OMICS-BASED CHARACTERIZATION OF COMPLEX ANAEROBIC METABOLISM IN METHANOGENIC WASTEWATER TREATMENT

BY

MASARU KONISHI NOBU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Environmental Engineering in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Wen-Tso Liu, Chair
Professor Isaac K. O. Cann
Associate Professor Thanh H. Nyugen
Assistant Professor Roland D. Cusick
Professor Michael J. McInerney, University of Oklahoma

ABSTRACT

Below the familiar oxygenated biosphere lay ecosystems teeming with "anaerobic" prokaryotes thriving in the absence of O_2. As anaerobes exhaust compounds for favorable respiration (*e.g.*, NO_3^- and SO_4^{2-}), microorganisms resort to fermentation and respiration of H^+ and CO_2. Across Earth, microbial communities under such environmental conditions are estimated to annually mineralize 1~2 GT of organic carbon to CH_4 and CO_2, thereby driving a critical step in the global carbon cycle. Since the discovery that we can tame such "methanogenic" (methane-generating) microbial communities to convert society's organic waste to CH_4 as a recoverable fuel, this biotechnology has become an essential component of managing municipal and industrial waste and development of sustainable energy. Driven by the environmental and technological significance, research has found four major niches form metabolic interactions to facilitate methanogenic degradation of organic carbon: hydrolyzers, fermenters, syntrophs, and methanogens. Despite this defined general ecological structure, many organisms and metabolism in methanogenic ecosystems remain uncharacterized due to challenges in handling and cultivating anaerobes. To tackle this issue, we can employ rapidly developing sequencing technology to recover genomes for uncultivated organisms directly from the environment ("metagenomics"), obtain insight into their physiology, and ultimately uncover hitherto overlooked ecological and biochemical processes taking place in methanogenic natural ecosystems and engineered systems.

In the series of studies presented in this dissertation, we use methanogenic wastewater treatment bioreactors as model ecosystems and implement cutting-edge bioinformatics with rigorous annotation of anaerobic metabolic capacities to investigate the ecological roles of uncultured syntrophs, methanogens, and organisms from other bacterial lineages. For syntrophs, we characterize novel aromatic compound degradation pathways and find that syntrophic catabolism and interactions are much more diverse and flexible than previously anticipated, opening new possibilities for ecological niches that syntrophs can exploit. In investigating methanogens, we successfully recover the first genomes for a methanogen-related *Euryarchaeota* class WSA2 found across various anaerobic environments and discover that they encode unique H_2-oxidizing methyl-compound-reducing methanogenesis, suggesting that this may be a major process in both natural and engineered methanogenic environments. As for uncharacterized bacterial lineages, we acquire genomes for populations spanning 15 phyla, of which 5 are bacterial phyla with no cultured representatives ("candidate phyla"). We find that these organisms may contribute to novel syntrophic, fermentative, and acetogenic processes and form intricate metabolic interactions to facilitate complete mineralization of organic matter to in methanogenic ecosystems. Finally, to expand the application of the approach used throughout these studies, we compile the accumulated insight into

genomics and complex metabolism and perform an unprecedentedly large-scale comparative genomics analysis on a bacterial phylum that contains both uncultivated lineages affiliated with methanogenic ecosystems and poorly understood lineages prevalent across Earth: *Bacteroidetes*. This reveals novel relationships between phylogeny, metabolism, and habitats and unnoticed ecological roles that *Bacteroidetes* can take in methanogenic environments, marine ecosystems, and even the human gastrointestinal tract. In total, we demonstrate that integration of metagenomics, comparative genomics, and strict annotation of metabolic capacity can effectively characterize the ecophysiology of uncultivated organisms and reveal novel ecological niches in methanogenic environments and beyond.

DEDICATION

I dedicate this thesis to my beloved family for their unconditional encouragement and love, my mentors and their family who not only nurtured me as an academic but also supported me as dear friends, and friends with whom I shared both unforgettable joys and hardships throughout my graduate career. I would like to especially thank my mother, Amy Yoshiko Nobu, and father, Motoo Konishi, for everything they have done to allow me to pursue the path I have chosen. I also express my deep gratitude to the Tamaki and Narihiro families for guiding my academic growth while also caring for me as if family. Without support from them, my graduate career and personal life would not have been nearly as enjoyable or accomplished as it was; I never cease to feel lucky and privileged to have been gifted with such inimitable bliss. I truly believe that it is thanks to their support that I have grown both as an academic and as a person. Any further success from this moment on will also be thanks to them. I only hope to continuously give back such joy to them and others that I cross paths with.

