hot spring CAZymes as we initially hoped to make.

Outlook

Hot springs are promising ecosystems for the discovery of biomass-degrading enzymes with high potential for unclassified or novel enzyme activities. The unique physicochemical extremes of thermal features make them an ideal testbed for identifying the next generation of highly lignocellulolytic organisms and their enzymes. In this study, we analyzed 71 metagenomes with a focus on four datasets from hot springs with mildly acidic to neutral pH. While these four metagenomes were among the most deeply sequenced, we stress that further bioprospecting of hot springs should occur across all temperature and pH ranges, given enough sequencing coverage. At this point, no particular geochemical niche (*e.g.*, specific range of temperature or pH) emerged as exceptional site for investigations into biomass degradation; we thus recommend being as inclusive as possible in future enzyme discovery studies. Improvements in genomic library creation and high throughput sequencing have provided an opportunity to query communities from low biomass hot spring samples at greater depth and

coverage for metagenomics (DeCastro et al., 2016). Challenging to cultivate organisms can now be accessed via single cell genomics (Rinke et al., 2014; Alteio et al., 2020) and metagenomics and genes of high interest can be expressed in a wider range of amenable hosts (Girfoglio et al., 2012; Wang et al., 2019) for lab characterization via functional metagenomics (Milshteyn et al., 2014; Van Der Helm et al., 2018). In order to screen the large diversity of yet uncharacterized CAZymes found in hot springs and other ecosystems, we recommend to combine the above approaches with high through-put screening tools, such as microfluidics-based functional enzyme assays, capable of screening thousands to tens of thousands of enzymes in parallel (Brouzes et al., 2009; Leung et al., 2012; Terekhov et al., 2017, 2018).

To help advance future bioprospecting studies, we call for considerable improvements to the minimum metadata collected, especially for geochemical parameters of sampling locations. More metadata would allow to contextualize and more directly compare different metagenome datasets as well as provide opportunities to identify possible trends of microbial community composition and function across various sampling efforts and studies. Several studies have attempted to test whether the microbial community composition of hot springs can be predicted based on geochemical parameters (Inskeep et al., 2013b; Campbell et al., 2017; Power et al., 2018; Kochetkova et al., 2020) but a lack in contextual data make comparison across datasets obtained by different laboratories challenging. Other than pH, temperature, location, and date of sample acquisition, which already need to be reported if a dataset is to be added to IMG/M, valuable parameters to report in future studies should include dissolved oxygen, salinity, sulfide, total and dissolved (in)organic carbon and nitrogen, as well as trace metal availability. Similarly, we were unable to comment on the presence of plant biomass (*e.g.*, tree trunks, pinecones, grass)

in specific features. While photos were available for most hot springs in primary publications or online databases, the exact location of DNA sample acquisition was typically not included; because of this, it was impossible to correlate the presence of plant biomass with the presence of cellulose degraders.

The majority of biomass-degrading enzymes in hot springs remains to be discovered and biochemically characterized. Metagenome-based discovery studies of CAZymes will benefit from further exploration and expansion of databases to link functional activity to genomic predictions. With continued focus on underexplored environments and taxa, novel or higher efficiency CAZymes will be discovered to aid in the industrial processing of lignocellulosic biomass for more efficient biofuel conversion.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquires can be directed to the corresponding author/s.

Author contributions

NJR, TW, and RH conceptualized and designed the study. NJR and RMB performed the data analyses. NJR drafted the manuscript. All authors edited the manuscript.

Funding

This work was supported through a NASA Exobiology award (80NSSC19K1633) to RH. A portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative (award 503546 to RH) and used resources at the Department of

Energy (DOE) Joint Genome Institute (JGI), the National Energy Research Scientific Computing Center (NERSC), and the Environmental Molecular Sciences Laboratory (EMSL), which are DOE Office of Science User Facilities. These facilities were sponsored by the Office of Biological and Environmental Research and operated under Contract Nos. DE-AC02-05CH11231 (JGI and NERSC) and DE-AC05-76RL01830 (EMSL). This material is based upon work supported by the U.S. DOE, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education (ORISE) for the DOE. ORISE is managed by ORAU under Contract No. DE-SC0014664.

Acknowledgements

We thank Hatzenpichler lab member Mackenzie Lynes for generating metagenomic and geochemical data for hot spring LCB-024. We also thank Drs. Eric Boyd and Daniel Colman for permitting analysis of metagenomic data from hot spring SJ3 and Drs. Allyson Brady, Peter Dunfield, and Ramunas Stepanauskas for allowing us to analyze metagenomes of features Larsen N4 and Dewar Creek DC9. We thank BC Parks (British Columbia, Canada) and the Ktunaxa Nation for their cooperating in sampling the Dewar Creek spring. We appreciate the US National Park Service, in particular Annie Carlson at the Yellowstone Center for Resources, for permitting work in YNP under Permit No. YELL-SCI-8010.

Supplementary material

The Supplementary Material for this article can be found online at:
https://www.frontiersin.org/articles/10.3389/fmib.2021.668238/full#supplementary-material

REFERENCES CITED

Allaire, J. J., Ellis, P., Gandrud, C., Kuo, K., Lewis, B. W., Owen, J., et al. (2017). networkD3: JavaScript Network Graphs from R. Available at: https://cran.r-project.org/package=networkD3.

Alteio, L. V., Schulz, F., Seshadri, R., Varghese, N., Rodriguez-Reillo, W., Ryan, E., et al. (2020). Complementary Metagenomic Approaches Improve Reconstruction of Microbial Diversity in a Forest Soil. *mSystems* 5, 1–18. doi:10.1128/msystems.00768-19.

Anderson, I., Abt, B., Lykidis, A., Klenk, H. P., Kyrpides, N., and Ivanova, N. (2012). Genomics of aerobic cellulose utilization systems in actinobacteria. *PLoS One* 7. doi:10.1371/journal.pone.0039331.

Ausland, C., Zheng, J., Yi, H., Yang, B., Li, T., Feng, X., et al. (2021). dbCAN-PUL: A database of experimentally characterized CAZyme gene clusters and their substrates. *Nucleic Acids Res.* 49, D523–D528. doi:10.1093/nar/gkaa742.

Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., and Lloyd, K. G. (2020). Diversity, ecology and evolution of Archaea. *Nat. Microbiol.* 5, 887–900. doi:10.1038/s41564-020-0715-z.

Bala, A., and Singh, B. (2019). Development of an environmental-benign process for efficient pretreatment and saccharification of Saccharum biomasses for bioethanol production. *Renew. Energy* 130, 12–24. doi:10.1016/j.renene.2018.06.033.

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67. doi:10.18637/jss.v067.i01.

Bayer, E. A., Shoham, Y., and Lamed, R. (2013). "Lignocellulose-Decomposing Bacteria and Their Enzyme Systems," in *The Prokaryotes: Prokaryotic Physiology and Biochemistry*, 1–662. doi:10.1007/978-3-642-30141-4.

Berlemont, R., and Martiny, A. C. (2013). Phylogenetic Distribution of Potential Cellulases in Bacteria. *Appl. Environ. Microbiol.* 79, 1545–1554. doi:10.1128/AEM.03305-12.

Bohra, V., Dafale, N. A., and Purohit, H. J. (2019). Understanding the alteration in rumen microbiome and CAZymes profile with diet and host through comparative metagenomic approach. *Arch. Microbiol.* 201, 1385–1397. doi:10.1007/s00203-019-01706-z.

Boyd, E. S., Fecteau, K. M., Havig, J. R., Shock, E. L., and Peters, J. W. (2012). Modeling the habitat range of phototrophs in Yellowstone National Park: Toward the development of a comprehensive fitness landscape. *Front. Microbiol.* 3, 1–11. doi:10.3389/fmicb.2012.00221.

Brouzes, E., Medkova, M., Savenelli, N., Marran, D., Twardowski, M., Hutchison, J. B., et al. (2009). Droplet microfluidic technology for single-cell high-throughput screening. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14195–14200. doi:10.1073/pnas.0903542106.

Brumm, P. J., Gowda, K., Robb, F. T., and Mead, D. A. (2016). The Complete Genome Sequence of Hyperthermophile Dictyoglomus turgidum DSM 6724 ™ Reveals a Specialized Carbohydrate Fermentor. *Front. Microbiol.* 7. doi:10.3389/fmicb.2016.01979.

Brunecky, R., Alahuhta, M., Xu, Q., Donohoe, B. S., Crowley, M. F., Kataeva, I. A., et al. (2013). Revealing Nature's Cellulase Diversity: The Digestion Mechanism of Caldicellulosiruptor bescii CelA. *Science* 342, 1513–1516.

Bugg, T. D. H., Ahmad, M., Hardiman, E. M., and Rahmanpour, R. (2011). Pathways for

degradation of lignin in bacteria and fungi. *Nat. Prod. Rep.* 28, 1883–1896. doi:10.1039/c1np00042j.

Cabrera, M. Á., and Blamey, J. M. (2018). Biotechnological applications of archaeal enzymes from extreme environments. *Biol. Res.* 51, 1–15. doi:10.1186/s40659-018-0186-3.

Campanharo, J. C., Kielak, A. M., Castellane, T. C. L., Kuramae, E. E., and Lemos, E. G. de M. (2016). Optimized medium culture for Acidobacteria subdivision 1 strains. *FEMS Microbiol. Lett.* 363, 1–7. doi:10.1093/femsle/fnw245.

Campbell, K. M., Kouris, A., England, W., Anderson, R. E., McCleskey, R. B., Nordstrom, D. K., et al. (2017). Sulfolobus islandicus meta-populations in Yellowstone National Park hot springs. *Environ. Microbiol.* 19, 2334–2347. doi:10.1111/1462-2920.13728.

Cheah, W. Y., Sankaran, R., Show, P. L., Ibrahim, T. N. B. T., Chew, K. W., Culaba, A., et al. (2020). Pretreatment methods for lignocellulosic biofuels production: Current advances, challenges and future prospects. *Biofuel Res. J.* 7, 1115–1127. doi:10.18331/BRJ2020.7.1.4.

Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2021). The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* 49, D751–D763. doi:10.1093/nar/gkaa939.

Chung, D., Cha, M., Guss, A. M., and Westpheling, J. (2014). Direct conversion of plant biomass to ethanol by engineered Caldicellulosiruptor bescii. *Proc. Natl. Acad. Sci.* 111, 8931–8936. doi:10.1073/pnas.1402210111.

Colman, D. R., Lindsay, M. R., and Boyd, E. S. (2019). Mixing of meteoric and geothermal fluids supports hyperdiverse chemosynthetic hydrothermal communities. *Nat. Commun.* 10. doi:10.1038/s41467-019-08499-1.

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* 3, 853–859. doi:10.1016/S0969-2126(01)00220-9.

De León, K. B., Gerlach, R., Peyton, B. M., and Fields, M. W. (2013). Archaeal and bacterial communities in three alkaline hot springs in heart lake geyser basin, yellowstone national park. *Front. Microbiol.* 4, 1–10. doi:10.3389/fmicb.2013.00330.

DeCastro, M. E., Rodríguez-Belmonte, E., and González-Siso, M. I. (2016). Metagenomics of thermophiles with a focus on discovery of novel thermozymes. *Front. Microbiol.* 7, 1–21. doi:10.3389/fmicb.2016.01521.

Doud, D. F. R., Bowers, R. M., Schulz, F., Raad, M. De, Deng, K., Tarver, A., et al. (2019). Function-driven single-cell genomics uncovers cellulose-degrading bacteria from the rare biosphere. *ISME J.* doi:10.1038/s41396-019-0557-y.

Eichorst, S. A., Trojan, D., Roux, S., Herbold, C., Rattei, T., and Woebken, D. (2018). Genomic insights into the Acidobacteria reveal strategies for their success in terrestrial environments. *Environ. Microbiol.* 20, 1041–1063. doi:10.1111/1462-2920.14043.

Eichorst, S. A., Varanasi, P., Stavila, V., Zemla, M., Auer, M., Singh, S., et al. (2013). Community dynamics of cellulose-adapted thermophilic bacterial consortia. 15, 2573–2587. doi:10.1111/1462-2920.12159.

Fusco, F. A., Ronca, R., Fiorentino, G., Pedone, E., Contursi, P., Bartolucci, S., et al. (2018). Biochemical characterization of a thermostable endomannanase / endoglucanase from Dictyoglomus turgidum. *Extremophiles* 22, 131–140. doi:10.1007/s00792-017-0983-6.

Girfoglio, M., Rossi, M., and Cannio, R. (2012). Cellulose degradation by Sulfolobus solfataricus requires a cell-anchored endo-β-1-4-glucanase. *J. Bacteriol.* 194, 5091–5100.

doi:10.1128/JB.00672-12.
Graham, J. E., Clark, M. E., Nadler, D. C., Huffer, S., Chokhawala, H. A., Rowland, S. E., et al. (2011). Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat. Commun.* 2, 375–379. doi:10.1038/ncomms1373.
Gupta, A., and Verma, J. P. (2015). Sustainable bio-ethanol production from agro-residues: A review. *Renew. Sustain. Energy Rev.* 41, 550–567. doi:10.1016/j.rser.2014.08.032.
He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R. H., et al. (2013). Comparative Metagenomic and Metatranscriptomic Analysis of Hindgut Paunch Microbiota in Wood- and Dung-Feeding Higher Termites. *PLoS One* 8. doi:10.1371/journal.pone.0061126.
He, Y.-C., Ma, C.-L., and Yang, B. (2018). "Pretreatment Process and Its Synergistic Effects on Enzymatic Digestion of Lignocellulosic Material," in *Fungal Cellulolytic Enzymes: Microbial Production and Application*, 1–25. doi:10.1007/978-981-13-0749-2.
Heins, R. A., Cheng, X., Nath, S., Deng, K., Bowen, B. P., Chivian, D. C., et al. (2014). Phylogenomically Guided Identi fi cation of Industrially Relevant GH1 β - Glucosidases through DNA Synthesis and Nanostructure-Initiator Mass Spectrometry. doi:10.1021/cb500244v.
Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P. H., et al. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* 5. doi:10.1038/srep14567.
Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., et al. (2011). Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science (80-.).* 463, 463–467. doi:10.1126/science.1200387.
Hou, W., Wang, S., Dong, H., Jiang, H., Briggs, B. R., Peacock, J. P., et al. (2013). A Comprehensive Census of Microbial Diversity in Hot Springs of Tengchong, Yunnan Province China Using 16S rRNA Gene Pyrosequencing. *PLoS One* 8. doi:10.1371/journal.pone.0053350.
Huber, R., Dyba, D., Huber, H., Burggraf, S., and Rachel, R. (1998). Sulfur-inhibited Thermosphaera aggregans sp. nov., a new genus of hyperthermophilic archaea isolated after its prediction from environmentally derived 16S rRNA sequences. *Int. J. Syst. Bacteriol.* 48, 31–38. doi:10.1099/00207713-48-1-31.
Inskeep, W. P., Jay, Z. J., Herrgard, M. J., Kozubal, M. A., Rusch, D. B., Tringe, S. G., et al. (2013a). Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. *Front. Microbiol.* 4. doi:10.3389/fmicb.2013.00095.
Inskeep, W. P., Jay, Z. J., Tringe, S. G., Herrgård, M. J., and Rusch, D. B. (2013b). The YNP metagenome project: Environmental parameters responsible for microbial distribution in the yellowstone geothermal ecosystem. *Front. Microbiol.* 4, 1–15. doi:10.3389/fmicb.2013.00067.
Inskeep, W. P., Rusch, D. B., Jay, Z. J., Herrgard, M. J., Kozubal, M. A., Richardson, T. H., et al. (2010). Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS One* 5. doi:10.1371/journal.pone.0009773.
Ishii, S., Shimoyama, T., Hotta, Y., and Watanabe, K. (2008). Characterization of a filamentous

biofilm community established in a cellulose-fed microbial fuel cell. *BMC Microbiol.* 8, 1–12. doi:10.1186/1471-2180-8-6.

Itoh, T., Suzuki, K., and Sanchez, P. C. (1999). new genus of rod-shaped crenarchaeote isolated from a hot spring in the Philippines. *Int. J. Syst. Bacteriol.*, 1–7.

Jones, R. T., Robeson, M. S., Lauber, C. L., Hamady, M., Knight, R., and Fierer, N. (2009). A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J.* 3, 442–453. doi:10.1038/ismej.2008.127.

Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A., and Kuramae, E. E. (2016). The ecology of Acidobacteria: Moving beyond genes and genomes. *Front. Microbiol.* 7, 1–16. doi:10.3389/fmicb.2016.00744.

Kim, S. K., Himmel, M. E., Bomble, Y. J., and Westpheling, J. (2018). Expression of a cellobiose phosphorylase from Thermotoga maritima in Caldicellulosiruptor bescii improves the phosphorolytic pathway and results in a dramatic increase in cellulolytic activity. *Appl. Environ. Microbiol.* 84, 1–9. doi:10.1128/AEM.02348-17.

Klatt, C. G., Inskeep, W. P., Herrgard, M. J., Jay, Z. J., Rusch, D. B., Tringe, S. G., et al. (2013). Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Front. Microbiol.* 4, 1–23. doi:10.3389/fmicb.2013.00106.

Klatt, C. G., Wood, J. M., Rusch, D. B., Bateson, M. M., Hamamura, N., Heidelberg, J. F., et al. (2011). Community ecology of hot spring cyanobacterial mats: Predominant populations and their functional potential. *ISME J.* 5, 1262–1278. doi:10.1038/ismej.2011.73.

Kochetkova, T. V, Toshchakov, S. V, Zayulina, K. S., Elcheninov, A. G., Zavarzina, D. G., Lavrushin, V. Y., et al. (2020). Hot in Cold : Microbial Life in the Hottest Springs in Permafrost. *Microorganisms* 8, 1–18. doi:doi:10.3390/microorganisms8091308.

Kolde, R. (2015). pheatmap : Pretty Heatmaps. *R Packag. version 1.0.8*, 1–7. Available at: https://cran.r-project.org/package=pheatmap.

Lauber, C. L., Hamady, M., Knight, R., and Fierer, N. (2009). Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale □ †. 75, 5111–5120. doi:10.1128/AEM.00335-09.

Lavrentyeva, E. V., Radnagurueva, A. A., Barkhutova, D. D., Belkova, N. L., Zaitseva, S. V., Namsaraev, Z. B., et al. (2018). Bacterial Diversity and Functional Activity of Microbial Communities in Hot Springs of the Baikal Rift Zone. *Microbiol. (Russian Fed.* 87, 272–281. doi:10.1134/S0026261718020078.

Leung, K., Zahn, H., Leaver, T., Konwar, K. M., Hanson, N. W., Pagé, A. P., et al. (2012). A programmable droplet-based microfluidic device applied to multiparameter analysis of single microbes and microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* 109, 7665–7670. doi:10.1073/pnas.1106752109.

Levasseur, A., Lomascolo, A., Chabrol, O., Ruiz-Dueñas, F. J., Boukhris-Uzan, E., Piumi, F., et al. (2014). The genome of the white-rot fungus Pycnoporus cinnabarinus: A basidiomycete model with a versatile arsenal for lignocellulosic biomass breakdown. *BMC Genomics* 15, 1–24. doi:10.1186/1471-2164-15-486.

Lewin, A., Zhou, J., Pham, V. T. T., Haugen, T., Zeiny, M. El, Aarstad, O., et al. (2017). Novel archaeal thermostable cellulases from an oil reservoir metagenome. *AMB Express* 7. doi:10.1186/s13568-017-0485-z.

Liu, L., Lv, A. P., Li, M. M., Ming, Y. Z., Jiao, J. Y., Fang, B. Z., et al. (2020). Seramator thermalis gen. Nov., sp. nov., a novel cellulose-and xylan-degrading member of the family dysgonamonadaceae isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* 70, 5717–5724. doi:10.1099/ijsem.0.004469.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, 490–495. doi:10.1093/nar/gkt1178.

López-Mondéjar, R., Tláskal, V., Větrovský, T., Štursová, M., Toscan, R., Nunes da Rocha, U., et al. (2020). Metagenomics and stable isotope probing reveal the complementary contribution of fungal and bacterial communities in the recycling of dead biomass in forest soil. *Soil Biol. Biochem.* 148. doi:10.1016/j.soilbio.2020.107875.

Lu, J., Breitwieser, F. P., Thielen, P., and Salzberg, S. L. (2017). Bracken: Estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* 2017, 1–17. doi:10.7717/peerj-cs.104.

Ma, L., Zhao, Y., Meng, L., Wang, X., Yi, Y., Shan, Y., et al. (2020). Isolation of Thermostable Lignocellulosic Bacteria From Chicken Manure Compost and a M42 Family Endocellulase Cloning From Geobacillus thermodenitrificans Y7. *Front. Microbiol.* 11, 1–15. doi:10.3389/fmicb.2020.00281.

Martins, L. F., Antunes, L. P., Pascon, R. C., de Oliveira, J. C. F., Digiampietri, L. A., Barbosa, D., et al. (2013). Metagenomic Analysis of a Tropical Composting Operation at the São Paulo Zoo Park Reveals Diversity of Biomass Degradation Functions and Organisms. *PLoS One* 8. doi:10.1371/journal.pone.0061928.

Marynowska, M., Goux, X., Sillam-Dussès, D., Rouland-Lefèvre, C., Halder, R., Wilmes, P., et al. (2020). Compositional and functional characterisation of biomass-degrading microbial communities in guts of plant fibre- and soil-feeding higher termites. *Microbiome* 8, 96. doi:10.1186/s40168-020-00872-3.

Mathrani, I. M., and Ahring, B. K. (1992). Applied Microbiology Biotechnology Thermophilic and alkalophilic xylanases from several Dictyoglomus isolates. 1, 23–27.

Menzel, P., Gudbergsd, R., Rike, G., Lin, L., Zhang, Q., Contursi, P., et al. (2015). Comparative Metagenomics of Eight Geographically Remote Terrestrial Hot Springs. *Environ. Microbiol.*, 411–424. doi:10.1007/s00248-015-0576-9.

Milshteyn, A., Schneider, J. S., and Brady, S. F. (2014). Mining the metabiome: Identifying novel natural products from microbial communities. *Chem. Biol.* 21, 1211–1223. doi:10.1016/j.chembiol.2014.08.006.

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Sundaramurthi, J. C., Lee, J., et al. (2021). Genomes OnLine Database (GOLD) v.8: overview and updates. *Nucleic Acids Res.* 49, D723–D733. doi:10.1093/nar/gkaa983.

Niederberger, T. D., Götz, D. K., McDonald, I. R., Ronimus, R. S., and Morgan, H. W. (2006). Ignisphaera aggregans gen. nov., sp. nov., a novel hyperthermophilic crenarchaeote isolated from hot springs in Rotorua and Tokaanu, New Zealand. *Int. J. Syst. Evol. Microbiol.* 56, 965–971. doi:10.1099/ijs.0.63899-0.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2019). vegan: Community Ecology Package. *R Packag. version 2.5-4*. Available at: https://cran.r-project/package=vegan.

Opdahl, L. J., Gonda, M. G., and St-Pierre, B. (2018). Identification of Uncultured Bacterial

Species from Firmicutes, Bacteroidetes and CANDIDATUS Saccharibacteria as Candidate Cellulose Utilizers from the Rumen of Beef Cows. *Microorganisms* 6, 17. doi:10.3390/microorganisms6010017.

Peacock, J. P., Cole, J. K., Murugapiran, S. K., Dodsworth, J. A., Fisher, J. C., Moser, D. P., et al. (2013). Pyrosequencing Reveals High-Temperature Cellulolytic Microbial Consortia in Great Boiling Spring after In Situ Lignocellulose Enrichment. *PLoS One* 8, 1–12. doi:10.1371/journal.pone.0059927.

Pedersen, M., Johansen, K. S., and Meyer, A. S. (2011). Low temperature lignocellulose pretreatment: Effects and interactions of pretreatment pH are critical for maximizing enzymatic monosaccharide yields from wheat straw. *Biotechnol. Biofuels* 4, 1–10. doi:10.1186/1754-6834-4-11.

Perevalova, A. A., Bidzhieva, S. K., Kublanov, I. V., Hinrichs, K. U., Liu, X. L., Mardanov, A. V., et al. (2010). Fervidicoccus fontis gen. nov., sp. nov., an anaerobic, thermophilic crenarchaeote from terrestrial hot springs, and proposal of Fervidicoccaceae fam. nov. and Fervidicoccales ord. nov. *Int. J. Syst. Evol. Microbiol.* 60, 2082–2088. doi:10.1099/ijs.0.019042-0.

Podar, P. T., Yang, Z., Björnsdóttir, S. H., and Podar, M. (2020). Comparative Analysis of Microbial Diversity Across Temperature Gradients in Hot Springs From Yellowstone and Iceland. *Front. Microbiol.* 11, 1–16. doi:10.3389/fmicb.2020.01625.

Power, J. F., Carere, C. R., Lee, C. K., Wakerley, G. L. J., Evans, D. W., Button, M., et al. (2018). Microbial biogeography of 925 geothermal springs in New Zealand. *Nat. Commun.* 9. doi:10.1038/s41467-018-05020-y.

R Core Team (2014). R: A language and environment for statistical computing. *R Found. Stat. Comput.* doi:10.18637/jss.v067.i01.R.

Reichart, N. J., Jay, Z. J., Krukenberg, V., Parker, A. E., Spietz, R. L., and Hatzenpichler, R. (2020). Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. *ISME J.* 14, 2851–2861. doi:10.1038/s41396-020-00749-1.

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., et al. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* 9, 1038–1048. doi:10.1038/nprot.2014.067.

Saiki, T., Kobayashi, Y., Kawagoe, K., and Beppu, T. (1985). Dictyoglomus thermophilum gen. nov., sp. nov., a Chemoorganotrophic, Anaerobic, Thermophilic Bacterium. *Int. J. Syst. Bacteriol.* 35, 253–259.

Sharp, C. E., Brady, A. L., Sharp, G. H., Grasby, S. E., Stott, M. B., and Dunfield, P. F. (2014). Humboldt's spa: Microbial diversity is controlled by temperature in geothermal environments. *ISME J.* 8, 1166–1174. doi:10.1038/ismej.2013.237.

Song, Z. Q., Wang, F. P., Zhi, X. Y., Chen, J. Q., Zhou, E. M., Liang, F., et al. (2013). Bacterial and archaeal diversities in Yunnan and Tibetan hot springs, China. *Environ. Microbiol.* 15, 1160–1175. doi:10.1111/1462-2920.12025.

Strazzulli, A., Cobucci-Ponzano, B., Iacono, R., Giglio, R., Maurelli, L., Curci, N., et al. (2020). Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments. *FEBS J.* 287, 1116–1137. doi:10.1111/febs.15080.

Suleiman, M., Krüger, A., and Antranikian, G. (2020). Biomass-degrading glycoside hydrolases of archaeal origin. *Biotechnol. Biofuels* 13, 1–14. doi:10.1186/s13068-020-01792-y.

Sun, Y., Liu, Y., Pan, J., Wang, F., and Li, M. (2020). Perspectives on Cultivation Strategies of Archaea. *Microb. Ecol.* 79, 770–784. doi:10.1007/s00248-019-01422-7.

Taillefer, M., Arntzen, M. Ø., Henrissat, B., Pope, P. B., and Larsbrink, J. (2018). Proteomic Dissection of the Cellulolytic Machineries Used by Soil-Dwelling Bacteroidetes . *mSystems* 3, 1–16. doi:10.1128/msystems.00240-18.

Takacs-Vesbach, C., Inskeep, W. P., Jay, Z. J., Herrgard, M. J., Rusch, D. B., Tringe, S. G., et al. (2013). Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct geothermal channels reveals specialization of three aquificales lineages. *Front. Microbiol.* 4, 1–25. doi:10.3389/fmicb.2013.00084.

Tange, O. (2018). *GNU Parallel 2018*.

Terekhov, S. S., Smirnov, I. V., Malakhova, M. V., Samoilov, A. E., Manolo, A. I., Nazarov, A. S., et al. (2018). Ultrahigh-throughput functional profiling of microbiota communities. *Proc. Natl. Acad. Sci. U. S. A.* 115, 9551–9556. doi:10.1073/pnas.1811250115.

Terekhov, S. S., Smirnov, I. V., Stepanova, A. V., Bobik, T. V., Mokrushina, Y. A., Ponomarenko, N. A., et al. (2017). Microfluidic droplet platform for ultrahigh-throughput single-cell screening of biodiversity. *Proc. Natl. Acad. Sci. U. S. A.* 114, 2550–2555. doi:10.1073/pnas.1621226114.

Toshchakov, S. V., Korzhenkov, A. A., Samarov, N. I., Mazunin, I. O., Mozhey, O. I., Shmyr, I. S., et al. (2015). Complete genome sequence of and proposal of Thermofilum uzonense sp. nov. A novel hyperthermophilic crenarchaeon and emended description of the genus Thermofilum. *Stand. Genomic Sci.* 10, 1–9. doi:10.1186/s40793-015-0105-y.

Van Der Helm, E., Genee, H. J., and Sommer, M. O. A. (2018). The evolving interface between synthetic biology and functional metagenomics. *Nat. Chem. Biol.* 14, 752–759. doi:10.1038/s41589-018-0100-x.

Wang, C., Dong, D., Wang, H., Müller, K., Qin, Y., Wang, H., et al. (2016). Metagenomic analysis of microbial consortia enriched from compost: New insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnol. Biofuels* 9, 1–17. doi:10.1186/s13068-016-0440-2.

Wang, G., Zhao, Z., Ke, J., Engel, Y., Shi, Y. M., Robinson, D., et al. (2019). CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. *Nat. Microbiol.* 4, 2498–2510. doi:10.1038/s41564-019-0573-8.

Wilhelm, R. C., Singh, R., Eltis, L. D., and Mohn, W. W. (2018). Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. *ISME J.*, 413–429. doi:10.1038/s41396-018-0279-6.

Wohlgemuth, R., Littlechild, J., Monti, D., Schnorr, K., Rossum, T. Van, Siebers, B., et al. (2018). Discovering novel hydrolases from hot environments. *Biotechnol. Adv.* 36, 2077–2100. doi:10.1016/j.biotechadv.2018.09.004.

Wong, M. T., Wang, W., Couturier, M., Razeq, F. M., Lombard, V., Lapebie, P., et al. (2017). Comparative metagenomics of cellulose- and poplar hydrolysate-degrading microcosms from gut microflora of the Canadian Beaver (Castor canadensis) and North American moose (Alces americanus) after long-term enrichment. *Front. Microbiol.* 8, 1–14. doi:10.3389/fmicb.2017.02504.

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 1–13. doi:10.1186/s13059-019-1891-0.

Yang, S. J., Kataeva, I., Hamilton-Brehm, S. D., Engle, N. L., Tschaplinski, T. J., Doeppke, C., et al. (2009). Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "Anaerocellum thermophilum" DSM 6725. *Appl. Environ. Microbiol.* 75, 4762–4769. doi:10.1128/AEM.00236-09.

Young, J., Chung, D., Bomble, Y. J., Himmel, M. E., and Westpheling, J. (2014). Deletion of *Caldicellulosiruptor bescii* CelA reveals its crucial role in the deconstruction of lignocellulosic biomass. *Biotechnol. Biofuels* 7, 142. doi:10.1186/s13068-014-0142-6.

Zarafeta, D., Kissas, D., Sayer, C., Gudbergsdottir, S. R., Ladoukakis, E., Isupov, M. N., et al. (2016). Discovery and characterization of a thermostable and highly halotolerant GH5 cellulase from an icelandic hot spring isolate. *PLoS One* 11, 1–18. doi:10.1371/journal.pone.0146454.

Zayulina, K. S., Kochetkova, T. V, Piunova, U. E., and Morgan, H. (2020). Novel Hyperthermophilic Crenarchaeon Thermofilum and Two Novel Glycosidases for Cellulose Hydrolysis. *Front. Microbiol.* 10, 1–12. doi:10.3389/fmicb.2019.02972.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi:10.1093/nar/gky418.

Zhao, C., Deng, Y., Wang, X., Li, Q., Huang, Y., and Liu, B. (2014). Identification and characterization of an anaerobic ethanol-producing cellulolytic bacterial consortium from great basin hot springs with agricultural residues and energy crops. *J. Microbiol. Biotechnol.* 24, 1280–1290. doi:10.4014/jmb.1401.01022.

CHAPTER FOUR

EFFECT OF CRYSTALLINE CELLULOSE SUPPLEMENTATION ON A HOT SPRING MICROBIAL COMMUNITY IN A SHORT TIME SCALE INCUBATION

Contribution of Authors and Co-Authors

Author: Nicholas J. Reichart

Contributions: Designed the study, performed the data analyses, and wrote the manuscript.

Author: Zackary J. Jay

Contributions: Provided bioinformatic support and helped edit the manuscript.

Author: Viola Krukenberg

Contributions: Designed the study, assisted with incubations, and edited the manuscript.

Author: Tanja Woyke

Contributions: Designed the study and edited the manuscript.

Author: Roland Hatzenpichler

Contributions: Designed the study and edited the manuscript.

Manuscript Information Page

Nicholas J. Reichart, Zackary J. Jay, Viola Krukenberg, Tanja Woyke, Roland Hatzenpichler

Status of Manuscript:

_X__ Prepared for submission to a peer-reviewed journal

____ Officially submitted to a peer-reviewed journal

____ Accepted by a peer-reviewed journal

____ Published in a peer-reviewed journal

CHAPTER FOUR

EFFECT OF CRYSTALLINE CELLULOSE SUPPLEMENTATION ON A HOT SPRING MICROBIAL COMMUNITY IN A SHORT TIME SCALE INCUBATION

Introduction

Lignocellulose is the most abundant biopolymer on Earth, comprising 35 - 40 % of agriculture-based cellulose components of this biomass (Rai et al., 2019). Cellulose degradation to monosaccharides and their subsequent fermentation are of high interest to researchers who are looking for more cost-effective methods of generating biofuels. The diversity of metabolisms of hot spring microbes, together with their enzymatic adaptations to extreme conditions, make these microbes strong targets for improvements in efficiency for plant biomass bioprocessing. Hot springs have previously been the target for data mining experiments (Reichart et al., 2021) for carbohydrate active enzymes (CAZymes) (Lombard et al., 2014) and several studies have been able to characterize novel hydrolases first discovered from hot springs in laboratory settings (Wohlgemuth et al., 2018; Strazzulli et al., 2020). Discovery of thermostable and pH tolerant enzymes will aid in future cellulose bioprocessing by allowing synergistic degradation of cellulose under extreme conditions with high activity enzymes. A cellulolytic organism, *Caldicellulosiruptor bescii* DSM 6725, has been tested to have a maximum growth temperature on cellulose at 90 °C and optimal activity at 75 °C (Yang et al., 2010). In addition to high temperature benefits for enzyme activity, the pH of industrial process is also influential for lignocellulose component degradation (He et al., 2018) and the wide pH range of hot springs

(Power et al., 2018) provides ample opportunity of discovery of cellulolytic organisms and novel CAZymes.

Development of single-cell studies have progressed to be able to study genomic predictions in high-throughput fashion for specific functions. Several studies have utilized bioorthogonal non-canonical amino acid tagging coupled with fluorescence-activated cell sorting (BONCAT-FACS) to study environmental microbial communities including deep sea sediment (Hatzenpichler et al., 2016), soil sediments (Couradeau et al., 2019), and even cystic fibrosis sputum (Valentini et al., 2020). While these studies were focused on the native, active community members, broad amendment screening with BONCAT-FACS has been used to measure changed in hot spring communities when exposed to a multitude of substrates (Reichart et al., 2020) and incubation of fecal matter with rutin to determine microbes involved in flavonol metabolism (Riva et al., 2020). Building upon these previous studies, BONCAT can be implemented to ascertain translational activity information of a desired function without the need for expensive compounds, such as isotopically labeled substrates.

The aim of this study is to demonstrate the use of bioorthogonal labeling for the specific function of cellulose degradation from a hot spring (77 °C, pH 7.6) microbial community. The selected hot spring contains several downed and submerged trees that have been present for an unknown amount of time (> 5 years) Samples were incubated with or without crystalline cellulose (Avicel) and subjected to BONCAT-FACS to determine microbes that were differentially active under the amendment condition. BONCAT-FACS enabled separation of active microbes from the overall bulk community and through subsequent amplicon sequencing, many of the active microbes were classified as uncultured taxa suggesting novel uncharacterized

populations were present in the sample and possibly involved in crystalline cellulose degradation. In addition to BONCAT-FACS functional data, a broad community analysis was performed using metagenomic sequencing of unamended hot spring sediment-slurry to bioinformatically screen for CAZymes that have the potential to be used in cellulose degradation. Together, using amplicon sequencing data from BONCAT-FACS and metagenomic based predictions, we connect the results to better understand the ecophysiology of a system seemingly rich in plant biomass. Our method can be applied to various other systems (e.g., compost or rumen microbiota) and other functions of interest (e.g., nitrification or chitin degradation).

Methods

Incubation setup

The sampling and experimental procedure for BONCAT-FACS amendment screening was developed in previous work (Reichart et al., 2020) and modified here to investigate for cellulose degradation. Sampling was performed in Yellowstone National Park, Wyoming, USA under US National Park Service permit no. YELL-SCI-8010 (2017-2019). A sediment slurry sample (~10 % sediment, 90 % water) was collected October 30th, 2019, using a sterile, stainless-steel measuring cup fixed to the end of a 12-foot telescoping pole from “Buffalo Pool” hot spring (77 °C, pH 7.6, 44.53246, -110.79748, Inventory ID LWCG154, https://rcn.montana.edu/Features/Detail.aspx?id=5620). The sample was transferred to a 1 L acid-washed and autoclaved sealed glass bottle with ambient air for a headspace and transported to the laboratory in an insulated heated container at an ambient temperature of 55 °C during the 4 h of transit. Incubation experiments were setup the following day after allowing the sample to adjust to 75 °C overnight in an incubator. Nine incubations were prepared that consisted of 30

mL of homogenized slurry in 70 mL acid-washed and autoclaved glass vials. Crystalline cellulose, Avicel (5 g/L, Fisher), was added to three vials, and all nine vials were incubated at 75 °C for 24 h. The following day 50 µM, *L*-homopropargylglycine (HPG, Click Chemistry Tools) was added to the three Avicel amended vials and three additional vials to serve as HPG-only controls. The remaining three vials served as negative controls (i.e., no-HPG control). All vials were incubated for an additional 12 h at 75 °C. At the end of the incubations, the content of each vial was split four ways (2x 5 mL, 2x 10 mL) for subsequent processing. The 5 mL aliquots were stored in 10 % glycerol-TE, while the 10 mL aliquots were directly flash frozen in liquid N_2. All aliquots were stored at -80 °C until further processing.

Sample processing and sorting

Sample processing was conducted over two consecutive days with cell extraction and BONCAT labeling occurring on the first day and cell sorting on the second day. Frozen samples were thawed at 4 °C, vortexed (Vortex Genie 2, Scientific Industries) for 10 seconds at maximum speed (setting 10), and 1 mL of sediment-water slurry was transferred to a 15 mL conical tube containing 4 mL of 1x phosphate buffered saline (PBS) with 0.01 % Tween 20 (Promega). The conical tubes containing the sample and Tween 20 mixture were vortexed at maximum speed for 5 minutes prior to being centrifuged at 500 *g* for 5 minutes to pellet the sediment particles. The supernatant was transferred evenly to three new 1.5 mL tubes after being passed through a 35 µm pore sized filter. The 1.5 mL tubes were centrifuged at 14,000 *g* for 5 minutes and the supernatant was discarded. Each pellet was resuspended and combined in a final volume of 300 µL 1x PBS. A stock of click reaction solution was prepared with the addition of SYBR I dye and 200 µL was aliquoted to each sample tube. The final reaction mix contained 5

mM amino guanidine hydrochloride (Sigma Aldrich), 5 mM sodium L-ascorbate (Sigma Aldrich), 100 µM copper sulfate pentahydrate (Sigma Aldrich), 500 µM THPTA (Click Chemistry Tools), 4 µM 594 picolyl-azide dye (Click Chemistry Tools), and 0.05 x SYBR I dye (Invitrogen) in 1x PBS. The samples were briefly vortexed to mix the solution and then rotated in the dark at room temperature for 1 hour. At the conclusion of the incubation time, samples were washed three times in 1 mL of 1x PBS and centrifugation at 14,000 *g* for 5 minutes. The final resuspension step was done in 500 µL of 1x PBS and stored at 4 °C in the dark overnight being sorted on a Sony SH800S fluorescence-activated cell sorter (FACS).

Prior to sorting, 50 µL of each sample was transferred to a new tube and stored at 4 °C to be used as a "Presort" sample. Four sorting gates were designed to exclude large particles based on forward and back scatter area and then to select and sort for all biomass-stained cells (SYBR samples) and translationally active cells (BONCAT samples). All gates for fluorescence signal were determined from first recording sorter event position of a no-HPG control sample compared to an HPG-only sample, with and without SYBR staining. The gates were designed to contain <0.1 % of events in samples that did not contain a fluorescence signal. Sorted cells were sorted directly into 1.5 mL tubes that contained 200 µL of 1x PBS that had previously been UV-sterilized and filtered through a pore size of 0.2 µm. All sorted samples were centrifuged at 14,000 *g* for 5 minutes and the supernatants discarded before resuspending the cell pellets in 20 µL of nuclease-free water to be used for amplicon sequencing.

Amplicon sequencing and processing

Prior to amplification of the 16S rRNA gene, samples stored in 20 µL nuclease-free water were transferred to a 96-well microtiter plate and subjected to three rounds of freeze-thaw cycles.