ACKNOWLEDGEMENTS

I would also like to thank my adviser Wen-Tso Liu, Ph.D. committee members Michael McInerney, Isaac Cann, Helen Nguyen, and Roland Cusick, mentors, colleagues, fellowships, and funding agencies for their support throughout my graduate career. I had entered the graduate program with minimal knowledge on environmental engineering; I merely had passion and curiosity. However, the tireless guidance from my adviser, mentors, and colleagues from the University of Illinois at Urbana-Champaign and the National Institute of Advanced Industrial Science and Technology allowed me to truly translate this passion into the successful and impactful research presented in this dissertation. I am also grateful to fellow students for their friendship, cooperation, and feedback. In addition, the presented research was made possible through exceptional support from Richard S. Engelbrecht and his family, the Japan Society for the Promotion of Science, Terracon, the Department of Energy Joint Genome Institute, British Petroleum, the Energy Bioscience Institute, and the University of Illinois.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 – Background

Although atmospheric oxygen makes Earth inhabitable for many types of life, below Earth's surface lies a plethora of "anaerobic" ecosystems where oxygen does not penetrate to. As biomass produced on the surface trickle down into such ecosystems, microorganisms decompose the organic compounds and release CO_2 back to the atmosphere. However, roughly 1~2 GT carbon remains after organisms deplete compounds that can support energetic respiration (*e.g.*, sulfate and ferric iron). Under such extreme conditions, microorganisms resort to using H^+, CO_2, and endogenous metabolites as electron acceptors. Thriving on such restricted metabolism, specialized *Bacteria* and *Archaea* cooperatively accomplish mineralization of organic compounds to CO_2 and CH_4, whereby releasing biomass-bound carbon back to the atmosphere (*1, 2*). In nature, methanogenic ecosystems span aquatic sediments, wetlands, permafrost, groundwater, and other subsurface environments; clearly, understanding the physiology and ecological contributions of these organisms is paramount for furthering our knowledge of global biogeochemical cycles. Moreover, beginning from the 17th century discovery by Van Helmont that decaying organic matter produces flammable gases, by 1895 in Exeter, England, society could harness these anaerobic microorganisms to accomplish conversion of sewage to biogas for fueling street lamps (*3, 4*). Now, anaerobic wastewater treatment is a widely employed biotechnology in harvesting energy from both municipal and industrial waste. Thus, advancing our comprehension of methanogenic microbial communities is also critical for environmental stewardship and development of sustainable energy.

Compared to respiration of favorable electron acceptors (*e.g.*, O_2, NO_3^-, SO_4^{2-}, and Fe^{3+}), electron disposal through fermentation and H^+/CO_2 respiration are much more complex due to energetic and thermodynamic limitations. For example, while aerobic propionate catabolism generates ~20 ATP and is quite exergonic (ΔG° = -1493 kJ mol^{-1}), anaerobic propionate degradation to acetate, CO_2, and H_2 only generates <1 ATP and is thermodynamically challenging (ΔG° = +76 kJ mol^{-1}). Given such drastic differences, it is no wonder that anaerobic metabolism tends to be more complex. Further, the correspondingly slow growth rates and difficulty in taming these organisms (*5*) has hindered progress in characterization of anaerobic metabolism and cultivation of uncharacterized microbial populations in methanogenic ecosystems. For instance, despite knowing that H_2 metabolism is a core component of anaerobic life since the 1890s, biochemical characterization of hydrogenases are on-going to this day due to difficulty in handling the organisms and their enzymes (*6, 7*). Thus, although major efforts have been poured into biochemical characterization of isolates (*8, 9*) and cultivation of the unknown (*10*), much of

how microorganisms accomplish energy-limited anaerobic metabolism and what the vast majority of microorganisms in nature accomplish remain unknown.

Although the details of many types of anaerobic metabolism remain unclear, the general scheme of microbial ecology in methanogenic ecosystems is well-known. Whether starting with organic macromolecules, simple organic molecules, or aromatic compounds, microorganisms cooperatively degrade organic compounds in a step-wise manner. Organic macromolecules (*e.g.*, polysaccharides and protein) are first decomposed to individual components (*e.g.*, sugars and amino acids) by hydrolyzers. Fermenters degrade sugars and labile amino acids to simple organic acids (*e.g.*, fatty acids), alcohols, H_2, formate, and CO_2. Finally, methanogens accomplish mineralization of acetate, H_2, and formate to CO_2 and CH_4. However, many compounds including alcohols, fatty acids, amino acids, and aromatic compounds are non-fermentable because accumulation of its metabolic byproducts (*e.g.*, acetate, H_2, and formate) is thermodynamically inhibitory (*11-13*). Thus, specialized organotrophs form metabolic partnerships with acetate-, H_2-, and formate-scavenging methanogens to accomplish "syntrophic" catabolism of non-fermentable compounds, whereby serving as a key component of carbon cycling in methanogenic ecosystems. In addition to this essential ecological contribution, syntrophic metabolizers play critical roles in methanogenic wastewater treatment processes. Syntrophic metabolizers degrade acids produced by hydrolyzers and fermenters to not only bridge the gap between fermentation and methanogenesis, but also prevent process failure by mitigating acid accumulation (*14, 15*). In addition, they are essential for driving remediation of aromatic compounds contained in waste from the petrochemical industry. However, despite this ecological and biotechnological importance, characterization of syntrophs' metabolism and interactions has been hampered by challenges in cultivation and biochemical characterization.

In addition to the aforementioned niches, methanogenic ecosystems harbor diverse organisms with unknown function. Since the advent of DNA sequencing and other molecular tools, we have become increasingly aware that much of the microorganisms present in nature remain to be characterized (*16, 17*). This also holds true for both natural and engineered methanogenic ecosystems (*18, 19*). Classification based on 16S ribosomal RNA (rRNA) sequences shows that the two microbial domains, *Bacteria* and *Archaea*, can be divided into at least 85 lineages, of which more than half lack cultivated representatives (55 lineages known as "candidate phyla"). Further inspection of the public 16S rRNA databases reveals an astounding diversity of uncultivated microorganisms affiliated with methanogenic ecosystems. Surprisingly, over 20 of the known candidate phyla have been observed in methanogenic environments. Moreover, many of these phyla inhabit wide ranges of anaerobic ecosystems (*e.g.*, SAR406 in marine and anaerobic digesters and OP9 in lakes, hot springs, and various methanogenic ecosystems), implicating that they make overlooked ecological contributions consistently relevant under diverse conditions. Thus, unraveling their metabolic

capacity and ecological contributions is necessary to fully understand the ecology of methanogenic decomposition in natural and engineered systems.

1.2 – (Meta)Genomic approaches to overcome cultivation-based obstacles

Investigation of syntrophs and candidate phyla has been challenging due to difficulty in cultivation. However, advances in sequencing technology has allowed researchers to obtain high-quality genomes for isolates and even uncultivated populations. In 1977, DNA sequencing technology reached a major breakthrough with development of Sanger sequencing, also known as the "chain termination" technique (*20*). Optimization of the method allowed development of machines that could sequence DNA fragments around one kilobase in length (*21*). Taking advantage of this technology, the first complete genome of a bacterium (*Haemophilus influenzae*) was sequenced in 1995 using whole-genome shotgun sequencing in which the genomic DNA is randomly sheared into short segments, sequenced segment-by-segment through Sanger sequencing, and reconstructed as a consensus sequence (*22*).

As another approach, "pyrosequencing" achieved a higher sequencing efficiency than Sanger sequencing capacity by creatively taking advantage of the pyrophosphate produced during DNA synthesis (*23*). During polymerization of a target DNA fragment by a DNA polymerase, this technique sequentially supplies individual nucleotides (*i.e.*, A/T/C/G) and detects whether a specific nucleotide was integrated into DNA polymerization through luminescently measuring the produced pyrophosphate using the ATP sulfurylase in combination with luciferase. 454 further promoted this technique by developing sequencing machines that could efficiently yield large amounts of sequences by attaching individual DNA fragments onto separate beads and running individual sequencing runs in parallel for each bead. In 2008, researchers combined Sanger- and pyro-sequencing to economically obtain unprecedentedly high-quality microbial genomes and obtain deep insight into soil and marine microbial ecology (*24*). In the same year, researchers developed an alternative parallelization approach known as the Solexa method that could also achieve high throughput sequencing by attaching billions of DNA fragments to a single flow-cell and simultaneously determining the sequence of all fragments with the use of specialized deoxynucleotides that release fluorescent signals when integrated into DNA polymerization reactions (*25, 26*). Illumina has produced various machines (*e.g.*, Genome Analyzer, MiSeq, and HiSeq) that can perform such massive DNA sequencing at different costs, throughput, run times, and read lengths. Development of pyrosequencing and Illumina sequencing has made genome sequencing much more accessible, leading to the acquisition of thousands of prokaryotic genomes over the past two decades.