Briefly, the microtiter plates were frozen at -80 °C for 20 minutes and then transferred to a thermal cycler (Eppendorf Mastercycler Nexus GSX1) at 99 °C for 10 minutes. A pulse centrifugation step was performed in between each cycle. Amplification of the 16S rRNA gene was performed by PCR using the Earth Microbiome protocol (Thompson et al., 2017) with updated 515F (Parada et al., 2016) (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (Apprill et al., 2015) (5'-GGACTACNVGGGTWTCTAAT-3') primers in the microtiter plate containing the extracted DNA. The final PCR solution contained 20 µL Invitrogen Platinum Taq II 2x Master Mix, 1 µL 515F primer (10 µM; final 0.2 µM), 1 µL 806R primer (10 µM; final 0.2 µM), 8 µL nuclease-free water, and 20 µL of extracted DNA lysate. The thermocycler conditions were: 94 °C for 3 minutes followed by 28 cycles of 94 °C for 45 seconds, 50 °C for 60 seconds, and 72 °C for 90 seconds before a final elongation step at 72 °C for 10 minutes. PCR products were purified using Ampure XP beads (Beckman Coulter) following manufacturer's protocol and a final resuspension in 40 µL nuclease-free water. A subsequent PCR was performed to attach dual barcode indices and sequencing adapters to the purified products. The barcoding PCR contained 5 L purified, amplified DNA, 12.5 µL Invitrogen Platinum Taq II 2X Master Mix, 2.5 µL i5 primer (final concentration: 0.25 µM), 2.5 µL i7 primer (final concentration: 0.25 µM), and 2.5 µL of nuclease-free water. The conditions for the thermocycler were as follows: 95 °C for 3 minutes followed by 8 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds, and 72 °C for seconds, followed by a final elongation step at 72 °C for 5 minutes. Each PCR product was again purified with Ampure XP beads with a final resuspension in 40 µL of nuclease-free water. Quant-iT Picogreen ds DNA Assay (Invitrogen) reactions were performed in triplicate for all samples and quantified with a Biotek Synergy H1 Hybrid microplate reader. For each sample, 10

ng of DNA was pooled, and half of this volume was concentrated with QIAquick PCR purification spin column (Qiagen). Concentrated DNA was sequenced at Idaho State University's Molecular Research Core Facility using Illumina 2x250 paired end read MiSeq sequencing.

Sequencing reads were processed using QIIME 2 (Bolyen et al., 2019) (version 2020.11) with DADA2 plugin. Taxonomy was assigned using Silva SSU database release 138 with classify-sklearn method. Contaminant ASVs were removed using the prevalence model in Decontam (Davis et al., 2018) with a threshold of 0.5. Microbial community analysis was performed in R using phyloseq (McMurdie and Holmes, 2013) and vegan (Oksanen et al., 2019) packages.

Metagenomic sequencing and processing

A sediment slurry sample (~10 % sediment, 90 % water) was collected using a stainless-steel measuring cup fixed to the end of a 12-foot telescoping pole from "Buffalo Pool" hot spring (44.5325, -110.7974) into a 50 mL conical tube. This sample was collected October 31st, 2018, one year prior to the samples collected for incubation and amplicon sequencing. The sample was immediately flash frozen using a dry ice-ethanol bath. Back in the laboratory, the sample was thawed on ice and DNA was extracted from 2 g of sediment slurry using the FAST DNA spin kit for soil (MP Biomedicals) according to manufacturer's protocol. Extracted DNA was shipped to the Joint Genome Institute for metagenomic shotgun sequencing on the Illumina NovaSeq platform. The metagenome was submitted to Integrated Microbial Genome and Microbiomes (IMG/M) and assembly was performed by metaspades v 3.13.0 and annotation by IMG pipeline v5.0.1.

Assembled scaffolds ≥ 2000 bp were clustered ('binned') with six implementations of four different programs including, Maxbin v2.2.4 (Wu et al., 2016), Concoct v1.0.0 (Alneberg et al., 2014), Metabat v2.12.1 with and without coverage (Kang et al., 2015), and Autometa v1 using both bacterial and archaeal modes and the machine learning step (Miller et al., 2019). Initial working bins were generated with DAS_Tool (Sieber et al., 2018) by dereplication, aggregation, and scoring of the bins produced by each program. Preliminary taxonomies were assigned with GTDB-Tk v1.2.0 (Chaumeil et al., 2020) and GTDB release 89 (Parks et al., 2018). Bin quality statistics were determined with CheckM v1.1.2 (Parks et al., 2015). CAZymes of the metagenomic bins were identified using the dbCAN2 package (Zhang et al., 2018) and only gene hits that were present in at least two out of the tools (HMMER, HotPep, and Diamond) were considered for further analysis.

Results and discussion

In the present work, the metabolic response of a hot spring (77 °C, pH 7.6) microbial community was tested with or without crystalline cellulose amendment through BONCAT-FACS analysis. Cellulolytic microbial communities have been previously investigated using multiple methods and approaches such as sorting from fluorescent cellulose particles (Doud et al., 2019), stable isotope probing with ^{13}C cellulose (Wilhelm et al., 2018), and enrichment cultures on various plant biomass (Peacock et al., 2013). Despite the viability of these methods, they are hampered by either expensive reagents or long incubation time frames. Recent developments in the use of BONCAT-FACS have shown the utility of this method for screening multiple environments (Hatzenpichler et al., 2016; Couradeau et al., 2019; Riva et al., 2020; Valentini et al., 2020; Chen et al., 2021) and for a variety of potential metabolic substrates

(Reichart et al., 2020). We aimed to leverage the expansion of BONCAT-based studies to be able to efficiently and cost effectively screen an extreme environment for potential cellulolytic microbes.

Amplicon sequencing

Amplicon sequencing resulted in 1,044,875 reads across 34 samples, including sequencing controls. Following processing of the sequencing reads with Qiime2 (v 2020.11), the 34 samples comprised 944,868 reads representing 474 amplicon sequence variants (ASVs). Decontam prevalence model with a threshold of 0.5 was used and removed 25 ASVs classified as contaminants and listed in Supplementary Table 1. Processing of the samples after decontam removed seven samples that were used as processing controls (five FACS sheath fluid collections, cell extraction blank, and PCR negative) and three samples that were used to set the sorter gates (no-HPG, BONCAT sorted). After removal of these 10 samples, 77 singleton ASVs were removed, leaving 372 ASVs representing 830,397 reads. The minimum amount of reads in a sample was 8,425 and the maximum read was 51,548 for an average of 34,600 reads across 24 samples.

Diversity among substrates and sample type categories

The biodiversity among the different substrate and sample types varied widely. Classification of substrates were separated into three categories, Avicel with HPG (here on referred to as Avicel), HPG only, and no addition (here on referred to as control), and three sample type categories, presort, SYBR sorted, and BONCAT sorted. Species richness was measured by observed ASVs and diversity was measured by Shannon diversity index (Table 1). The highest average α-diversity among triplicate samples was control presort for both metrics

tested at 92.0 and 3.50 for observed and Shannon diversity, respectively. Avicel samples were highest in observed ASVs in BONCAT sorted samples (90.3) whereas the HPG only samples (87.0) were highest in presort samples. Shannon diversity ranged among all sample averages from a lowest value of 3.04 in Avicel SYBR sorted to the highest value of 3.50 in control presort.

Table 1. Alpha diversity mean values for three replicates.

	Presort			SYBR			BONCAT	
	Control	HPG	Avicel	Control	HPG	Avicel	HPG	Avicel
Observed taxa	92.0	87.0	79.3	84.3	82.0	79.3	72.7	90.3
Shannon diversity	3.50	3.42	3.32	3.29	3.16	3.04	3.23	3.26

BONCAT sorted samples are similar to SYBR sorted but have been gated and sorted through the FACS instrument according to positive labeling of click-attached fluorescent dye. As recently shown for a soil sediment microbiome (Couradeau et al., 2019), the active fraction of cells separated using BONCAT-FACS, was distinct from the presort bulk community. This was also evident in our dataset using hot spring sediment-water slurry.

Grouping samples according to substrate only, the control samples had the tightest range of α-diversity for both metrics, while the Avicel samples had the largest range of values (Supplementary Figure 1). The α-diversity of the samples ranged widely among individual treatment types and no significance was detected among the samples, presumably due to the standard deviation range of our triplicate samples.

Community composition through β-diversity and ASV relative abundance

While the difference for α-diversity metrics were not significant among sample types and substrates, the β-diversity throughout multiple samples was significant (Supplementary Table 2). The community composition was examined using Bray Curtis dissimilarity and plotted on a PCoA ordination plot (Figure 1). Permutational multivariate analysis of variance (PERMANOVA) was used to pairwise compare substrates and sample types. Overall, when all samples were compared by substrate (P = 0.050) or sample type (P = 0.001), there were significant differences in community composition (Supplementary Table 2). When the samples were further split by sample type, presort samples were significantly different from SYBR sorted (P = 0.008) and BONCAT sorted (P = 0.003) samples, as well as SYBR sorted being distinct from BONCAT sorted samples (P = 0.004). However, when Avicel was compared to HPG samples for all sample types, they were not significantly different (P = 0.241). Further pairwise tests were calculated between sample types of presort and BONCAT sorted samples separately for Avicel (P = 0.028) and HPG (P = 0.061). The composition was also compared between Avicel and HPG samples for both Presort (P = 0.183) and BONCAT (P = 0.100) sample types to determine the effect Avicel had on the translational activity. The full list of comparisons made, and resultant P values are summarized in Supplementary Table 2.

Interestingly, SYBR sorted samples for all three substrate conditions combined were distinct from presort samples according to β-diversity calculated with Bray Curtis dissimilarity (Figure 1, p = 0.008). It is possible that the low concentration used in SYBR to minimize fluorescent signal bleedover into BONCAT gates could have selected against ASVs present in the presort but not in the resultant SYBR sorted samples. Similarly, presort samples were also significantly different from the BONCAT sorted sample (p = 0.003) and SYBR sorted compared

to BONCAT sorted (p = 0.004). The difference in community composition observed could have been due to extended incubation times with the HPG or Avicel amendment, however, comparisons made among presort samples only were not significant, suggesting the overall community did not drastically shift during the time from of the experiment. Previous work on cellulolytic consortia from compost (Eichorst et al., 2013) and landfill (Ransom-Jones et al., 2017) enrichments have shown shifts in microbial taxa over incubation duration with both compost and landfill samples becoming less complex in α-diversity compared to starting inoculum. In the compost enrichment with Avicel, Firmicutes were observed as early degraders and overtime the Thermus and Bacteroidetes populations became abundant as late-stage polysaccharide degraders (Eichorst et al., 2013).

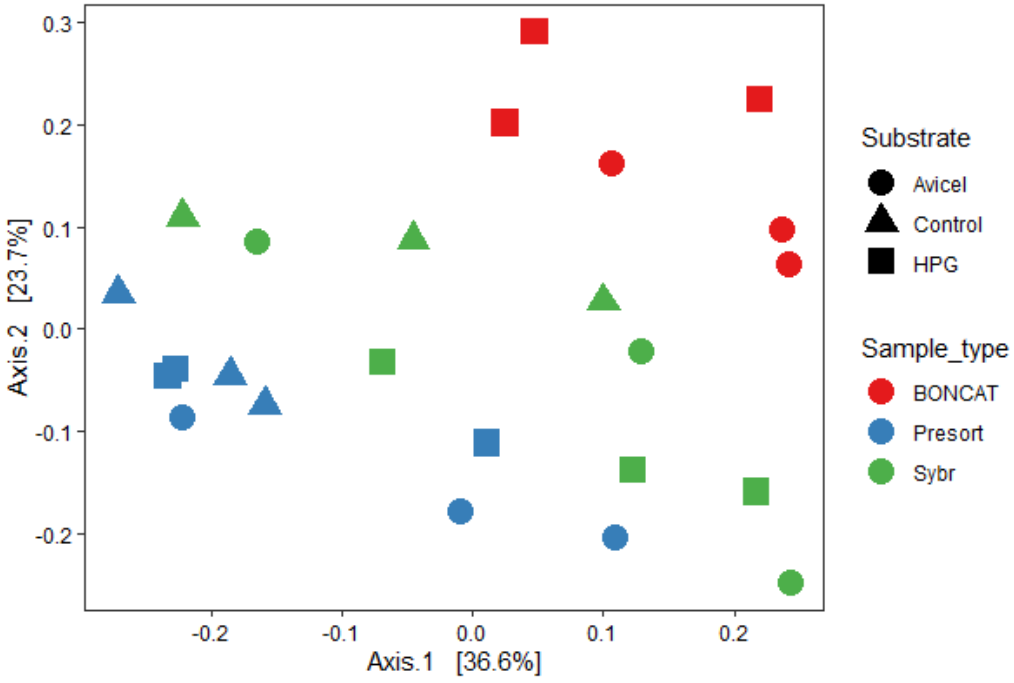


Figure 1. Principle coordinate analysis of Bray Curtis dissimilarity. Samples are shown as corresponding shapes to substrate and colored according to sample type categories.

The difference in community composition when accounting for all 376 ASVs was not significant when using PERMANOVA and Bray Curtis dissimilarity. We followed up this test with Kruskal-Wallis analysis to identify individual ASVs that were significantly different between BONCAT sorted Avicel and HPG only samples for each of the top 20 most abundant ASVs present in the dataset (Supplementary Table 3). Beyond the top 20 ASVs, the relative abundance of taxa across the dataset decreased to marginal amounts for comparison. Out of the top 20 most abundant ASVs across the entire dataset, seven ASVs were significantly different, suggesting the close to *in situ* activity of HPG only samples can be perturbed with Avicel supplementation ($P < 0.05$, Figure 2).

Several taxa from BONCAT sorted samples were significantly more abundant in HPG only samples. The relative abundances of the significant ASVs are shown in Figure 2 and ranged from 0.17 % to 11.6 % of the overall community depending on the phylum and sample. *Kapabacteriales* (previously classified within *Chlorobi* or *Ignavibacteria,* now classified within *Bacteroidota*) is yet uncultured; however, an additional organism from this lineage, *Melioribacter roseus,* was previously isolated using crystalline cellulose (Podosokorskaya et al., 2013). In both Avicel and HPG only samples, *Kapabacteriales* related ASV were more abundant in presort fractions opposed to the BONCAT sorted fractions, which could suggest a limiting substrate or being outcompeted in the incubation. In a survey of CAZyme from hot spring metagenomes, *Calescasmantes* were seen to encode oligosaccharide-degrading enzymes but not enzymes able to act early upon crystalline cellulose (Reichart et al., 2021). This taxon was significantly higher in the BONCAT sorted, HPG only samples within our short incubation time frame. Oligosaccharide degrading enzymes might require more time of degradation prior to

improving the activity of the resultant taxon, suggesting if we continued incubation samples, the ASVs related to *Calescasmantes* may have increased in relative abundance within the Avicel incubated samples. Two ASVs related to Chloroflexi were more abundant in HPG only samples, one of which could be classified to Thermus at the genus level. *Thermus hugenholtzii*, a previously cultured member of the *Chloroflexi*, was unable to grow with insoluble or soluble cellulose as a sole carbon source (Dodsworth et al., 2014). Alternatively, the three ASVs that were more abundant in BONCAT sorted fractions of Avicel than in HPG only samples belonged to one *Deinococcota Thermus* ASV and two ASVs of uncultured *Fervidibacteria*. Three *Deinococcota Thermus* ASVs were present within the top 20 ASVs, however, only one was significantly different, which could represent heterogeneity in the *Deinococcota* populations within Buffalo Pool. In contrast, three *Fervidibacteria* ASVs were present of which two were significant, and one of the significant ASVs could be classified to candidatus *Fervidibacteria* as opposed to uncultured. The cellulolytic nature of *Fervidibacteria* has remained an enigma with conflicting studies suggesting cellulolytic potential through genomics (Peacock et al., 2013) or previous BONCAT-FACS studies seeing *Fervidibacteria* decrease in the presence of cellulose amendment (Reichart et al., 2020). This phylum could be diverse in the metabolic capabilities of its members and requires further research to both obtain an isolate and to understand its cellulolytic nature.

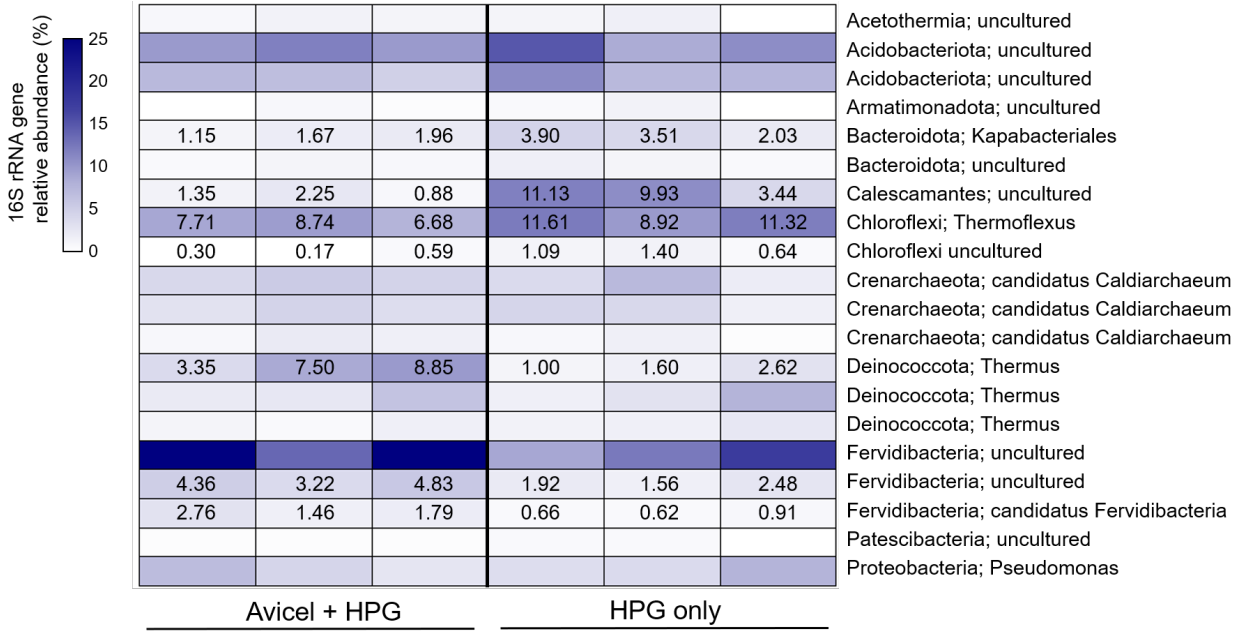


Figure 2. Heatmap of 16S rRNA gene relative abundance of the top 20 abundant amplicon sequence variants (ASVs) for triplicate BONCAT sorted samples incubated with or without Avicel. ASVs naming corresponds to Phylum followed by the lowest taxonomic rank provided by Silva database release version 138. Each row represents unique ASVs. Values in the plot represent relative abundance of significant differential abundance between Avicel and HPG only samples using Kruskal-Wallis test ($p < 0.05$).

Our analysis was confined by the taxonomic resolution of short read amplicon sequencing and comparatively limited 16S rRNA gene database from hot springs. Out of the top 20 ASVs across all samples, six ASVs could be classified to the genera level, and the remaining 14 ASVs were classified as uncultured genera (Supplementary Table 3).

Metagenomic sequencing and CAZyme potential

A Buffalo Pool DNA sample was sequenced at the Joint Genome Institute (JGI) and submitted to Integrated Microbial Genome and Microbiomes (IMG/M) under Taxonoid 3300034696. The assembled metagenomic size was 544 Mbp with an average GC content of

50.0 %. Of the 66 total bins recovered, 31 were considered medium or high-quality bins according to MiMAG standards (Bowers et al., 2017). All MAGs were subjected to CheckM and GTDB-tk analysis for bin statistics and phylogenetic identification. The average completeness and contamination of the medium and high-quality MAGs were 81.7 % and 1.92 %, respectively. They had an average length of 1.61 Mbp and 641-fold coverage. Bin statistics and identification was listed in Supplementary Table 4.

To further screen for cellulolytic potential in Buffalo Pool, CAZymes were detected using the dbCAN2 package on the 31 medium and high-quality MAGs. Only gene hits that were present in two or three of the dbCAN2 tools, were used for further analysis. A total of 16 CAZymes families were present with at least 10 gene hits across the entire 31 MAG dataset (Figure 3). For the 31 MAGs, there was a total of 406 CAZyme gene hits. Glycosyl transferase (GT) 4 and 2 were the most abundant CAZymes with glycosyl hydrlase (GH) 57 and carboxyl esterase (CE) 4 being the most abundant for their respective CAZyme class.