In 2008, researchers acquired the first genomic insight into syntrophic metabolism through Sanger sequencing of the *Syntrophus aciditrophicus* (*27*), *Moorella thermoacetica* (*28*), and *Pelotomaculum*

thermopropionicium (*29, 30*) genomes. With the breakthroughs of 454 and Illumina sequencing, researchers have been able to obtain additional high-quality genomes at low cost, leading to major efforts in genomic characterization of syntrophic metabolizers, including *Aminobacterium colombiense* (*31*), *Thermoacetogenium phaeum* (*32*), *Syntrophorhabdus aromaticivorans* (*33*), and many others. Collectively, these syntrophic metabolizer genomes have elucidated remarkable insight into the metabolic pathways and energy conservation involved in sustaining life at the thermodynamic limit (*12*). Such efforts have effectively characterized how syntrophic metabolizers catabolize acetate (*34*), propionate (*29*), butyrate (*35*), and benzoate (*27*). Genomic investigations reveal that these organisms often conserve energy and streamline catabolism by pairing exergonic and endergonic reactions during catabolism, such as decarboxylation with carboxylation (*e.g.*, carboxyltransferase) and thioester hydrolysis with synthesis (*e.g.*, CoA transferase). This theme also carries over into syntrophic metabolizers' strategies for accomplishing thermodynamically restricted H_2 production. Genomics revealed that they take advantage of two unique electron transfer reactions to optimize energy conservation: reverse electron transport and electron confurcation. To achieve H_2 production, a high-energy electron donor (*e.g.*, ferredoxin) is necessary, but catabolism of syntrophic substrates often yields low-energy electron carriers (*e.g.*, NADH). Reverse electron transport uses the cross-membrane H^+ gradient (*i.e.*, proton motive force / PMF) to drive vectorial electron transfer to a carrier with sufficiently low reduction potential for H_2 generation. Electron confurcation rather uses exergonic H_2 generation from a high-energy carrier to drive endergonic H_2 generation from a low-energy carrier. Comparative genomics shows that these strategies are prevalent among syntrophic metabolizers, making large steps forward in our understanding of syntrophic metabolism. Moreover, recent studies complement genome-derived knowledge with biochemical and proteomic experiments, leading to even more rigorous and accurate characterization (*8, 36, 37*).

In parallel with sequencing isolate genomes, advances in bioinformatics has permitted researchers to directly obtain genomic sequences for organisms in their natural habitat using the same sequencing technologies. In 2004, studies accomplished revolutionary acquisition of genomes for individual uncultivated microbial populations directly from acidophilic biofilms and open ocean samples simply using Sanger sequencing. They accomplished this by sequencing DNA extracted directly environmental samples (*38-40*), assembling an ecosystem genome comprised of many genomes of the innate microbial populations ("metagenome") (*38*), and phylogenetically classifying the genomic fragments to specific microorganisms ("binning") (*41*). Unlike assembly of isolate genomes, metagenomes contain mixed sequences from different populations at varying abundances that need to be differentiated. Thus, classical genome assembly tools (*e.g.*, Velvet, SOAPdenovo, and CLCBio) (*42-44*) were not sufficiently accurate. Over the past decade, bioinformaticians created assemblers that address these obstacles and generate high quality metagenomic

assemblies (*e.g.*, metaSPAdes and MEGAHIT) (*45-48*). In parallel, binning also significantly matured over the past decade. In early stages, binning depended on sequence similarity to genomes of known organisms and, thus, could only categorize sequences to known phylogenetic lineages. By 2007, a similarity-independent *de novo* "composition-based" method was proposed (*41*) based on the concept that genomes of different phylogenetic lineages have differentiable compositional signatures (*e.g.*, frequency of tetranucleotides) (*49*). This method accomplished categorization of assembled metagenomic sequences into bins with homogeneous compositional signatures and, thereby, recover genomes for uncultivated organisms without any reference genomes (*50-52*). To further improve the accuracy of binning, software now also integrate information from (a) essential single-copy genes (*e.g.*, MaxBin) and (b) multiple metagenomes of related samples (*e.g.*, MetaBAT and GroopM) (*53-56*). In addition, researchers can also now determine individual bins' phylogeny ("phylogenomics"; *e.g.*, PhyloPhlAn) (*57*) and genome completeness/contamination (*e.g.*, CheckM) (*58*) using marker genes. The combined advances enable accurate metagenomic assembly, binning, and ultimately recovery of genomes for phylogenetically novel organisms without cultivation (*52, 55, 59*).

Indeed, metagenomics has proven to be an incredibly effective tool in obtaining genomic for uncultivated organisms across various disciplines and ecosystems. In early implementations, studies recovered genomes for symbionts of gutless worms, tubeworms, and pathogenic fungi to reveal novel symbiotic behavior (*60-63*). In parallel, investigation of an anaerobic wastewater treatment system revealed that a novel candidate phylum member "*Candidatus* Cloacamonas" could potentially be an amino acid fermenter and syntrophic propionate degrader, both critical ecological roles (*64*). Reflecting advances in bioinformatics tools, recent analyses of groundwater microbial communities yield genomes for more than 30 candidate bacterial phyla that all encode fermentative metabolism and are curiously lacking many components of typical housekeeping enzymes or biosynthesis, suggesting novel ecological niches and evolutionary directions of anaerobic *Bacteria* (*59, 65*). Other studies even anomalously discover methanogen-like methane metabolism in uncultured archaeal lineages outside of *Euryarchaeota*, "*Ca.* Bathyarchaeota" and "*Ca.* Verstraetearchaeota" *(66, 67)*. Thus, there are clearly many holes in our knowledge that need to be filled to properly understand the ecology of microbial communities.

1.3 – Directions for progress in understanding syntrophs and microbial dark matter

For syntrophic metabolizers, catabolic pathways of several model substrates have been successfully characterized through cultivation, biochemical analyses, and gene/protein expression (*i.e.*, transcriptomics and proteomics), but how syntrophs degrade many other environmentally relevant metabolites (*e.g.*, amino acids) and pollutants (*e.g.*, aromatic compounds) remain unclear. In addition, many syntrophic

metabolizer-related lineages in *Deltaproteobacteria* and *Firmicutes* have yet to be cultured (*19, 68-70*), especially those related to *Syntrophus*, *Smithella*, *Syntrophobacter*, *Syntrophorhabdus*, and *Pelotomaculum* (Fig. 1.1). Incidentally, most of these genera can degrade petroleum-related compounds (*i.e.*, alkanes and aromatic acids), indicating there remains much to be learned about catabolism of such compounds from the uncultured lineages. To overcome the obstacle of cultivation, metagenomics is an effective tool to acquire genomic insight into these lineages. Moreover, pairing this approach with whole-ecosystem gene expression analyses (*i.e.*, "metatranscriptomics") allows us to inspect syntroph metabolizers' behavior *in situ*, a critical step forward for truly comprehending their ecological contribution.