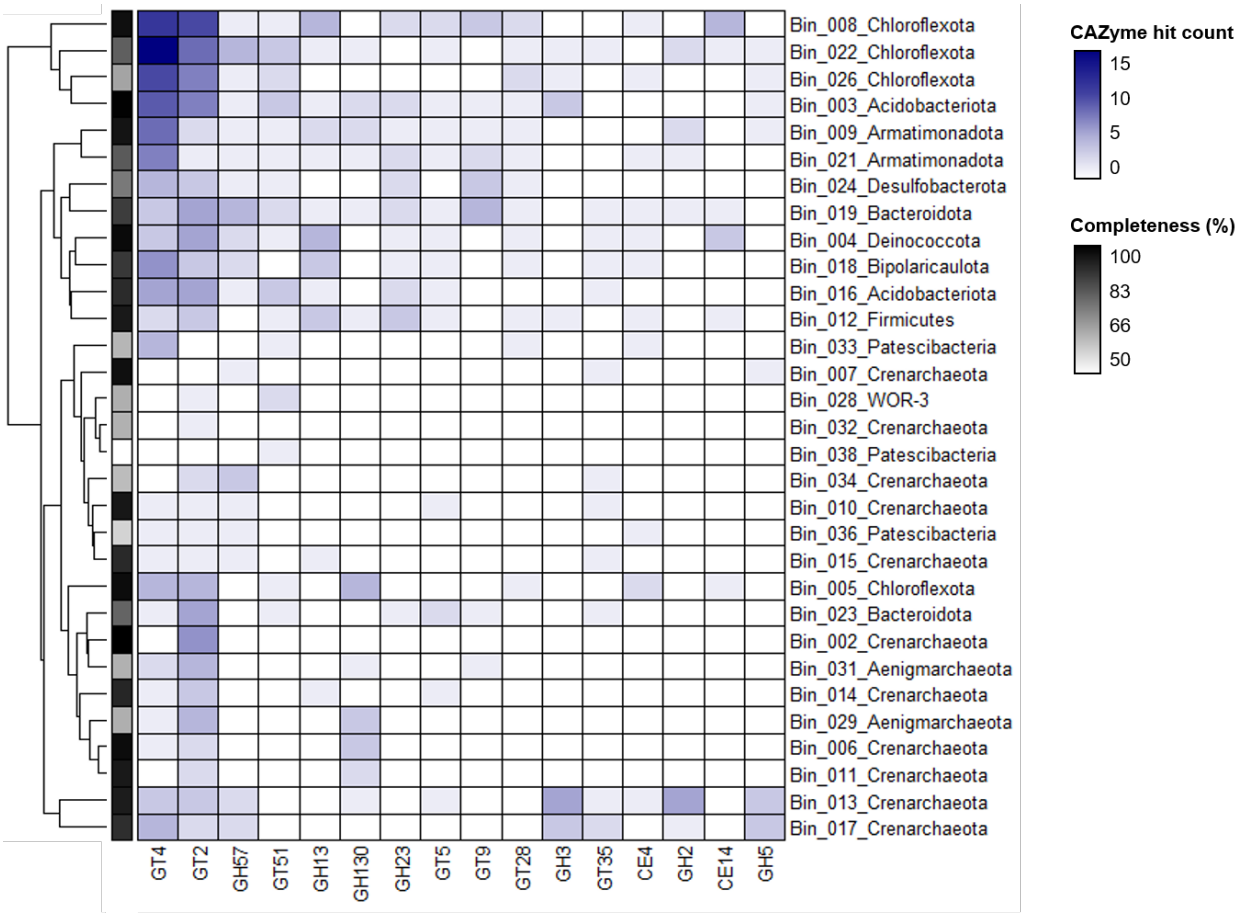


Figure 3**.** Heatmap of CAZyme domain hits in medium- and high-quality bins from Buffalo Pool metagenome bins. CAZymes shown had 10 or more hits across the entire dataset. Dendrogram based on clustering of total CAZymes detected with dbCAN2. Completeness (%) based on CheckM and taxonomy determined from GTDB-tk.

The total CAZyme hits spanned across 100 families and the most abundant CAZymes included GH 3 and 5 which are associated with cellulose degradation. Interestingly, two MAGs of Crenarchaeotal lineage, contained the most gene hits within these CAZyme families. MAG bins 013 and 017 belong to *Thermofilales* and *Ignisphaera*, respectively. A member of *Ignisphaera* was determined to be able to degrade crystalline cellulose as part of an enrichment culture from a geothermal source at 90 °C (Graham et al., 2011). *Ignisphaera* has also been enriched significantly on plant biomass from an 85 °C hot spring (Peacock et al., 2013). Following previous enrichment methods observing strong cellulolytic activity from *Ignisphaera*, Buffalo Pool's *Ignisphaera* population would be of interest for culturing efforts to expand upon this taxa's cellulose degradation capabilities.

In conclusion, this work confirmed that hot springs remain a promising environment for discovering novel cellulolytic organisms and BONCAT-FACS is a viable approach for detecting such microbes as was seen with the different ASV response to incubation amendment with or without crystalline cellulose. The results indicate microbial species respond with metabolic activity to substrate amendment and BONCAT-FACS can separate fluorescently labeled populations from environmental samples. The application of BONCAT-FACS to the discovery of biotechnologically relevant functions such as cellulose degradation, opens many new avenues of research such as cell sorting active microbes to directly culture cells of interest or single-cell genomic sequencing of cells directly involved in functions of interest.

Author contributions

NJR, TW, and RH conceptualized and designed the study. NJR and VK performed the experiments. ZJJ performed metagenomic binning. NJR analyzed the data and drafted the manuscript. All authors edited the manuscript.

Funding

This work was supported through a NASA Exobiology award (80NSSC19K1633) to RH and a DOE Office of Science Graduate Student Research award to NJR. A portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative (award 503546 to RH) and used resources at the Department of Energy (DOE) Joint Genome Institute (JGI), the National Energy Research Scientific Computing Center (NERSC), and the Environmental Molecular Sciences Laboratory (EMSL), which are DOE Office of Science User Facilities. These facilities were sponsored by the Office of Biological and Environmental Research and operated under Contract Nos. DE-AC02-05CH11231 (JGI and NERSC) and DE-AC05-76RL01830 (EMSL). This material is based upon work supported by the U.S. DOE, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education (ORISE) for the DOE. ORISE is managed by ORAU under Contract No. DE-SC0014664.

Acknowledgements

We appreciate the US National Park Service, in particular Annie Carlson at the Yellowstone Center for Resources, for permitting work in YNP under Permit No. YELL-SCI-8010.

REFERENCES CITED

Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi:10.1038/nmeth.3103.

Apprill, A., Mcnally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* 75, 129–137. doi:10.3354/ame01753.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi:10.1038/s41587-019-0209-9.

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35, 725–731. doi:10.1038/nbt.3893.

Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi:10.1093/bioinformatics/btz848.

Chen, L., Zhao, B., Li, X., Cheng, Z., Wu, R., and Xia, Y. (2021). Isolating and characterizing translationally active fraction of anammox microbiota using bioorthogonal non-canonical amino acid tagging. *Chem. Eng. J.* 418, 129411. doi:10.1016/j.cej.2021.129411.

Couradeau, E., Sasse, J., Goudeau, D., Nath, N., Hazen, T. C., Bowen, B. P., et al. (2019). Probing the active fraction of soil microbiomes using BONCAT-FACS. *Nat. Commun.* 10. doi:10.1038/s41467-019-10542-0.

Davis, N. M., Proctor, Di. M., Holmes, S. P., Relman, D. A., and Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 6, 1–14. doi:10.1186/s40168-018-0605-2.

Dodsworth, J. A., Gevorkian, J., Despujos, F., Cole, J. K., Murugapiran, S. K., Ming, H., et al. (2014). Thermoflexus hugenholtzii gen. nov., sp. nov., a thermophilic, microaerophilic, filamentous bacterium representing a novel class in the Chloroflexi, Thermoflexia classis nov., and description of Thermoflexaceae fam. nov. and Thermoflexales ord. nov. *Int. J. Syst. Evol. Microbiol.* 64, 2119–2127. doi:10.1099/ijs.0.055855-0.

Doud, D. F. R., Bowers, R. M., Schulz, F., Raad, M. De, Deng, K., Tarver, A., et al. (2019). Function-driven single-cell genomics uncovers cellulose-degrading bacteria from the rare biosphere. *ISME J.* doi:10.1038/s41396-019-0557-y.

Eichorst, S. A., Varanasi, P., Stavila, V., Zemla, M., Auer, M., Singh, S., et al. (2013). Community dynamics of cellulose-adapted thermophilic bacterial consortia. 15, 2573–2587. doi:10.1111/1462-2920.12159.

Graham, J. E., Clark, M. E., Nadler, D. C., Huffer, S., Chokhawala, H. A., Rowland, S. E., et al. (2011). Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat. Commun.* 2, 375–379. doi:10.1038/ncomms1373.

Hatzenpichler, R., Connon, S. A., Goudeau, D., Malmstrom, R. R., Woyke, T., and Orphan, V. J. (2016). Visualizing in situ translational activity for identifying and sorting slow-growing

archaeal−bacterial consortia. *Proc. Natl. Acad. Sci.* 113, E4069–E4078. doi:10.1073/pnas.1603757113.

He, Y.-C., Ma, C.-L., and Yang, B. (2018). "Pretreatment Process and Its Synergistic Effects on Enzymatic Digestion of Lignocellulosic Material," in *Fungal Cellulolytic Enzymes: Microbial Production and Application*, 1–25. doi:10.1007/978-981-13-0749-2.

Kang, D. D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 2015, 1–15. doi:10.7717/peerj.1165.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, 490–495. doi:10.1093/nar/gkt1178.

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8. doi:10.1371/journal.pone.0061217.

Miller, I. J., Rees, E. R., Ross, J., Miller, I., Baxa, J., Lopera, J., et al. (2019). Autometa: Automated extraction of microbial genomes from individual shotgun metagenomes. *Nucleic Acids Res.* 47, 1–12. doi:10.1093/nar/gkz148.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2019). vegan: Community Ecology Package. *R Packag. version 2.5-4*. Available at: https://cran.r-project/package=vegan.

Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. doi:10.1111/1462-2920.13023.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996. doi:10.1038/nbt.4229.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi:10.1101/gr.186072.114.

Peacock, J. P., Cole, J. K., Murugapiran, S. K., Dodsworth, J. A., Fisher, J. C., Moser, D. P., et al. (2013). Pyrosequencing Reveals High-Temperature Cellulolytic Microbial Consortia in Great Boiling Spring after In Situ Lignocellulose Enrichment. *PLoS One* 8, 1–12. doi:10.1371/journal.pone.0059927.

Podosokorskaya, O. A., Kadnikov, V. V., Gavrilov, S. N., Mardanov, A. V., Merkel, A. Y., Karnachuk, O. V., et al. (2013). Characterization of Melioribacter roseus gen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class Ignavibacteria, and a proposal of a novel bacterial phylum Ignavibacteriae. *Environ. Microbiol.* 15, 1759–1771. doi:10.1111/1462-2920.12067.

Power, J. F., Carere, C. R., Lee, C. K., Wakerley, G. L. J., Evans, D. W., Button, M., et al. (2018). Microbial biogeography of 925 geothermal springs in New Zealand. *Nat. Commun.* 9. doi:10.1038/s41467-018-05020-y.

Rai, R., Agrawal, D., and Chadha, B. S. (2019). *New paradigm in degradation of lignocellulosic biomass and discovery of novel microbial strains*. doi:10.1007/978-981-13-8487-5_16.

Ransom-Jones, E., McCarthy, A. J., Haldenby, S., Doonan, J., and McDonald, J. E. (2017). Lignocellulose-degrading microbial communities in landfill sites represent a repository of unexplored biomass-degrading diversity. *mSphere* 2, 1–13.

Reichart, N. J., Bowers, R. M., Woyke, T., and Hatzenpichler, R. (2021). High Potential for Biomass-Degrading Enzymes Revealed by Hot Spring Metagenomics. *Front. Microbiol.* 12, 1–13. doi:10.3389/fmicb.2021.668238.

Reichart, N. J., Jay, Z. J., Krukenberg, V., Parker, A. E., Spietz, R. L., and Hatzenpichler, R. (2020). Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. *ISME J.* 14, 2851–2861. doi:10.1038/s41396-020-00749-1.

Riva, A., Kolimár, D., Spittler, A., Wisgrill, L., Herbold, C. W., Abrankó, L., et al. (2020). Conversion of Rutin, a Prevalent Dietary Flavonol, by the Human Gut Microbiota. *Front. Microbiol.* 11, 1–11. doi:10.3389/fmicb.2020.585428.

Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., et al. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* 3, 836–843. doi:10.1038/s41564-018-0171-1.

Strazzulli, A., Cobucci-Ponzano, B., Iacono, R., Giglio, R., Maurelli, L., Curci, N., et al. (2020). Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments. *FEBS J.* 287, 1116–1137. doi:10.1111/febs.15080.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. doi:10.1038/nature24621.

Valentini, T. D., Lucas, S. K., Binder, K. A., Cameron, L. C., Motl, J. A., Dunitz, J. M., et al. (2020). Bioorthogonal non-canonical amino acid tagging reveals translationally active subpopulations of the cystic fibrosis lung microbiota. *Nat. Commun.* 18, S36. doi:10.1016/s1569-1993(19)30233-4.

Wilhelm, R. C., Singh, R., Eltis, L. D., and Mohn, W. W. (2018). Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. *ISME J.*, 413–429. doi:10.1038/s41396-018-0279-6.

Wohlgemuth, R., Littlechild, J., Monti, D., Schnorr, K., Rossum, T. van, Siebers, B., et al. (2018). Discovering novel hydrolases from hot environments. *Biotechnol. Adv.* 36, 2077–2100. doi:10.1016/j.biotechadv.2018.09.004.

Wu, Y. W., Simmons, B. A., and Singer, S. W. (2016). MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32, 605–607. doi:10.1093/bioinformatics/btv638.

Yang, S. J., Kataeva, I., Wiegel, J., Yin, Y., Dam, P., Xu, Y., et al. (2010). Classification of "Anaerocellum thermophilum" strain DSM 6725 as Caldicellulosiruptor bescii sp. nov. *Int. J. Syst. Evol. Microbiol.* 60, 2011–2015. doi:10.1099/ijs.0.017731-0.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi:10.1093/nar/gky418.

CHAPTER FIVE

METAGENOMIC ASSEMBLIES OF SUBSTRATE-AMENDED HOT SPRING SEDIMENT INCUBATIONS FROM YELLOWSTONE NATIONAL PARK

Contribution of Authors and Co-Authors

Author: Nicholas J. Reichart

Contributions: Designed the study, performed the data analyses, and wrote the manuscript.

Author: Robert M. Bowers

Contributions: Provided bioinformatic support and helped edit the manuscript.

Author: Tanja Woyke

Contributions: Designed the study and edited the manuscript.

Author: Roland Hatzenpichler

Contributions: Designed the study and wrote the manuscript.

Manuscript Information Page

Nicholas J. Reichart, Robert M. Bowers, Tanja Woyke, Roland Hatzenpichler

Frontiers in Microbiology

Status of Manuscript:

_X__ Prepared for submission to a peer-reviewed journal

____ Officially submitted to a peer-reviewed journal

____ Accepted by a peer-reviewed journal

____ Published in a peer-reviewed journal

CHAPTER FIVE

METAGENOMIC ASSEMBLIES OF SUBSTRATE-AMENDED HOT SPRING SEDIMENT INCUBATIONS FROM YELLOWSTONE NATIONAL PARK

Abstract

Here, we examined eight sediment metagenome data sets obtained from an alkaline hot spring in Yellowstone National Park, WY, USA. Samples were incubated for two days with various substrate amendments in conjunction with homopropargylglycine in an activity study targeted at identifying anabolically active uncultured bacteria and archaea. Metagenome-assembled genomes were constructed and the sequencing statistics were compared between samples.

Introduction

Hot spring ecosystems harbor diverse microbial lifeforms, many of which currently do not have a cultured representative (1, 2). The complex metabolic networks within the environment make it difficult to determine conditions conducive for culturing. However, incubations of hot springs material with bioorthogonal compounds can be used in conjunction with fluorescent labeling to identify translationally active microbial members within a community (3–5). Connected with a study (6) to describe microbial activity via 16S rRNA gene amplicon sequencing of translationally active cells, incubated hot spring material was subjected to metagenomic sequencing to accompany previous 16S rRNA gene amplicon data. The dataset

generated herein aids to describe the potential genomic functionally and physiology observed in this previous study (6).

A sediment-water microbiome from a hot spring within the Five Sisters hot spring group in Yellowstone National Park (44.532619, −110.797272; 74 °C, pH 8.4) was amended with potential growth substrates and the bioorthogonal amino acid homopropargylglycine (HPG) and incubated at 74 °C for 48 hours in the dark. We previously identified samples that had significant shifts in active microbial populations using fluorescence-activated cell sorting (FACS; (6)). At the time of incubation, left-over material that had not been processed for FACS had been flash frozen for metagenomic processing. Seven bulk, unsorted samples were selected for sequencing based on significant shifts in their active microbial populations determined via FACS (6). In addition, one unamended, unincubated sample was sequenced (Table 1). Genomic DNA was extracted using MP Biomedical FAST DNA Spin Kit for Soil according to the manufacturer's directions.

Metagenome samples were added to the Joint Genome Institute's (JGI) Integrated Microbial Genomes and Microbiomes (IMG/M) database (7) between May 2020 and November 2020 and as such, were processed with different versions of IMG/M's processing pipelines. Sequences were assembled using metaspades v3.13.0 or v3.14.1 and genes were annotated using IMG/M pipeline v5.0.14 or v5.0.19. After assembly and annotation, metagenomes were binned according to IMG/M's methods which used MetaBat v2.12.1 or v2.2.15. MAG statistics were generated using CheckM v1.0.12 or v1.1.3. Only MAGs that are classified as medium or high quality according to MIMAG standards (8) are represented. Finally, MAGs were phylogenetically classified using GTDB-tk v0.2.2 or v1.3.0 which used GTDB database release

86 and 95, respectively. Metagenome processing versions and metagenome statistics are available in Supplementary Table 1. All IMG/M bin ID numbers, statistics, and processing method for MAGs can be found in Supplementary Table 2. The MAGs represented 12 phyla and 18 classes.

Table 1. Incubation information for metagenomic samples.

Taxon OID	Sample Name	Substrate	Headspace	HPG	Incubation time
3300038414	FS5_t0	None	atmospheric	No	-
3300040930	FS5_Glycine_1	Glycine	atmospheric	Yes	48 hours
3300040931	FS5_Glycine_2	Glycine	atmospheric	Yes	48 hours
3300040932	FS5_Glycine_3	Glycine	atmospheric	Yes	48 hours
3300040933	FS5_Cellobiose_2	Cellobiose	atmospheric	Yes	48 hours
3300040934	FS_N2_1	None	100% N_2	Yes	48 hours
3300040935	FS_N2_2	None	100% N_2	Yes	48 hours
3300041472	FS5_HPG_2	None	atmospheric	Yes	48 hours

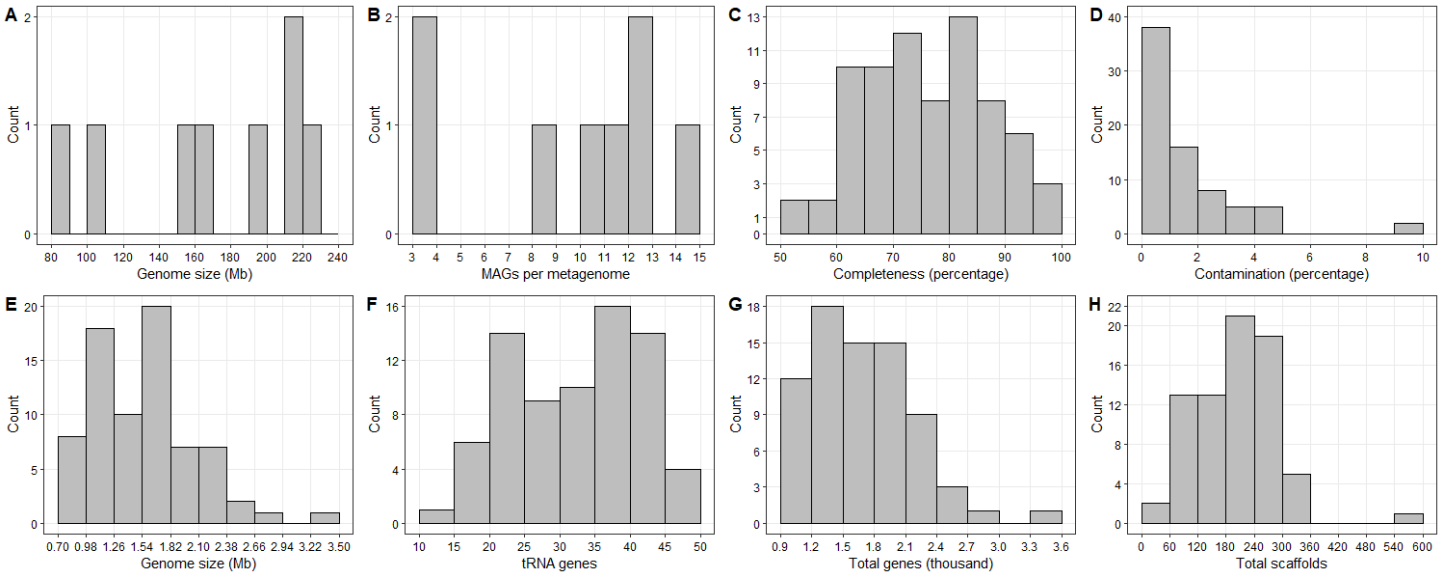


Figure 1. Sequencing statistics for (A-B) assembled metagenomes and (C-H) medium and high-quality MAGs generated by IMG/M pipeline.

Data availability

The assembled metagenomes are deposited on IMG/M according to the Taxon Object ID listed in Table 1 and described in further detail in Supplementary Table 1. Details of the MAGs are in Supplementary Table 2 with corresponding IMG/M Bin ID.

Acknowledgements

This work was supported through a NSF RII Track-2 FEC award (DBI-1736255) and a NASA Early Career Fellowship to R.H. (80NSSC19K0449). A portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative (award 503546 to R.H.) and used resources at the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory, which are DOE Office of Science User Facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under Contract Nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL). This material is based upon work supported by the U.S. DOE, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education (ORISE) for the DOE. ORISE is managed by ORAU under contract number DE-SC0014664. All opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of DOE, ORAU, or ORISE. We appreciate the US National Park Service, in particular Annie Carlson at the Yellowstone Center for Resources, for permitting work in YNP under permit number YELL-SCI-8010.

REFERENCES CITED

1. Eloe-Fadrosh EA, Paez-Espino D, Jarett J, Dunfield PF, Hedlund BP, Dekas AE, Grasby SE, Brady AL, Dong H, Briggs BR, Li WJ, Goudeau D, Malmstrom R, Pati A, Pett-Ridge J, Rubin EM, Woyke T, Kyrpides NC, Ivanova NN. 2016. Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. Nat Commun 7:1–10.
2. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. 2018. Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. mSystems 3:1–12.
3. Hatzenpichler R, Scheller S, Tavormina PL, Babin BM, Tirrell DA, Orphan VJ. 2014. In situ visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. Environ Microbiol 16:2568–2590.
4. Hatzenpichler R, Connon SA, Goudeau D, Malmstrom RR, Woyke T, Orphan VJ. 2016. Visualizing in situ translational activity for identifying and sorting slow-growing archaeal−bacterial consortia. Proc Natl Acad Sci 113:E4069–E4078.
5. Hatzenpichler R, Krukenberg V, Spietz RL, Jay ZJ. 2020. Next-generation physiology approaches to study microbiome function at single cell level. Nat Rev Microbiol 18.
6. Reichart NJ, Jay ZJ, Krukenberg V, Parker AE, Spietz RL, Hatzenpichler R. 2020. Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. ISME J 14:2851–2861.
7. Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter S, Varghese N, Seshadri R, Roux S, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC. 2021. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. Nucleic Acids Res 49:D751–D763.
8. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731.

CHAPTER SIX

CONCLUSION AND OUTLOOK FOR BONCAT EXPERIMENTS

This section will briefly recap the work developed in this thesis on BONCAT experiments and expand upon the difficulties and troubleshooting encountered. I hope to assist any future researchers who may wish to implement BONCAT methodology into their research plans. BONCAT was selected as an appropriate approach due to its relatively affordable reagents, high-throughput screening, and availability of necessary instrumentation. As with all methodology, trade-offs were present when selecting this approach. The number of incubations performed varied depending on experiment and reached a maximum of 75 samples, without controls included. Using other methods, such as Raman cell sorting or nanoSIMS, this would have been restricted due to processing time and costs. With proper benchmarking of BONCAT and BONCAT-FACS, the speed and affordability of processing outweighed any slight lack of precision.

When establishing a BONCAT-FACS experiment several steps need to be developed for each sample type; incubation time, cell extraction protocol, and sorting conditions are just a few. BONCAT is used to screen a sample for translational activity through incorporation of a synthetic amino acid which can later be fluorescently identified by “clicking” a dye to the newly synthesized proteins. Incubation time of the experiment needs to be sufficient to achieve labeling that is detectable by either epifluorescence microscopy for cell counts or FACS instrumentation for sorting. In contrast to running an experiment long enough to get usable fluorescence, too long of an incubation introduces the chance that most cells could become labeled. With too much of the sample being BONCAT labeled, differential analysis of treatment to control groups would

not be possible. A time series is recommended for each sample type or environment as a preliminary test to determine the earliest timepoint of sufficient BONCAT labeling to be used in final experiments.

In addition to the length of BONCAT incubation, the time point in which the synthetic amino acid is added may influence results. Tests are ongoing but allowing for a "preincubation" period for complex substrate may allow for more specific results to the substrate to be analyzed. In development experiments, complex substrates such as crystalline cellulose or native plant biomass were added along with HPG at time point 0. It is difficult with just BONCAT-FACS to determine if translational activity at this point was due to the substrate amendment or the previous conditions of the sample. As previously discussed, lignocellulose degradation requires a multitude of enzymes that may not be constitutively expressed. Allowing for a certain span of time to occur while in the presence of the substrate and then subsequent addition of HPG to start the incubation timeframe, may better survey the translational activity in response to the substrate and not the original sample conditions. Like the incubation time series, a time series would also be required at this point to find a preincubation time. Fluorescence checks or cell counts may not be adequate to successfully determine if a cell is beginning to consume a complex substrate. Methods to track the presence of the substrate would be informative to indicate when a sample is beginning to process the amendment. Together, the preincubation of a substrate prior to HPG addition and the time length of incubation are the two major first steps required to properly establishing a BONCAT experiment.

Once the length of incubation is decided on for a BONCAT experiment, substrate conditions must be determined. In previous work, the geochemistry of the studied hot springs

was not analyzed prior. Hot spring selection for cellulose degradation was determined based on the visual presence of plant biomass, present for an unknown amount of time. Along with the tests to determine incubation time based on substrate consumption, tests to check natural concentration could benefit the understanding of effects of amendment concentration. Too high of a concentration of amendments may have a detrimental effect on the microbial population and too little may not elicit a response. Hot springs geochemistry can vary depending on seasonal conditions (e.g., snow melt and rain runoff) and the residence time of features with runoffs can impact conditions. I would recommend future BONCAT experiments to survey the geochemistry of the sample site prior to starting incubations to better understand the native conditions. In addition to determining the substrate concentrations, the number of replicates should be considered for the sample provided. Permits within Yellowstone National Park limited the amount of material that could be retrieved for each sample trip. However, I would recommend the importance of replicates over additional sample conditions. Triplicates were used in all the BONCAT-FACS experiments conducted in this thesis, but future experiments will focus on more replicates to provide better statistical support.

Processing of BONCAT-FACS samples will vary depending on sample type and incubation time. Multiple cell extraction methods should be tested to improve upon cell to sediment ratios for reduced particle noise once on the FACS instrument while maintaining high cell yields. Dye selection will depend on the FACS instrument's lasers and will influence the experimental setup for both possible usage of a nucleic acid counterstain and concentration of dyes. Fluorescence detection should be optimized while maintaining low levels of fluorescence channel bleedover. Purity of the sorts should be checked to confirm confidence in sorting gates.

A sorted sample, placed back onto the FACS instrument plotting entirely back in the sorted gate confirms successful gating and sorting efficiency. In addition to this check, an epifluorescence microscopy check of the sample should be done, considering the FACS instrument and the microscope may have different detection levels for fluorescence. Once proper sorting is confirmed, the number of events sorted will need to be tested with downstream processing steps. Sorting for amplicon sequencing or metagenomics may require different amounts of cells to recover sufficient DNA material. A test sample sorted at increasing event numbers and carried through cell lysis and DNA recovery, will provide a working range of events for subsequent experiments of similar sample type. BONCAT-FACS experiments can provide vital information of translational activity for a microbial community, but much is needed for proper optimization. Following the steps and recommendation laid out here will help future researchers identify trouble areas with BONCAT and hopefully provide ease of planning and conducting BONCAT-FACS experimentations.

CUMULATIVE REFERENCES

Abby, S. S., Melcher, M., Kerou, M., Krupovic, M., Stieglmeier, M., Rossel, C., et al. (2018). Candidatus Nitrosocaldus cavascurensis , an Ammonia Oxidizing , Extremely Thermophilic Archaeon with a Highly Mobile Genome. *Front. Microbiol.* 9, 1–19. doi:10.3389/fmicb.2018.00028.

Agbor, V. B., Cicek, N., Sparling, R., Berlin, A., and Levin, D. B. (2011). Biomass pretreatment : Fundamentals toward application. 29, 675–685. doi:10.1016/j.biotechadv.2011.05.005.

Allaire, J. J., Ellis, P., Gandrud, C., Kuo, K., Lewis, B. W., Owen, J., et al. (2017). networkD3: JavaScript Network Graphs from R. Available at: https://cran.r-project.org/package=networkD3.

Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi:10.1038/nmeth.3103.

Alteio, L. V., Schulz, F., Seshadri, R., Varghese, N., Rodriguez-Reillo, W., Ryan, E., et al. (2020). Complementary Metagenomic Approaches Improve Reconstruction of Microbial Diversity in a Forest Soil. *mSystems* 5, 1–18. doi:10.1128/msystems.00768-19.

Anderson, I., Abt, B., Lykidis, A., Klenk, H. P., Kyrpides, N., and Ivanova, N. (2012). Genomics of aerobic cellulose utilization systems in actinobacteria. *PLoS One* 7. doi:10.1371/journal.pone.0039331.

Apprill, A., Mcnally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* 75, 129–137. doi:10.3354/ame01753.

Ausland, C., Zheng, J., Yi, H., Yang, B., Li, T., Feng, X., et al. (2021). dbCAN-PUL: A database of experimentally characterized CAZyme gene clusters and their substrates. *Nucleic Acids Res.* 49, D523–D528. doi:10.1093/nar/gkaa742.

Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., and Lloyd, K. G. (2020). Diversity, ecology and evolution of Archaea. *Nat. Microbiol.* 5, 887–900. doi:10.1038/s41564-020-0715-z.

Bala, A., and Singh, B. (2019). Development of an environmental-benign process for efficient pretreatment and saccharification of Saccharum biomasses for bioethanol production. *Renew. Energy* 130, 12–24. doi:10.1016/j.renene.2018.06.033.

Bang-Andreasen, T., Anwar, M. Z., Lanzén, A., Kjøller, R., Rønn, R., Ekelund, F., et al. (2020). Total RNA sequencing reveals multilevel microbial community changes and functional responses to wood ash application in agricultural and forest soil. *FEMS Microbiol. Ecol.* 96, 1–13. doi:10.1093/femsec/fiaa016.

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67. doi:10.18637/jss.v067.i01.

Bayer, E. A., Shoham, Y., and Lamed, R. (2013). "Lignocellulose-Decomposing Bacteria and Their Enzyme Systems," in *The Prokaryotes: Prokaryotic Physiology and Biochemistry*, 1–662. doi:10.1007/978-3-642-30141-4.

Beam, J. P., Jay, Z. J., Schmid, M. C., Rusch, D. B., Romine, M. F., M Jennings, R. De, et al. (2016). Ecophysiology of an uncultivated lineage of Aigarchaeota from an oxic, hot spring

filamentous “streamer” community. *ISME J.* 10, 210–224. doi:10.1038/ismej.2015.83.
Berlemont, R., and Martiny, A. C. (2013). Phylogenetic Distribution of Potential Cellulases in Bacteria. *Appl. Environ. Microbiol.* 79, 1545–1554. doi:10.1128/AEM.03305-12.
Berry, D., Mader, E., Lee, T. K., Woebken, D., Wang, Y., Zhu, D., et al. (2015). Tracking heavy water (D_2O) incorporation for identifying and sorting active microbial cells. *Proc. Natl. Acad. Sci.* 112, E194–E203. doi:10.1073/pnas.1420406112.
Bohra, V., Dafale, N. A., and Purohit, H. J. (2019). Understanding the alteration in rumen microbiome and CAZymes profile with diet and host through comparative metagenomic approach. *Arch. Microbiol.* 201, 1385–1397. doi:10.1007/s00203-019-01706-z.
Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi:10.1038/s41587-019-0209-9.
Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35, 725–731. doi:10.1038/nbt.3893.
Boyd, E. S., Fecteau, K. M., Havig, J. R., Shock, E. L., and Peters, J. W. (2012). Modeling the habitat range of phototrophs in Yellowstone National Park: Toward the development of a comprehensive fitness landscape. *Front. Microbiol.* 3, 1–11. doi:10.3389/fmicb.2012.00221.
Brethauer, S., and Studer, M. H. (2014). Consolidated bioprocessing of lignocellulose by a microbial consortium. *Energy Environ. Sci.* 7, 1446–1453. doi:10.1039/c3ee41753k.
Brouzes, E., Medkova, M., Savenelli, N., Marran, D., Twardowski, M., Hutchison, J. B., et al. (2009). Droplet microfluidic technology for single-cell high-throughput screening. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14195–14200. doi:10.1073/pnas.0903542106.
Brumm, P. J., Gowda, K., Robb, F. T., and Mead, D. A. (2016). The Complete Genome Sequence of Hyperthermophile Dictyoglomus turgidum DSM 6724 ™ Reveals a Specialized Carbohydrate Fermentor. *Front. Microbiol.* 7. doi:10.3389/fmicb.2016.01979.
Brunecky, R., Alahuhta, M., Xu, Q., Donohoe, B. S., Crowley, M. F., Kataeva, I. A., et al. (2013). Revealing Nature’s Cellulase Diversity: The Digestion Mechanism of Caldicellulosiruptor bescii CelA. *Science* 342, 1513–1516.
Bugg, T. D. H., Ahmad, M., Hardiman, E. M., and Rahmanpour, R. (2011). Pathways for degradation of lignin in bacteria and fungi. *Nat. Prod. Rep.* 28, 1883–1896. doi:10.1039/c1np00042j.
Cabrera, M. Á., and Blamey, J. M. (2018). Biotechnological applications of archaeal enzymes from extreme environments. *Biol. Res.* 51, 1–15. doi:10.1186/s40659-018-0186-3.
Campanharo, J. C., Kielak, A. M., Castellane, T. C. L., Kuramae, E. E., and Lemos, E. G. de M. (2016). Optimized medium culture for Acidobacteria subdivision 1 strains. *FEMS Microbiol. Lett.* 363, 1–7. doi:10.1093/femsle/fnw245.
Campbell, K. M., Kouris, A., England, W., Anderson, R. E., McCleskey, R. B., Nordstrom, D. K., et al. (2017). Sulfolobus islandicus meta-populations in Yellowstone National Park hot springs. *Environ. Microbiol.* 19, 2334–2347. doi:10.1111/1462-2920.13728.
Castelle, C. J., and Banfield, J. F. (2018). Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. *Cell* 172, 1181–1197.

doi:10.1016/j.cell.2018.02.016.
Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi:10.1093/bioinformatics/btz848.
Cheah, W. Y., Sankaran, R., Show, P. L., Ibrahim, T. N. B. T., Chew, K. W., Culaba, A., et al. (2020). Pretreatment methods for lignocellulosic biofuels production: Current advances, challenges and future prospects. *Biofuel Res. J.* 7, 1115–1127. doi:10.18331/BRJ2020.7.1.4.
Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2021a). The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* 49, D751–D763. doi:10.1093/nar/gkaa939.
Chen, L., Zhao, B., Li, X., Cheng, Z., Wu, R., and Xia, Y. (2021b). Isolating and characterizing translationally active fraction of anammox microbiota using bioorthogonal non-canonical amino acid tagging. *Chem. Eng. J.* 418, 129411. doi:10.1016/j.cej.2021.129411.
Chung, D., Cha, M., Guss, A. M., and Westpheling, J. (2014). Direct conversion of plant biomass to ethanol by engineered Caldicellulosiruptor bescii. *Proc. Natl. Acad. Sci.* 111, 8931–8936. doi:10.1073/pnas.1402210111.
Colman, D. R., Lindsay, M. R., and Boyd, E. S. (2019). Mixing of meteoric and geothermal fluids supports hyperdiverse chemosynthetic hydrothermal communities. *Nat. Commun.* 10. doi:10.1038/s41467-019-08499-1.
Couradeau, E., Sasse, J., Goudeau, D., Nath, N., Hazen, T. C., Bowen, B. P., et al. (2019). Probing the active fraction of soil microbiomes using BONCAT-FACS. *Nat. Commun.* 10. doi:10.1038/s41467-019-10542-0.
Cozen, A. E., Weirauch, M. T., Pollard, K. S., Bernick, D. L., Stuart, J. M., Lowe, T. M., et al. (2009). Transcriptional Map of Respiratory Versatility in the Hyperthermophilic Crenarchaeon Pyrobaculum aerophilum □ †. 191, 782–794. doi:10.1128/JB.00965-08.
Daebeler, A., Herbold, C. W., Vierheilig, J., Sedlacek, C. J., Pjevac, P., Albertsen, M., et al. (2018). Cultivation and Genomic Analysis of " Candidatus Nitrosocaldus islandicus ," an Obligately Thaumarchaeon from a Hot Spring. *Front. Microbiol.* 9, 1–16. doi:10.3389/fmicb.2018.00193.
Daims, H., Lücker, S., and Wagner, M. (2006). daime , a novel image analysis program for microbial ecology and biofilm research. *Environ. Microbiol.* 8, 200–213. doi:10.1111/j.1462-2920.2005.00880.x.
Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* 3, 853–859. doi:10.1016/S0969-2126(01)00220-9.
Davis, N. M., Proctor, Di. M., Holmes, S. P., Relman, D. A., and Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 6, 1–14. doi:10.1186/s40168-018-0605-2.
De León, K. B., Gerlach, R., Peyton, B. M., and Fields, M. W. (2013). Archaeal and bacterial communities in three alkaline hot springs in heart lake geyser basin, yellowstone national park. *Front. Microbiol.* 4, 1–10. doi:10.3389/fmicb.2013.00330.
DeCastro, M. E., Rodríguez-Belmonte, E., and González-Siso, M. I. (2016). Metagenomics of thermophiles with a focus on discovery of novel thermozymes. *Front. Microbiol.* 7, 1–21. doi:10.3389/fmicb.2016.01521.
Dekas, A. E., Poretsky, R. S., and Orphan, V. J. (2009). Deep-Sea Archaea Fix and Share

Nitrogen in Methane-Consuming Microbial Consortia. *Science (80-.).* 326. doi:doi:10.1126/science.1178223.

Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A., and Schuman, E. M. (2006). Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* 103, 9482–9487. doi:10.1073/pnas.0601637103.

Dodsworth, J. A., Gevorkian, J., Despujos, F., Cole, J. K., Murugapiran, S. K., Ming, H., et al. (2014). Thermoflexus hugenholtzii gen. nov., sp. nov., a thermophilic, microaerophilic, filamentous bacterium representing a novel class in the Chloroflexi, Thermoflexia classis nov., and description of Thermoflexaceae fam. nov. and Thermoflexales ord. nov. *Int. J. Syst. Evol. Microbiol.* 64, 2119–2127. doi:10.1099/ijs.0.055855-0.

Doud, D. F. R., Bowers, R. M., Schulz, F., Raad, M. De, Deng, K., Tarver, A., et al. (2019). Function-driven single-cell genomics uncovers cellulose-degrading bacteria from the rare biosphere. *ISME J.* doi:10.1038/s41396-019-0557-y.

Eichorst, S. A., Trojan, D., Roux, S., Herbold, C., Rattei, T., and Woebken, D. (2018). Genomic insights into the Acidobacteria reveal strategies for their success in terrestrial environments. *Environ. Microbiol.* 20, 1041–1063. doi:10.1111/1462-2920.14043.

Eichorst, S. A., Varanasi, P., Stavila, V., Zemla, M., Auer, M., Singh, S., et al. (2013). Community dynamics of cellulose-adapted thermophilic bacterial consortia. 15, 2573–2587. doi:10.1111/1462-2920.12159.

Ellepola, K., Truong, T., Liu, Y., Lin, Q., Lim, T. K., Lee, Y. M., et al. (2019). Multi-omics Analyses Reveal Synergistic Carbohydrate Metabolism in Streptococcus mutans-Candida albicans Mixed-Species Biofilms. *Infect. Immun.* 87, 1–20. doi:10.1128/IAI.00339-19.

Eloe-Fadrosh, E. A., Ivanova, N. N., Woyke, T., and Kyrpides, N. C. (2016a). Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nat. Microbiol.* 1, 15032. doi:10.1038/nmicrobiol.2015.32.

Eloe-Fadrosh, E. A., Paez-Espino, D., Jarett, J., Dunfield, P. F., Hedlund, B. P., Dekas, A. E., et al. (2016b). Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat. Commun.* 7, 1–10. doi:10.1038/ncomms10476.

Fusco, F. A., Ronca, R., Fiorentino, G., Pedone, E., Contursi, P., Bartolucci, S., et al. (2018). Biochemical characterization of a thermostable endomannanase / endoglucanase from Dictyoglomus turgidum. *Extremophiles* 22, 131–140. doi:10.1007/s00792-017-0983-6.

García-Timermans, C., Props, R., Zacchetti, B., Sakarika, M., Delvigne, F., and Boon, N. (2020). Raman spectroscopy-based measurements of single-cell phenotypic diversity in microbial communities. *mSphere* 5, 1–14. doi:10.1101/2020.05.21.109934.

Girfoglio, M., Rossi, M., and Cannio, R. (2012). Cellulose degradation by Sulfolobus solfataricus requires a cell-anchored endo-β-1-4-glucanase. *J. Bacteriol.* 194, 5091–5100. doi:10.1128/JB.00672-12.

Graham, J. E., Clark, M. E., Nadler, D. C., Huffer, S., Chokhawala, H. A., Rowland, S. E., et al. (2011). Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat. Commun.* 2, 375–379. doi:10.1038/ncomms1373.

Gupta, A., and Verma, J. P. (2015). Sustainable bio-ethanol production from agro-residues: A review. *Renew. Sustain. Energy Rev.* 41, 550–567. doi:10.1016/j.rser.2014.08.032.

Hatzenpichler, R., Connon, S. A., Goudeau, D., Malmstrom, R. R., Woyke, T., and Orphan, V. J.

(2016). Visualizing in situ translational activity for identifying and sorting slow-growing archaeal–bacterial consortia. *Proc. Natl. Acad. Sci.* 113, E4069–E4078. doi:10.1073/pnas.1603757113.

Hatzenpichler, R., Krukenberg, V., Spietz, R. L., and Jay, Z. J. (2020). Next-generation physiology approaches to study microbiome function at single cell level. *Nat. Rev. Microbiol.* 18. doi:10.1038/s41579-020-0323-1.

Hatzenpichler, R., Lebedeva, E. V., Spieck, E., Stoecker, K., Richter, A., Daims, H., et al. (2008). A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc. Natl. Acad. Sci.* 105, 2134–2139. doi:10.1073/pnas.0708857105.

Hatzenpichler, R., Scheller, S., Tavormina, P. L., Babin, B. M., Tirrell, D. A., and Orphan, V. J. (2014). In situ visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. *Environ. Microbiol.* 16, 2568–2590. doi:10.1111/1462-2920.12436.

Hatzenpichler, R., Orphan, V. J. (2015). Detection of protein-synthesizing microorganisms in the environment via bioorthogonal noncanonical amino acid tagging (BONCAT). *Hydrocarb. Lipid Microbiol. Protoc. - Springer Protoc. Handbooks*, 1–29. doi:10.1007/8623.

He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R. H., et al. (2013). Comparative Metagenomic and Metatranscriptomic Analysis of Hindgut Paunch Microbiota in Wood- and Dung-Feeding Higher Termites. *PLoS One* 8. doi:10.1371/journal.pone.0061126.

He, Y.-C., Ma, C.-L., and Yang, B. (2018). "Pretreatment Process and Its Synergistic Effects on Enzymatic Digestion of Lignocellulosic Material," in *Fungal Cellulolytic Enzymes: Microbial Production and Application*, 1–25. doi:10.1007/978-981-13-0749-2.

Hedlund, B. P., Dodsworth, J. A., Murugapiran, S. K., Rinke, C., and Woyke, T. (2014). Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter." *Extremophiles* 18, 865–875. doi:10.1007/s00792-014-0664-7.

Hedlund, B. P., Murugapiran, S. K., Alba, T. W., Levy, A., Dodsworth, J. A., Goertz, G. B., et al. (2015). Uncultivated thermophiles: Current status and spotlight on "Aigarchaeota." *Curr. Opin. Microbiol.* 25, 136–145. doi:10.1016/j.mib.2015.06.008.

Heins, R. A., Cheng, X., Nath, S., Deng, K., Bowen, B. P., Chivian, D. C., et al. (2014). Phylogenomically Guided Identi fi cation of Industrially Relevant GH1 β - Glucosidases through DNA Synthesis and Nanostructure-Initiator Mass Spectrometry. doi:10.1021/cb500244v.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P. H., et al. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* 5. doi:10.1038/srep14567.

Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., et al. (2011). Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science (80-.).* 463, 463–467. doi:10.1126/science.1200387.

Heyer, R., Kohrs, F., Benndorf, D., Rapp, E., Kausmann, R., Heiermann, M., et al. (2013). Metaproteome analysis of the microbial communities in agricultural biogas plants. *N. Biotechnol.* 30, 614–622. doi:10.1016/j.nbt.2013.01.002.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical J.* 50, 346–363. doi:10.1002/bimj.200810425.

Hou, W., Wang, S., Dong, H., Jiang, H., Briggs, B. R., Peacock, J. P., et al. (2013). A Comprehensive Census of Microbial Diversity in Hot Springs of Tengchong, Yunnan Province China Using 16S rRNA Gene Pyrosequencing. *PLoS One* 8. doi:10.1371/journal.pone.0053350.

Hua, Z. S., Qu, Y. N., Zhu, Q., Zhou, E. M., Qi, Y. L., Yin, Y. R., et al. (2018). Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nat. Commun.* 9. doi:10.1038/s41467-018-05284-4.

Huang, W. E., Griffiths, R. I., Thompson, I. P., Bailey, M. J., and Whiteley, A. S. (2004). Raman microscopic analysis of single microbial cells. *Anal. Chem.* 76, 4452–4458. doi:10.1021/ac049753k.

Huang, W. E., Ward, A. D., and Whiteley, A. S. (2009). Raman tweezers sorting of single microbial cells. *Environ. Microbiol. Rep.* 1, 44–49. doi:10.1111/j.1758-2229.2008.00002.x.

Huber, R., Dyba, D., Huber, H., Burggraf, S., and Rachel, R. (1998). Sulfur-inhibited Thermosphaera aggregans sp. nov., a new genus of hyperthermophilic archaea isolated after its prediction from environmentally derived 16S rRNA sequences. *Int. J. Syst. Bacteriol.* 48, 31–38. doi:10.1099/00207713-48-1-31.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nat. Microbiol.*, 1–6. doi:10.1038/nmicrobiol.2016.48.

Inskeep, W. P., Jay, Z. J., Herrgard, M. J., Kozubal, M. A., Rusch, D. B., Tringe, S. G., et al. (2013a). Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. *Front. Microbiol.* 4. doi:10.3389/fmicb.2013.00095.

Inskeep, W. P., Jay, Z. J., Tringe, S. G., Herrgård, M. J., and Rusch, D. B. (2013b). The YNP metagenome project: Environmental parameters responsible for microbial distribution in the yellowstone geothermal ecosystem. *Front. Microbiol.* 4, 1–15. doi:10.3389/fmicb.2013.00067.

Inskeep, W. P., Rusch, D. B., Jay, Z. J., Herrgard, M. J., Kozubal, M. A., Richardson, T. H., et al. (2010). Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS One* 5. doi:10.1371/journal.pone.0009773.

Ishii, S., Shimoyama, T., Hotta, Y., and Watanabe, K. (2008). Characterization of a filamentous biofilm community established in a cellulose-fed microbial fuel cell. *BMC Microbiol.* 8, 1–12. doi:10.1186/1471-2180-8-6.

Itoh, T., Suzuki, K., and Sanchez, P. C. (1999). new genus of rod-shaped crenarchaeote isolated from a hot spring in the Philippines. *Int. J. Syst. Bacteriol.*, 1–7.

Jay, Z. J., Beam, J. P., Dlakić, M., Rusch, D. B., Kozubal, M. A., and Inskeep, W. P. (2018). Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. *Nat. Microbiol.* 3, 732–740. doi:10.1038/s41564-018-0163-1.

Jay, Z. J., Beam, J. P., Dohnalkova, A., Lohmayer, R., Bodle, B., Planer-Friedrich, B., et al. (2015). Pyrobaculum yellowstonensis strain WP30 respires on elemental sulfur and/or arsenate in circumneutral sulfidic geothermal sediments of Yellowstone National Park. *Appl. Environ. Microbiol.* 81, 5907–5916. doi:10.1128/AEM.01095-15.

Jay, Z. J., Beam, J. P., Kozubal, M. A., Jennings, R. de M., Rusch, D. B., and Inskeep, W. P.

(2016). The distribution, diversity and function of predominant Thermoproteales in high-temperature environments of Yellowstone National Park. *Environ. Microbiol.* 18, 4755–4769. doi:10.1111/1462-2920.13366.

Jensen, C. U., Rodriguez Guerrero, J. K., Karatzos, S., Olofsson, G., and Iversen, S. B. (2017). Fundamentals of Hydrofaction^{TM}: Renewable crude oil from woody biomass. *Biomass Convers. Biorefinery* 7, 495–509. doi:10.1007/s13399-017-0248-8.

Jones, R. T., Robeson, M. S., Lauber, C. L., Hamady, M., Knight, R., and Fierer, N. (2009). A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J.* 3, 442–453. doi:10.1038/ismej.2008.127.

Kang, D. D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 2015, 1–15. doi:10.7717/peerj.1165.

Karst, S. M., Dueholm, M. S., McIlroy, S. J., Kirkegaard, R. H., Nielsen, P. H., and Albertsen, M. (2018). Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nat. Biotechnol.* 36, 190–195. doi:10.1038/nbt.4045.

Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A., and Kuramae, E. E. (2016). The ecology of Acidobacteria: Moving beyond genes and genomes. *Front. Microbiol.* 7, 1–16. doi:10.3389/fmicb.2016.00744.

Kiick, K. L., Saxon, E., Tirrell, D. A., and Bertozzi, C. R. (2002). Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U. S. A.* 99, 19–24. doi:10.1073/pnas.012583299.

Kim, S. K., Himmel, M. E., Bomble, Y. J., and Westpheling, J. (2018). Expression of a cellobiose phosphorylase from Thermotoga maritima in Caldicellulosiruptor bescii improves the phosphorolytic pathway and results in a dramatic increase in cellulolytic activity. *Appl. Environ. Microbiol.* 84, 1–9. doi:10.1128/AEM.02348-17.

Klatt, C. G., Inskeep, W. P., Herrgard, M. J., Jay, Z. J., Rusch, D. B., Tringe, S. G., et al. (2013). Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Front. Microbiol.* 4, 1–23. doi:10.3389/fmicb.2013.00106.

Klatt, C. G., Wood, J. M., Rusch, D. B., Bateson, M. M., Hamamura, N., Heidelberg, J. F., et al. (2011). Community ecology of hot spring cyanobacterial mats: Predominant populations and their functional potential. *ISME J.* 5, 1262–1278. doi:10.1038/ismej.2011.73.

Kochetkova, T. V, Toshchakov, S. V, Zayulina, K. S., Elcheninov, A. G., Zavarzina, D. G., Lavrushin, V. Y., et al. (2020). Hot in Cold : Microbial Life in the Hottest Springs in Permafrost. *Microorganisms* 8, 1–18. doi:doi:10.3390/microorganisms8091308.

Kolde, R. (2015). pheatmap : Pretty Heatmaps. *R Packag. version 1.0.8*, 1–7. Available at: https://cran.r-project.org/package=pheatmap.

Lauber, C. L., Hamady, M., Knight, R., and Fierer, N. (2009). Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale □ †. 75, 5111–5120. doi:10.1128/AEM.00335-09.

Lavrentyeva, E. V., Radnagurueva, A. A., Barkhutova, D. D., Belkova, N. L., Zaitseva, S. V., Namsaraev, Z. B., et al. (2018). Bacterial Diversity and Functional Activity of Microbial Communities in Hot Springs of the Baikal Rift Zone. *Microbiol. (Russian Fed.* 87, 272–281. doi:10.1134/S0026261718020078.

Lee, J. Y., Iglesias, B., Chu, C. E., Lawrence, D. J. P., Jerome, E., and Iii, C. (2017). Pyrobaculum igneiluti sp . nov ., a novel anaerobic hyperthermophilic archaeon that reduces thiosulfate and ferric iron. 1714–1719. doi:10.1099/ijsem.0.001850.

Lee, K. S., Palatinszky, M., Pereira, F. C., Nguyen, J., Fernandez, V. I., Mueller, A. J., et al. (2019). An automated Raman-based platform for the sorting of live cells by functional properties. *Nat. Microbiol.* 4, 1035–1048. doi:10.1038/s41564-019-0394-9.

Leizeaga, A., Estrany, M., Forn, I., and Sebastián, M. (2017). Using click-chemistry for visualizing in situ changes of translational activity in planktonic marine bacteria. *Front. Microbiol.* 8, 1–11. doi:10.3389/fmicb.2017.02360.

Leung, K., Zahn, H., Leaver, T., Konwar, K. M., Hanson, N. W., Pagé, A. P., et al. (2012). A programmable droplet-based microfluidic device applied to multiparameter analysis of single microbes and microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* 109, 7665–7670. doi:10.1073/pnas.1106752109.

Levasseur, A., Lomascolo, A., Chabrol, O., Ruiz-Dueñas, F. J., Boukhris-Uzan, E., Piumi, F., et al. (2014). The genome of the white-rot fungus Pycnoporus cinnabarinus: A basidiomycete model with a versatile arsenal for lignocellulosic biomass breakdown. *BMC Genomics* 15, 1–24. doi:10.1186/1471-2164-15-486.

Lewin, A., Zhou, J., Pham, V. T. T., Haugen, T., Zeiny, M. El, Aarstad, O., et al. (2017). Novel archaeal thermostable cellulases from an oil reservoir metagenome. *AMB Express* 7. doi:10.1186/s13568-017-0485-z.

Liu, L., Lv, A. P., Li, M. M., Ming, Y. Z., Jiao, J. Y., Fang, B. Z., et al. (2020a). Seramator thermalis gen. Nov., sp. nov., a novel cellulose-and xylan-degrading member of the family dysgonamonadaceae isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* 70, 5717–5724. doi:10.1099/ijsem.0.004469.

Liu, Y., Xu, J., Tao, Y., Fang, T., Du, W., and Ye, A. (2020b). Rapid and accurate identification of marine microbes with single-cell Raman spectroscopy. *Analyst* 145, 3297–3305. doi:10.1039/c9an02069a.

Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J., and Crosby, L. (2018). Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. *mSystems* 3, 1–12. doi:10.1128/msystems.00055-18.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, 490–495. doi:10.1093/nar/gkt1178.

López-Mondéjar, R., Tláskal, V., Větrovský, T., Štursová, M., Toscan, R., Nunes da Rocha, U., et al. (2020). Metagenomics and stable isotope probing reveal the complementary contribution of fungal and bacterial communities in the recycling of dead biomass in forest soil. *Soil Biol. Biochem.* 148. doi:10.1016/j.soilbio.2020.107875.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 1–21. doi:10.1186/s13059-014-0550-8.

Lu, J., Breitwieser, F. P., Thielen, P., and Salzberg, S. L. (2017). Bracken: Estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* 2017, 1–17. doi:10.7717/peerj-cs.104.

Ma, L., Zhao, Y., Meng, L., Wang, X., Yi, Y., Shan, Y., et al. (2020). Isolation of Thermostable Lignocellulosic Bacteria From Chicken Manure Compost and a M42 Family Endocellulase

Cloning From Geobacillus thermodenitrificans Y7. *Front. Microbiol.* 11, 1–15. doi:10.3389/fmicb.2020.00281.

Martins, L. F., Antunes, L. P., Pascon, R. C., de Oliveira, J. C. F., Digiampietri, L. A., Barbosa, D., et al. (2013). Metagenomic Analysis of a Tropical Composting Operation at the São Paulo Zoo Park Reveals Diversity of Biomass Degradation Functions and Organisms. *PLoS One* 8. doi:10.1371/journal.pone.0061928.

Marynowska, M., Goux, X., Sillam-Dussès, D., Rouland-Lefèvre, C., Halder, R., Wilmes, P., et al. (2020). Compositional and functional characterisation of biomass-degrading microbial communities in guts of plant fibre- and soil-feeding higher termites. *Microbiome* 8, 96. doi:10.1186/s40168-020-00872-3.

Mathrani, I. M., and Ahring, B. K. (1992). Applied Microbiology Biotechnology Thermophilic and alkalophilic xylanases from several Dictyoglomus isolates. *Appl. Microbiol. Biotechnol.* 1, 23–27.

Mathrani, I. M., Nielsen, P., Sonne-Hansen, J., Kristjansson, J. K., and Ahring, B. K. (1993). Influence of pH and temperature on enumeration of cellulose- and hemicellulose-degrading thermophilic anaerobes in neutral and alkaline Icelandic hot springs. *Appl. Environ. Microbiol.* 59, 1963–1965. doi:10.1128/aem.59.6.1963-1965.1993.

McKay, L. J., Dlakić, M., Fields, M. W., Delmont, T. O., Eren, A. M., Jay, Z. J., et al. (2019). Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. *Nat. Microbiol.* 4, 614–622. doi:10.1038/s41564-019-0362-4.

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8. doi:10.1371/journal.pone.0061217.

Menzel, P., Gudbergsd, R., Rike, G., Lin, L., Zhang, Q., Contursi, P., et al. (2015). Comparative Metagenomics of Eight Geographically Remote Terrestrial Hot Springs. *Environ. Microbiol.*, 411–424. doi:10.1007/s00248-015-0576-9.

Michels, D. E., Lomenick, B., Chou, T.-F., Sweredoski, M. J., and Pasulka, A. (2021). Amino acid analog induces stress response in marine Synechococcus . *Appl. Environ. Microbiol.* doi:10.1128/aem.00200-21.

Miller, I. J., Rees, E. R., Ross, J., Miller, I., Baxa, J., Lopera, J., et al. (2019). Autometa: Automated extraction of microbial genomes from individual shotgun metagenomes. *Nucleic Acids Res.* 47, 1–12. doi:10.1093/nar/gkz148.

Milshteyn, A., Schneider, J. S., and Brady, S. F. (2014). Mining the metabiome: Identifying novel natural products from microbial communities. *Chem. Biol.* 21, 1211–1223. doi:10.1016/j.chembiol.2014.08.006.

Minich, J. J., Zhu, Q., Janssen, S., Hendrickson, R., Amir, A., Vetter, R., et al. (2018). KatharoSeq Enables High-Throughput Microbiome Analysis from Low-Biomass Samples. *mSystems* 3, 1–16. doi:10.1128/msystems.00218-17.

Mladenovska, Z., Mathrani, I. M., and Ahring, B. K. (1995). Isolation and characterization of Caldicellulosiruptor lactoaceticus sp. nov., an extremely thermophilic, cellulolytic, anaerobic bacterium. *Arch. Microbiol.* 163, 223–230. doi:10.1007/BF00305357.

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Sundaramurthi, J. C., Lee, J., et al. (2021). Genomes OnLine Database (GOLD) v.8: overview and updates. *Nucleic Acids Res.*

49, D723–D733. doi:10.1093/nar/gkaa983.
Niederberger, T. D., Götz, D. K., McDonald, I. R., Ronimus, R. S., and Morgan, H. W. (2006a). Ignisphaera aggregans gen. nov., sp. nov., a novel hyperthermophilic crenarchaeote isolated from hot springs in Rotorua and Tokaanu, New Zealand. *Int. J. Syst. Evol. Microbiol.* 56, 965–971. doi:10.1099/ijs.0.63899-0.
Nunoura, T., Sako, Y., Wakagi, T., and Uchida, A. (2003). Regulation of the aerobic respiratory chain in the facultatively aerobic and hyperthermophilic archaeon Pyrobaculum oguniense. 673–688. doi:10.1099/mic.0.26000-0.
Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2019). vegan: Community Ecology Package. *R Packag. version 2.5-4*. Available at: https://cran.r-project/package=vegan.
Opdahl, L. J., Gonda, M. G., and St-Pierre, B. (2018). Identification of Uncultured Bacterial Species from Firmicutes, Bacteroidetes and CANDIDATUS Saccharibacteria as Candidate Cellulose Utilizers from the Rumen of Beef Cows. *Microorganisms* 6, 17. doi:10.3390/microorganisms6010017.
Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. doi:10.1111/1462-2920.13023.
Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996. doi:10.1038/nbt.4229.
Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi:10.1101/gr.186072.114.
Pasulka, A. L., Thamatrakoln, K., Kopf, S. H., Guan, Y., Poulos, B., Moradian, A., et al. (2018). Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. *Environ. Microbiol.* 20, 671–692. doi:10.1111/1462-2920.13996.
Peacock, J. P., Cole, J. K., Murugapiran, S. K., Dodsworth, J. A., Fisher, J. C., Moser, D. P., et al. (2013). Pyrosequencing Reveals High-Temperature Cellulolytic Microbial Consortia in Great Boiling Spring after In Situ Lignocellulose Enrichment. *PLoS One* 8, 1–12. doi:10.1371/journal.pone.0059927.
Pedersen, M., Johansen, K. S., and Meyer, A. S. (2011). Low temperature lignocellulose pretreatment: Effects and interactions of pretreatment pH are critical for maximizing enzymatic monosaccharide yields from wheat straw. *Biotechnol. Biofuels* 4, 1–10. doi:10.1186/1754-6834-4-11.
Perevalova, A. A., Bidzhieva, S. K., Kublanov, I. V., Hinrichs, K. U., Liu, X. L., Mardanov, A. V., et al. (2010). Fervidicoccus fontis gen. nov., sp. nov., an anaerobic, thermophilic crenarchaeote from terrestrial hot springs, and proposal of Fervidicoccaceae fam. nov. and Fervidicoccales ord. nov. *Int. J. Syst. Evol. Microbiol.* 60, 2082–2088. doi:10.1099/ijs.0.019042-0.
Podar, P. T., Yang, Z., Björnsdóttir, S. H., and Podar, M. (2020). Comparative Analysis of Microbial Diversity Across Temperature Gradients in Hot Springs From Yellowstone and Iceland. *Front. Microbiol.* 11, 1–16. doi:10.3389/fmicb.2020.01625.

Podosokorskaya, O. A., Kadnikov, V. V., Gavrilov, S. N., Mardanov, A. V., Merkel, A. Y., Karnachuk, O. V., et al. (2013). Characterization of Melioribacter roseus gen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class Ignavibacteria, and a proposal of a novel bacterial phylum Ignavibacteriae. *Environ. Microbiol.* 15, 1759–1771. doi:10.1111/1462-2920.12067.

Pohlschroeder, M., Leschine, S. B., and Canale-Parola, E. (1994). Spirochaeta caldaria sp. nov., a thermophilic bacterium that enhances cellulose degradation by Clostridium thermocellum. *Arch. Microbiol.* 161, 17–24. doi:10.1007/BF00248889.

Power, J. F., Carere, C. R., Lee, C. K., Wakerley, G. L. J., Evans, D. W., Button, M., et al. (2018). Microbial biogeography of 925 geothermal springs in New Zealand. *Nat. Commun.* 9. doi:10.1038/s41467-018-05020-y.

R Core Team (2014). R: A language and environment for statistical computing. *R Found. Stat. Comput.* doi:10.18637/jss.v067.i01.R.

Rai, R., Agrawal, D., and Chadha, B. S. (2019). *New paradigm in degradation of lignocellulosic biomass and discovery of novel microbial strains*. doi:10.1007/978-981-13-8487-5_16.

Ransom-Jones, E., McCarthy, A. J., Haldenby, S., Doonan, J., and McDonald, J. E. (2017). Lignocellulose-degrading microbial communities in landfill sites represent a repository of unexplored biomass-degrading diversity. *mSphere* 2, 1–13.

Reichart, N. J., Bowers, R. M., Woyke, T., and Hatzenpichler, R. (2021). High Potential for Biomass-Degrading Enzymes Revealed by Hot Spring Metagenomics. *Front. Microbiol.* 12, 1–13. doi:10.3389/fmicb.2021.668238.

Reichart, N. J., Jay, Z. J., Krukenberg, V., Parker, A. E., Spietz, R. L., and Hatzenpichler, R. (2020). Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. *ISME J.* 14, 2851–2861. doi:10.1038/s41396-020-00749-1.

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., et al. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* 9, 1038–1048. doi:10.1038/nprot.2014.067.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437. doi:10.1038/nature12352.

Riva, A., Kolimár, D., Spittler, A., Wisgrill, L., Herbold, C. W., Abrankó, L., et al. (2020). Conversion of Rutin, a Prevalent Dietary Flavonol, by the Human Gut Microbiota. *Front. Microbiol.* 11, 1–11. doi:10.3389/fmicb.2020.585428.

Sadler, N. C., and Wright, A. T. (2015). Activity-based protein profiling of microbes. *Curr. Opin. Chem. Biol.* 24, 139–144. doi:10.1016/j.cbpa.2014.10.022.

Saiki, T., Kobayashi, Y., Kawagoe, K., and Beppu, T. (1985). Dictyoglomus thermophilum gen. nov., sp. nov., a Chemoorganotrophic, Anaerobic, Thermophilic Bacterium. *Int. J. Syst. Bacteriol.* 35, 253–259.

Samo, T. J., Smriga, S., Malfatti, F., Sherwood, B. P., and Azam, F. (2014). Broad distribution and high proportion of protein synthesis active marine bacteria revealed by click chemistry at the single cell level. *Front. Mar. Sci.* 1. doi:10.3389/fmars.2014.00048.

Saxena, R., Dhakan, D. B., Mittal, P., Waiker, P., Chowdhury, A., Ghatak, A., et al. (2017). Metagenomic analysis of hot springs in central india reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Front.*

Microbiol. 7. doi:10.3389/fmicb.2016.02123.
Sharp, C. E., Brady, A. L., Sharp, G. H., Grasby, S. E., Stott, M. B., and Dunfield, P. F. (2014). Humboldt's spa: Microbial diversity is controlled by temperature in geothermal environments. *ISME J.* 8, 1166–1174. doi:10.1038/ismej.2013.237.
Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., et al. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* 3, 836–843. doi:10.1038/s41564-018-0171-1.
Singer, E., Wagner, M., and Woyke, T. (2017). Capturing the genetic makeup of the active microbiome in situ. *ISME J.* 11, 1949–1963. doi:10.1038/ismej.2017.59.
Singh, N., Mathur, A. S., Tuli, D. K., Gupta, R. P., Barrow, C. J., and Puri, M. (2017). Cellulosic ethanol production via consolidated bioprocessing by a novel thermophilic anaerobic bacterium isolated from a Himalayan hot spring. *Biotechnol. Biofuels* 10, 1–18. doi:10.1186/s13068-017-0756-6.
Slobodkina, G. B., Lebedinsky, A. V, Chernyh, N. A., and Slobodkin, A. I. (2015). Pyrobaculum ferrireducens sp . nov ., a hyperthermophilic Fe (III) - , selenate- and arsenate- reducing crenarchaeon isolated from a hot spring. 851–856. doi:10.1099/ijs.0.000027.
Song, Z. Q., Wang, F. P., Zhi, X. Y., Chen, J. Q., Zhou, E. M., Liang, F., et al. (2013). Bacterial and archaeal diversities in Yunnan and Tibetan hot springs, China. *Environ. Microbiol.* 15, 1160–1175. doi:10.1111/1462-2920.12025.
Sonne-Hansen, J., Mathrani, I. M., and Ahring, B. K. (1993). Xylanolytic anaerobic thermophiles from Icelandic hot-springs. *Appl. Microbiol. Biotechnol.* 38, 537–541. doi:10.1007/BF00242951.
Steward, K. F., Eilers, B., Tripet, B., Fuchs, A., Dorle, M., Rawle, R., et al. (2020). Metabolic Implications of Using BioOrthogonal Non-Canonical Amino Acid Tagging (BONCAT) for Tracking Protein Synthesis. *Front. Microbiol.* 11, 1–12. doi:10.3389/fmicb.2020.00197.
Strazzulli, A., Cobucci-Ponzano, B., Iacono, R., Giglio, R., Maurelli, L., Curci, N., et al. (2020). Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments. *FEBS J.* 287, 1116–1137. doi:10.1111/febs.15080.
Suleiman, M., Krüger, A., and Antranikian, G. (2020). Biomass-degrading glycoside hydrolases of archaeal origin. *Biotechnol. Biofuels* 13, 1–14. doi:10.1186/s13068-020-01792-y.
Sun, Y., Liu, Y., Pan, J., Wang, F., and Li, M. (2020). Perspectives on Cultivation Strategies of Archaea. *Microb. Ecol.* 79, 770–784. doi:10.1007/s00248-019-01422-7.
Taillefer, M., Arntzen, M. Ø., Henrissat, B., Pope, P. B., and Larsbrink, J. (2018). Proteomic Dissection of the Cellulolytic Machineries Used by Soil-Dwelling Bacteroidetes . *mSystems* 3, 1–16. doi:10.1128/msystems.00240-18.
Takacs-Vesbach, C., Inskeep, W. P., Jay, Z. J., Herrgard, M. J., Rusch, D. B., Tringe, S. G., et al. (2013). Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct geothermal channels reveals specialization of three aquificales lineages. *Front. Microbiol.* 4, 1–25. doi:10.3389/fmicb.2013.00084.
Tange, O. (2018). *GNU Parallel 2018*.
Terekhov, S. S., Smirnov, I. V., Malakhova, M. V., Samoilov, A. E., Manolo, A. I., Nazarov, A. S., et al. (2018). Ultrahigh-throughput functional profiling of microbiota communities. *Proc. Natl. Acad. Sci. U. S. A.* 115, 9551–9556. doi:10.1073/pnas.1811250115.
Terekhov, S. S., Smirnov, I. V., Stepanova, A. V., Bobik, T. V., Mokrushina, Y. A.,

Ponomarenko, N. A., et al. (2017). Microfluidic droplet platform for ultrahigh-throughput single-cell screening of biodiversity. *Proc. Natl. Acad. Sci. U. S. A.* 114, 2550–2555. doi:10.1073/pnas.1621226114.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. doi:10.1038/nature24621.

Torre, J. R. De, Walker, C. B., Ingalls, A. E., Könneke, M., and Stahl, D. A. (2008). Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ. Microbiol.* 10, 810–818. doi:10.1111/j.1462-2920.2007.01506.x.

Toshchakov, S. V., Korzhenkov, A. A., Samarov, N. I., Mazunin, I. O., Mozhey, O. I., Shmyr, I. S., et al. (2015). Complete genome sequence of and proposal of Thermofilum uzonense sp. nov. A novel hyperthermophilic crenarchaeon and emended description of the genus Thermofilum. *Stand. Genomic Sci.* 10, 1–9. doi:10.1186/s40793-015-0105-y.

Valentini, T. D., Lucas, S. K., Binder, K. A., Cameron, L. C., Motl, J. A., Dunitz, J. M., et al. (2020). Bioorthogonal non-canonical amino acid tagging reveals translationally active subpopulations of the cystic fibrosis lung microbiota. *Nat. Commun.* 18, S36. doi:10.1016/s1569-1993(19)30233-4.

Van Der Helm, E., Genee, H. J., and Sommer, M. O. A. (2018). The evolving interface between synthetic biology and functional metagenomics. *Nat. Chem. Biol.* 14, 752–759. doi:10.1038/s41589-018-0100-x.

Wagner, M. (2009). Single-Cell Ecophysiology of Microbes as Revealed by Raman Microspectroscopy or Secondary Ion Mass Spectrometry Imaging. *Annu. Rev. Microbiol.* 63, 411–429. doi:10.1146/annurev.micro.091208.073233.

Wang, C., Dong, D., Wang, H., Müller, K., Qin, Y., Wang, H., et al. (2016). Metagenomic analysis of microbial consortia enriched from compost: New insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnol. Biofuels* 9, 1–17. doi:10.1186/s13068-016-0440-2.

Wang, G., Zhao, Z., Ke, J., Engel, Y., Shi, Y. M., Robinson, D., et al. (2019). CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. *Nat. Microbiol.* 4, 2498–2510. doi:10.1038/s41564-019-0573-8.

Wilhelm, R. C., Singh, R., Eltis, L. D., and Mohn, W. W. (2018). Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. *ISME J.*, 413–429. doi:10.1038/s41396-018-0279-6.

Wohlgemuth, R., Littlechild, J., Monti, D., Schnorr, K., Rossum, T. van, Siebers, B., et al. (2018). Discovering novel hydrolases from hot environments. *Biotechnol. Adv.* 36, 2077–2100. doi:10.1016/j.biotechadv.2018.09.004.

Wong, M. T., Wang, W., Couturier, M., Razeq, F. M., Lombard, V., Lapebie, P., et al. (2017). Comparative metagenomics of cellulose- and poplar hydrolysate-degrading microcosms from gut microflora of the Canadian Beaver (Castor canadensis) and North American moose (Alces americanus) after long-term enrichment. *Front. Microbiol.* 8, 1–14. doi:10.3389/fmicb.2017.02504.

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 1–13. doi:10.1186/s13059-019-1891-0.

Wu, Y. W., Simmons, B. A., and Singer, S. W. (2016). MaxBin 2.0: An automated binning

algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32, 605–607. doi:10.1093/bioinformatics/btv638.

Xue, M. Y., Sun, H. Z., Wu, X. H., Liu, J. X., and Guan, L. L. (2020). Multi-omics reveals that the rumen microbiome and its metabolome together with the host metabolome contribute to individualized dairy cow performance. *Microbiome* 8, 1–19. doi:10.1186/s40168-020-00819-8.

Yang, S. J., Kataeva, I., Hamilton-Brehm, S. D., Engle, N. L., Tschaplinski, T. J., Doeppke, C., et al. (2009). Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "Anaerocellum thermophilum" DSM 6725. *Appl. Environ. Microbiol.* 75, 4762–4769. doi:10.1128/AEM.00236-09.

Yang, S. J., Kataeva, I., Wiegel, J., Yin, Y., Dam, P., Xu, Y., et al. (2010). Classification of "Anaerocellum thermophilum" strain DSM 6725 as Caldicellulosiruptor bescii sp. nov. *Int. J. Syst. Evol. Microbiol.* 60, 2011–2015. doi:10.1099/ijs.0.017731-0.

Young, J., Chung, D., Bomble, Y. J., Himmel, M. E., and Westpheling, J. (2014). Deletion of *Caldicellulosiruptor bescii* CelA reveals its crucial role in the deconstruction of lignocellulosic biomass. *Biotechnol. Biofuels* 7, 142. doi:10.1186/s13068-014-0142-6.

Zarafeta, D., Kissas, D., Sayer, C., Gudbergsdottir, S. R., Ladoukakis, E., Isupov, M. N., et al. (2016). Discovery and characterization of a thermostable and highly halotolerant GH5 cellulase from an icelandic hot spring isolate. *PLoS One* 11, 1–18. doi:10.1371/journal.pone.0146454.

Zayulina, K. S., Kochetkova, T. V, Piunova, U. E., and Morgan, H. (2020). Novel Hyperthermophilic Crenarchaeon Thermofilum and Two Novel Glycosidases for Cellulose Hydrolysis. *Front. Microbiol.* 10, 1–12. doi:10.3389/fmicb.2019.02972.

Zegeye, E. K., Sadler, N. C., Lomas, G. X., Attah, I. K., Jansson, J. K., Hofmockel, K. S., et al. (2020). Activity-Based Protein Profiling of Chitin Catabolism. *ChemBioChem* 21, 1–8. doi:10.1002/cbic.202000616.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi:10.1093/nar/gky418.

Zhao, C., Deng, Y., Wang, X., Li, Q., Huang, Y., and Liu, B. (2014). Identification and characterization of an anaerobic ethanol-producing cellulolytic bacterial consortium from great basin hot springs with agricultural residues and energy crops. *J. Microbiol. Biotechnol.* 24, 1280–1290. doi:10.4014/jmb.1401.01022.

APPENDICES

APPENDIX A

SUPPLEMENTARY TABLES

Supplementary Table 1. Incubation conditions. Description of sample names with substrate and HPG amendment. Final concentration listed, "-" represents no addition. All samples were incubated in triplicate vials at 74 °C. Unless otherwise indicated, all incubations were done under atmospheric conditions (21 % O_2).

Sample name	Amendment	Substrate concentration	HPG concentration
No-HPG	-	-	-
HPG-only	-	-	50 µM
Acetate	Acetate	10 µM	50 µM
Aspartate	Aspartate	10 µM	50 µM
Biotin	Biotin	10 µM	50 µM
Cellobiose	Cellobiose	10 µM	50 µM
Cellulose	Cellulose	10 µM	50 µM
Glucose	Glucose	10 µM	50 µM
Glycerol	Glycerol	100 µM	50 µM
Glycine	Glycine	10 µM	50 µM
Isoleucine	Isoleucine	10 µM	50 µM
Leucine	Leucine	10 µM	50 µM
NH_4^+ 0.1 mg/L	NH_4^+	5.6 µM	50 µM
NH_4^+ 2 mg/L	NH_4^+	112 µM	50 µM
NH_4^+ 5 mg/L	NH_4^+	280 µM	50 µM
Nitrate	Nitrate	100 µM	50 µM
Nitrite	Nitrite	100 µM	50 µM
Pyruvate	Pyruvate	10 µM	50 µM
Riboflavin	Riboflavin	10 µM	50 µM
Ribose	Ribose	10 µM	50 µM
Serine	Serine	10 µM	50 µM
Thiamine	Thiamine	10 µM	50 µM
Valine	Valine	10 µM	50 µM
Anoxic	N_2 gas	100 % N_2	50 µM
Microoxic	N_2 gas and atmospheric air	90 % N_2, 10 % atmospheric air (2 % O_2)	50 µM

Supplementary Table 2. Sorter metrics for BONCAT incubations. Events sorted for each substrate amendment and all replicated listed. Efficiency (%) represents how many positive events were sorted compared to how many positive events had to be aborted and discarded in the FACS prior to sorting. Percent of total events is the number of positive events that comprised the total events before any gating restrictions were applied.

Sample name	Replicate 1			Replicate 2			Replicate 3		
	BONCAT+ events sorted	Efficiency (%)	Percent of total events	BONCAT+ events sorted	Efficiency (%)	Percent of total events	BONCAT+ events sorted	Efficiency (%)	Percent of total events
No-HPG	66	97	0.02	43	95	0.01	16	94	0.01
HPG-only	250,000	95	28.76	250,000	92	23.25	191549	97	33.89
Acetate	250,144	95	25.32	250,000	93	28.14	250000	98	64.67
Aspartate	250,000	96	33.32	250,000	93	26.08	207048	99	69.49
Biotin	250,000	95	20.53	250,000	93	30.56	118346	98	59.14
Cellobiose	125,073	93	6.61	199,739	92	10.50	38283	98	16.18
Cellulose	250,000	93	19.52	250,000	93	23.34	162525	98	58.16
Glucose	250,000	95	36.65	250,000	93	32.69	250000	97	56.30
Glycerol	250,000	92	25.44	250,000	93	28.08	250000	97	45.97
Glycine	99,964	93	4.57	250,000	92	16.44	66074	97	19.85
Isoleucine	250,000	92	16.77	179,666	93	10.24	250000	98	60.61
Leucine	250,000	93	25.66	250,000	93	17.52	250000	98	54.54
NH_4^+ 0.1 mg/L	250,000	96	27.37	250,000	91	22.96	250000	97	54.99
NH_4^+ 2 mg/L	250,000	94	21.03	250,000	93	27.88	75177	98	33.22
NH_4^+ 5 mg/L	250,000	94	26.80	250,000	93	27.16	250000	98	51.76
Nitrate	250,000	95	27.33	85,132	96	9.69	250000	97	43.79
Nitrite	175,598	95	24.93	80,164	97	15.69	250000	98	49.55
Pyruvate	250,000	93	25.60	250,000	94	30.93	177582	98	48.67
Riboflavin	250,000	93	20.89	250,000	93	23.46	250000	98	58.44
Ribose	250,000	94	24.03	250,000	93	19.27	250000	98	58.32
Serine	250,000	95	33.29	250,000	94	28.42	250000	97	49.13
Thiamine	250,000	95	33.42	250,000	94	26.88	119730	98	36.01
Valine	250,000	92	20.65	250,000	91	23.02	250000	97	50.00
Anoxic	150,302	86	6.5	125,452	88	4.85	78192	97	16.20
Microoxic	250,000	89	20.38	250,000	91	23.56	250000	97	49.77

Supplementary Table 3. Amplicon sequence variants (ASVs) for all taxa shown in sorted samples of Figure 4. Data processing grouped ASVs for similar taxa into one entry. Shown are the total number of ASVs for each taxon represented as well as the number of ASVs that had 16S rRNA gene sequence counts. Dashed lines represent taxa that were not statistically significant and thus not shown in Figure for that substrate.

Taxon	Cellobiose ASVs with sequencing count	N_2 100 % ASVs with sequencing count	Total ASVs
Crenarchaeota Pyrobaculum	1	3	37
Thaumarchaeota	-	1	4
Acidobacteria	4	-	5
Armatimonadetes	9	-	36
Aquificae Thermocrinis	17	14	58
BP4	13	-	34
Chloroflexi Thermoflexus	3	5	32
Cyanobacteria Synechococcus	-	2	29
Deinococcus-Thermus Thermus	-	27	127
Fervidibacteria	19	10	128
Gal15	5	3	13

Supplementary Table 1. Metagenome metadata. This file contains the TaxonOID for correct identification to the IMG database along with pertinent metadata on sample collection and any publication information.

Supplementary Table 2. Taxonomic profile at the genus level from Kraken2/Bracken for all metagenomes.

Supplementary Table 3. Output of α-diversity results for all metagenome taxonomic profile. Taxonomy of the metagenomic contigs was assigned using Kraken2 and the relative abundance calculated with Bracken. Diversity metrics were measured using Phyloseq R package.

Supplementary Table 4. Pairwise comparisons for α- and β-diversity for the taxonomic profile of temperature and pH categories. The α-diversity compared observed taxa and Shannon diversity metrics using the lme4 R package. The β-diversity compared Bray Curtis dissimilarity of taxonomic composition using the vegan R package. The p values were adjusted for pairwise comparisons using the Holm method using the base R stats function p.adjust().

Category type	Diversity metric	Comparison	p value	p value adjust holm
Temperature	Observed	Mesophilic vs thermophilic	0.000363	0.000726
Temperature	Observed	Mesophilic vs hyperthermophilic	0.000105	0.000315
Temperature	Observed	Thermophilic vs hyperthermophilic	0.181	0.181
Temperature	Shannon	Mesophilic vs thermophilic	0.000137	0.000274
Temperature	Shannon	Mesophilic vs hyperthermophilic	0.0000124	0.0000372
Temperature	Shannon	Thermophilic vs hyperthermophilic	0.0619	0.0619
Temperature	β-diversity	Mesophilic vs thermophilic	0.003	0.003
Temperature	β-diversity	Mesophilic vs hyperthermophilic	0.001	0.003
Temperature	β-diversity	Thermophilic vs hyperthermophilic	0.001	0.003
pH	Observed	Highly acidic vs mildly acidic	0.221	0.882
pH	Observed	Highly acidic vs neutral	0.0121	0.0726
pH	Observed	Highly acidic vs alkaline	0.0274	0.137
pH	Observed	Mildly acidic vs neutral	0.369	1
pH	Observed	Mildly acidic vs alkaline	0.468	1
pH	Observed	Neutral vs alkaline	0.866	1
pH	Shannon	Highly acidic vs mildly acidic	0.451	1
pH	Shannon	Highly acidic vs neutral	0.082	0.492
pH	Shannon	Highly acidic vs alkaline	0.539	1
pH	Shannon	Mildly acidic vs neutral	0.476	1
pH	Shannon	Mildly acidic vs alkaline	0.771	1
pH	Shannon	Neutral vs alkaline	0.142	0.71
pH	β-diversity	Highly acidic vs mildly acidic	0.149	0.208
pH	β-diversity	Highly acidic vs neutral	0.001	0.006
pH	β-diversity	Highly acidic vs alkaline	0.001	0.006
pH	β-diversity	Mildly acidic vs neutral	0.017	0.068
pH	β-diversity	Mildly acidic vs alkaline	0.025	0.075
pH	β-diversity	Neutral vs alkaline	0.104	0.208

Supplementary Table 5. CAZyme hits corresponding to IMG/M gene IDs for all IMG/M taxon OIDs. Only CAZyme genes with hits to two or more of dbCAN2 tools are listed and used within the analysis.

Supplementary Table 6. Sankey links used for generation of Figure 4. All taxa present in the enzyme output included in this file.

Supplementary Table 1. Decontam identified contaminant ASVs removed from sequencing count data.

AVSID	Sequence	Taxonomy
ASV36	d829bee4984f82ffc2453212157caf96	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Xanthobacteraceae; g__Bradyrhizobium
ASV55	8e09caaee61bd482522f087e96e093de	d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Bacillaceae; g__Geobacillus
ASV65	6376ea6dab7d0fde3cd66f53b57e1484	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas
ASV72	5648dccee530d68ceb3e4d7d22cf8756	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas
ASV73	0ba0d4c5a289facc80e0e481377a9e18	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Mesorhizobium
ASV77	f7447c8e079023f4f2579d8580575dd3	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus; s__Lactobacillus_iners
ASV87	0920dcf0f62fb2b3ab9e32f1c4edec37	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Burkholderiales; f__Comamonadaceae
ASV93	65d43491988bfe557da4d86a5ba25dae	d__Bacteria; p__Firmicutes; c__Bacilli; o__Staphylococcales; f__Staphylococcaceae; g__Staphylococcus
ASV94	fd496fd32dc8c08ade2e8b6c9d8ee13d	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus; s__Streptococcus_salivarius
ASV104	a7da93a9744a7863292bcf97d47b6a10	d__Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Micrococcales; f__Intrasporangiaceae
ASV105	06f825b512d903b9230e1a55d87359ee	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus
ASV116	b060079a052d8ef0971aa8faf2adb85c	d__Bacteria; p__Firmicutes; c__Bacilli; o__Staphylococcales; f__Staphylococcaceae; g__Staphylococcus
ASV129	d7348b86110af99c20176a0c0aa0697a	d__Bacteria; p__Cyanobacteria; c__Vampirivibrionia; o__Obscuribacterales; f__Obscuribacteraceae; g__Candidatus_Obscuribacter; s__uncultured_bacterium
ASV135	1735122f6452284cae1574c71c9a6eec	d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Bacillaceae; g__Aeribacillus
ASV148	65641452be75528754b5ae258ba661c1	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Aerococcaceae; g__Aerococcus
ASV159	acebaf0e1d8b741c8a7b36c5bf093863	d__Bacteria; p__Deinococcota; c__Deinococci; o__Thermales; f__Thermaceae; g__Thermus; s__Thermus_thermophilus
ASV168	2ddb215ff668b6a24a9ccf2bb076a453	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Enterococcaceae; g__Tetragenococcus; s__Tetragenococcus_halophilus
ASV173	fccd274e718e26b32c71f188d97d9e59	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__Paracoccus
ASV185	6aae9a17fbd9c7a4f55a65ef18d63320	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Beijerinckiaceae; g__Methylobacterium-Methylorubrum
ASV204	aa9b3a1418d146c262ec63305292065a	d__Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Corynebacteriales; f__Corynebacteriaceae; g__Corynebacterium
ASV208	e5c19d7800b18015f3a917fc015fc42f	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Enhydrobacter
ASV209	5a7b179b1b45f0fe2282f260bf073f60	d__Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Propionibacteriales; f__Propionibacteriaceae; g__Cutibacterium
ASV220	103c3f14abf946f37daa78195c99a002	d__Bacteria; p__Deinococcota; c__Deinococci; o__Deinococcales; f__Deinococcaceae; g__Deinococcus; s__Deinococcus_geothermalis
ASV255	e8386d3a307c208c4b9f0a756259cd6b	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter
ASV267	882b5b49ef18215727a0900958cdec5d	d__Bacteria; p__Deinococcota; c__Deinococci; o__Thermales; f__Thermaceae; g__Meiothermus

Supplementary Table 2. Pairwise comparisons made for β-diversity of sample type categories and substrates used in BONCAT-FACS experiments amended with or without crystalline cellulose.

		Comparison	P value
		Substrate	0.050
		Sample type	0.001
	Sample type	Pre vs SYBR	0.008
		Pre vs BON	0.003
		SYBR vs BON	0.004
	Substrate	Control vs HPG	0.220
		Control vs Avicel	0.168
		Avicel vs HPG	0.241
Sample type	Presort	Avicel vs HPG	0.183
	BONCAT	Avicel vs HPG	0.100
Substrate	Avicel	Pre vs BON	0.028
	HPG	Pre vs BON	0.061

Supplementary Table 3. Top 20 ASVs with full taxonomy identified by V4 amplicon sequencing and Silva 138 database.

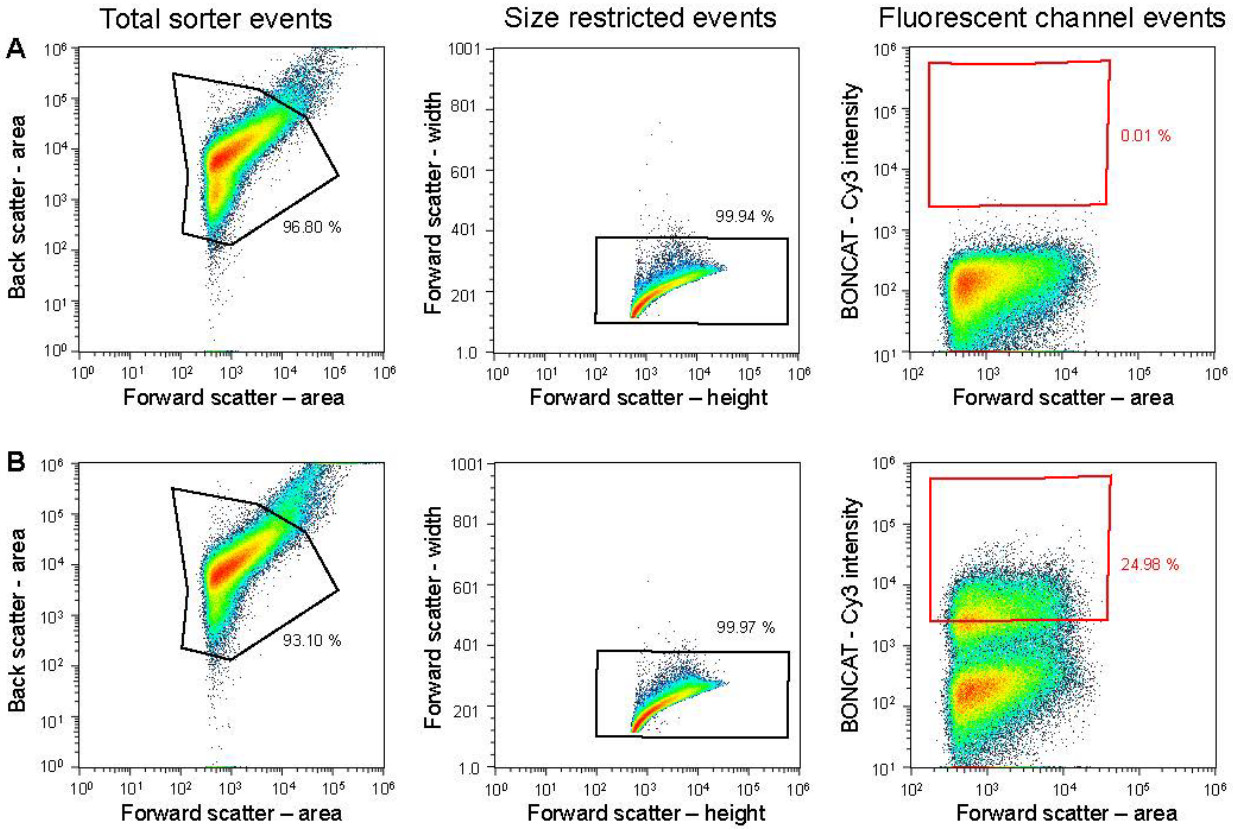
Supplementary Table 4. Sequencing statistics for bins generated from Buffalo Pool metagenomic sample.

Supplementary Table 1. Metagenomic statistics and processing information.

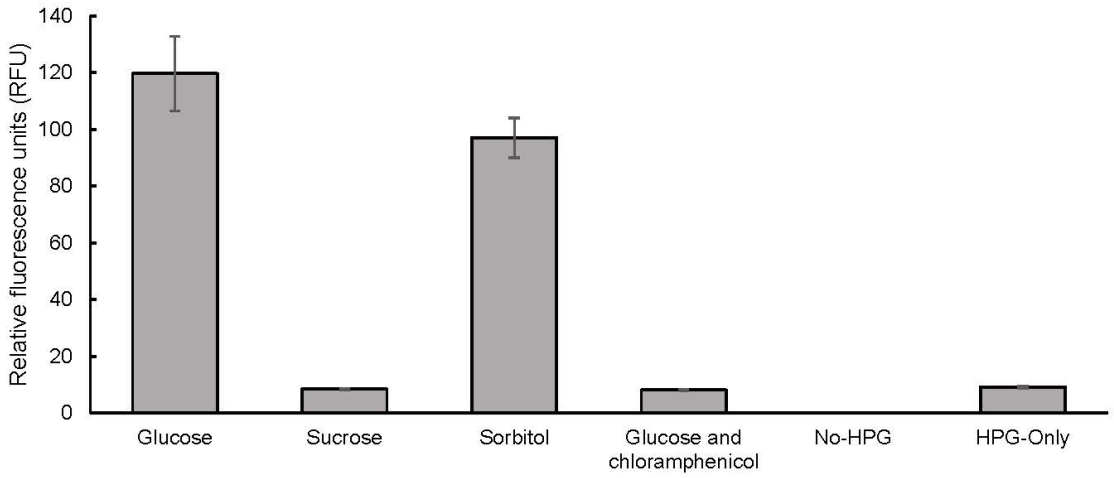
Supplementary Table 2. Bin statistics generated from IMG/M.

APPENDIX B

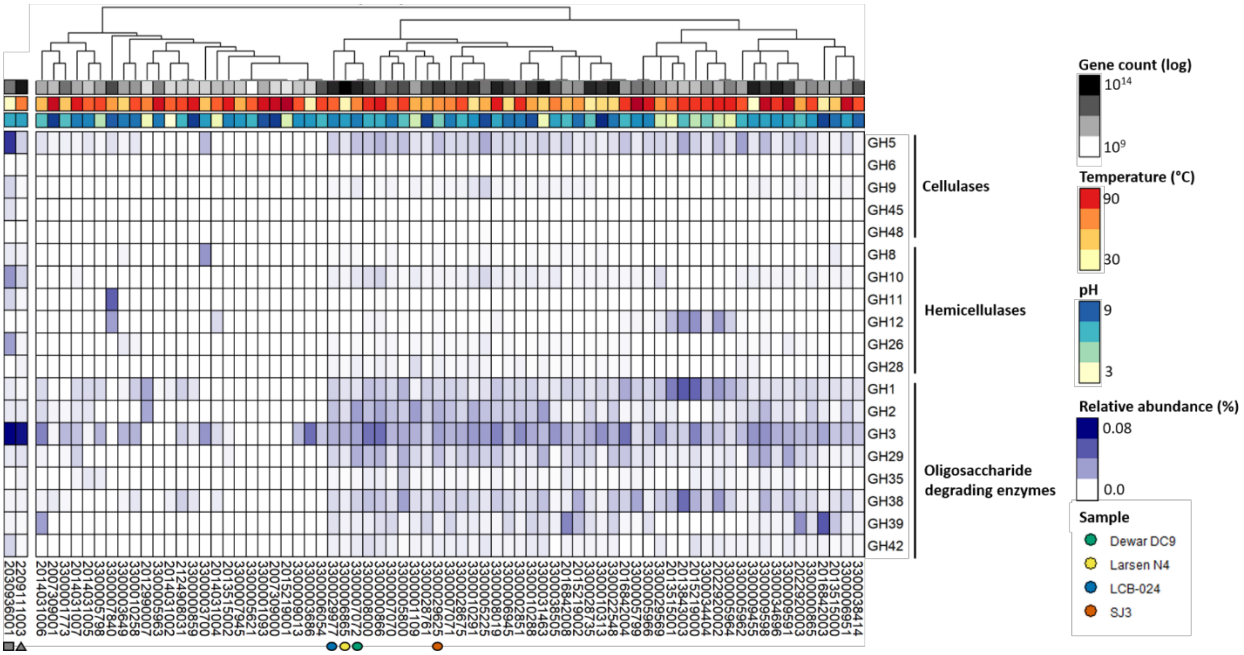
SUPPLEMENTARY FIGURES



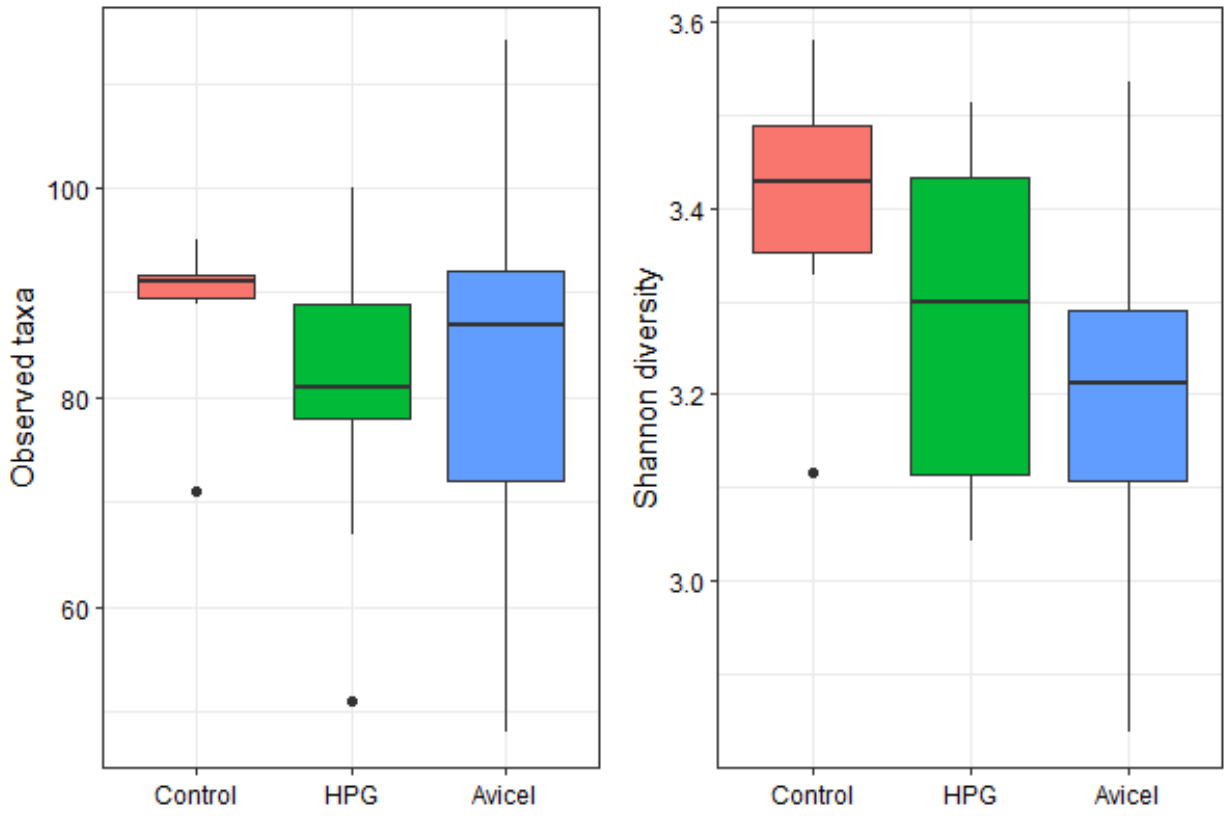
Supplementary Figure 1. BONCAT sorting gates. (a) No-HPG control. (b) HPG-only control. First gate is selecting cells based on size. Second gate again selects cells based on size. These two gates restrict larger particles and clumps of cells. The final gate is sorted into collection tubes and selects for BONCAT positive fluorescent cells. The BONCAT gate was drawn conservatively to reduce the number of false positives being collected. The gate was first drawn on the no-HPG sample to collect as few cells as possible. The gate settings were then transferred to all subsequent samples. Values shown next to each gate represent the proportion of events contained within that gate. The red gates are the final gates for each sample that were used for sorting.



Supplementary Figure 2. BONCAT substrate specificity tests with *E. coli*. Cultures of *E. coli* were grown in the presence of different substrates and HPG. Relative fluorescence units (RFU) were averaged from three replicates of each condition with error bars representing standard deviation. The cells in the no-HPG condition did not receive HPG during the incubation and thus could not be click stained and the RFU did not measure above the background fluorescence. Cells grown in the presence of sucrose showed only low levels of fluorescence, similar to the HPG-only control, consistent with the observations that *E. coli* is incapable of metabolizing sucrose.



Supplementary Figure 1. Relative abundance of CAZyme families out of total CAZyme gene hits. Specifically, 19 CAZymes with predicted cellulase, hemicellulase, and oligosaccharide degrading enzyme activities for 71 hot spring metagenomes and two additional non-hot spring metagenomes from cellulolytic sources. Sample clustering was kept consistent based on the dendrogram output from main text Figure 3. Symbols are as follows: grey square (2030936001 - *Nasutitermes corniger* P3 gut compartment), grey triangle (2209111003 - São Paulo Zoo compost), blue circle (3300029977 - LCB-024), yellow circle (3300006865 - Larsen N4), green circle (3300007072 - Dewar Creek DC9), and orange circle (3300029625 - SJ3). We caution the reader that direct comparisons of CAZyme abundances between features cannot be made due to the different sequencing technologies used during the 13-year timespan.



Supplementary Figure 1. α-diversity grouped by substrate and shown as box and whisker plots for observed taxa and Shannon diversity.

APPENDIX C

CURRICULUM VITAE

Nicholas J. Reichart
Chemistry and Biochemistry
Montana State University
Bozeman, MT 59715

Education

July 2021 Ph.D., Biochemistry
Montana State University, Bozeman
Advisor: Dr. Roland Hatzenpichler
Committee: Drs. Matthew Fields, Brent Peyton, Valérie Copié, and James Wilking
Dissertation title: Identification of cellulolytic hot spring organisms through bioorthogonal labeling

Publications

Reichart NJ, Jay ZJ, Krukenberg V, Woyke T, and Hatzenpichler R (2021) Effect of crystalline cellulose supplementation on a hot spring microbial community in a short time scale incubation. In preparation.

Reichart NJ, Bowers RM, Woyke T, and Hatzenpichler R (2021) Metagenomic assemblies of substrate-amended hot spring sediment incubations from Yellowstone National Park. In preparation.

Reichart NJ, Bowers RM, Woyke T, and Hatzenpichler R (2021) High Potential for Biomass-Degrading Enzymes Revealed by Hot Spring Metagenomics. *Front. Microbiol*. 12:668238. doi: 10.3389/fmicb.2021.668238.

Reichart, N. J., Jay, Z. J., Krukenberg, V., Parker, A. E., Spietz, R. L., Hatzenpichler, R. Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. *ISMEJ* 14, 2851-2861 (2020). https://doi.org/10.1038/s41396-020-00749-1

Awards

01/2020 – 12/2020 Department of Energy Graduate Student Research Fellowship
Host: Lawrence Berkeley National Laboratory, Joint Genome Institute, Dr. Tanja Woyke
08/2016 – 07/2021 Molecular Bioscience Fellowship, five years graduate tuition support
05/2018 – Montana Academy of Science Grant, $1500 research award

Presentations

07/2021 – Montana State University PhD Defense, Bozeman, Montana
01/2021 – Pacific Northwest National Lab seminar series, Richland, Washington
11/2020 – Chemistry and biochemistry departmental seminar, Bozeman, Montana
06/2020 – Integrative Genomics science forum seminar series, Berkeley, California
06/2019 – Geobiology society conference poster presentation, Banff, Canada
04/2019 – Thermal Biology Institute seminar, Montana State University, Bozeman, Montana
04/2019 – Montana Academy of Science oral presentation, Butte, Montana

08/2018 – International Symposium of Microbial Ecology poster presentation, Leipzig, Germany
04/2017 – Molecular Bioscience Fellowship oral presentation, Bozeman, Montana

Workshops

02/2018 – Integrated Microbial Genomics and Microbiome workshop, Walnut Creek, California
02/2020 – D-Labs R statistics course, Berkeley University, Berkeley, California

Teaching and Leadership Experience

06/2019 – Geobiology society conference session co-chair, Banff, Canada
01/2019 – 12/2019. Molecular bioscience program student representative, Montana State University, Bozeman, Montana
08/2017 – 12/2017 Teaching assistant. Biochemistry 300 level lab course, Montana State University, Bozeman, Montana