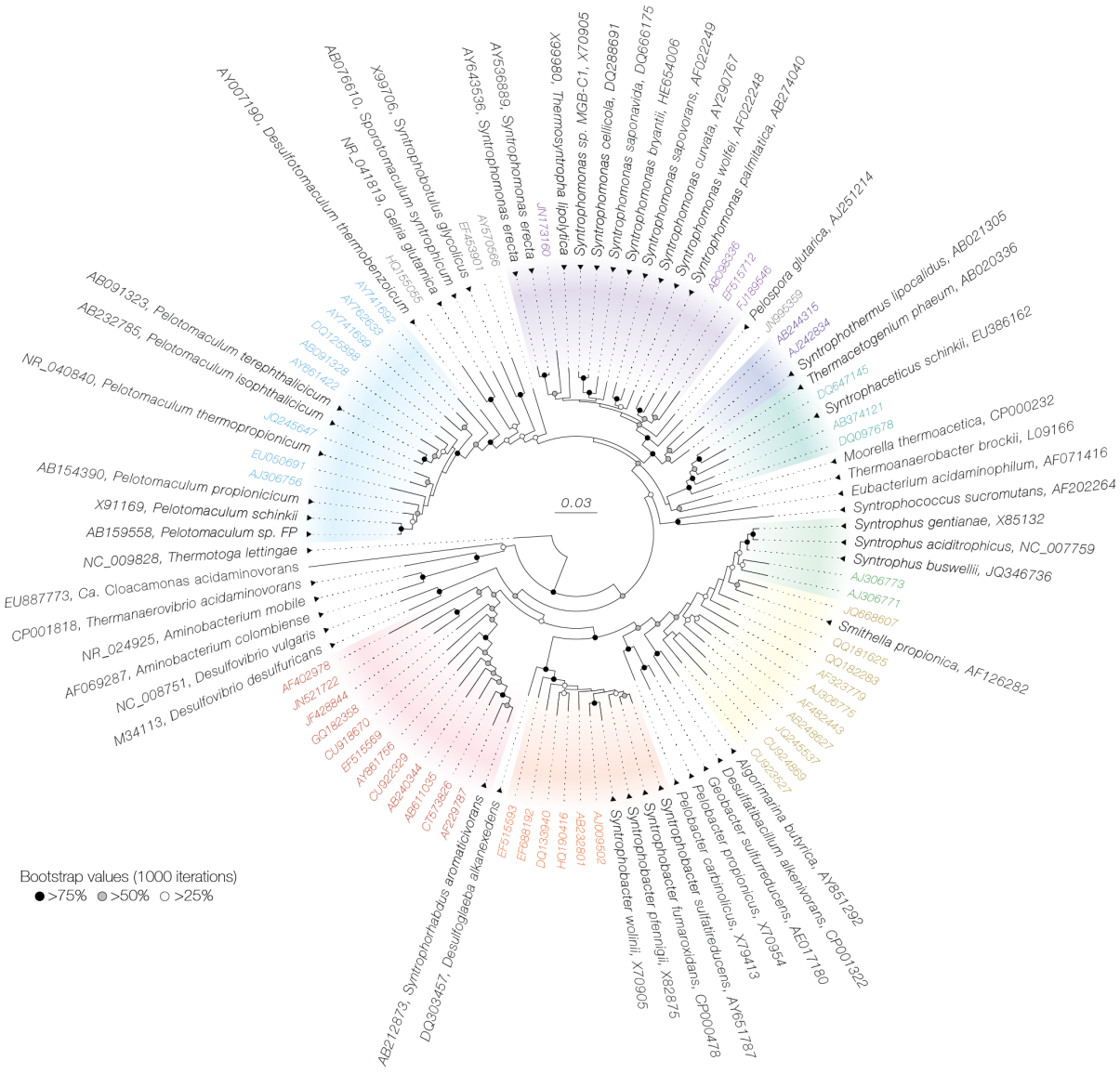

Figure 1.1. Phylogenetic tree of isolated *Bacteroidia* capable of syntrophy and representatives of phylogenetically related uncultured lineages affiliated with methanogenic environments. All representatives form clear phylogenetic clusters with syntrophic isolates distinct from other genera. The bootstrap values of 1000 iterations are shown as circles (black >75%, gray >50%, and white >25%) at the corresponding branches.

As for microbial dark matter lineages, previous studies have thus far only obtained genomic insight for one third of the candidate phyla affiliated with methanogenic environments, leaving 22 unaddressed (Fig. 1.2). Even for candidate phyla with genomes available, only several genomes have been acquired in most cases. Given the functional diversity of known phyla (*e.g.*, *Proteobacteria*, *Firmicutes*, and *Bacteroidetes*), it is reasonable to infer that the majority of such phyla's functional capacity remains unknown. In addition, we believe that genome-guided prediction of the metabolic capacity for uncultured organisms requires greater finesse than applied to previous studies. For organisms without cultivation and/or enrichment, we ought to take a cautious approach to avoid misguided interpretation and reports because often little non-bioinformatics-based background information is available to guide the analysis. This is especially critical for genomic prediction of organisms affiliated with anaerobic environments due to the aforementioned complexity of anaerobic metabolism.

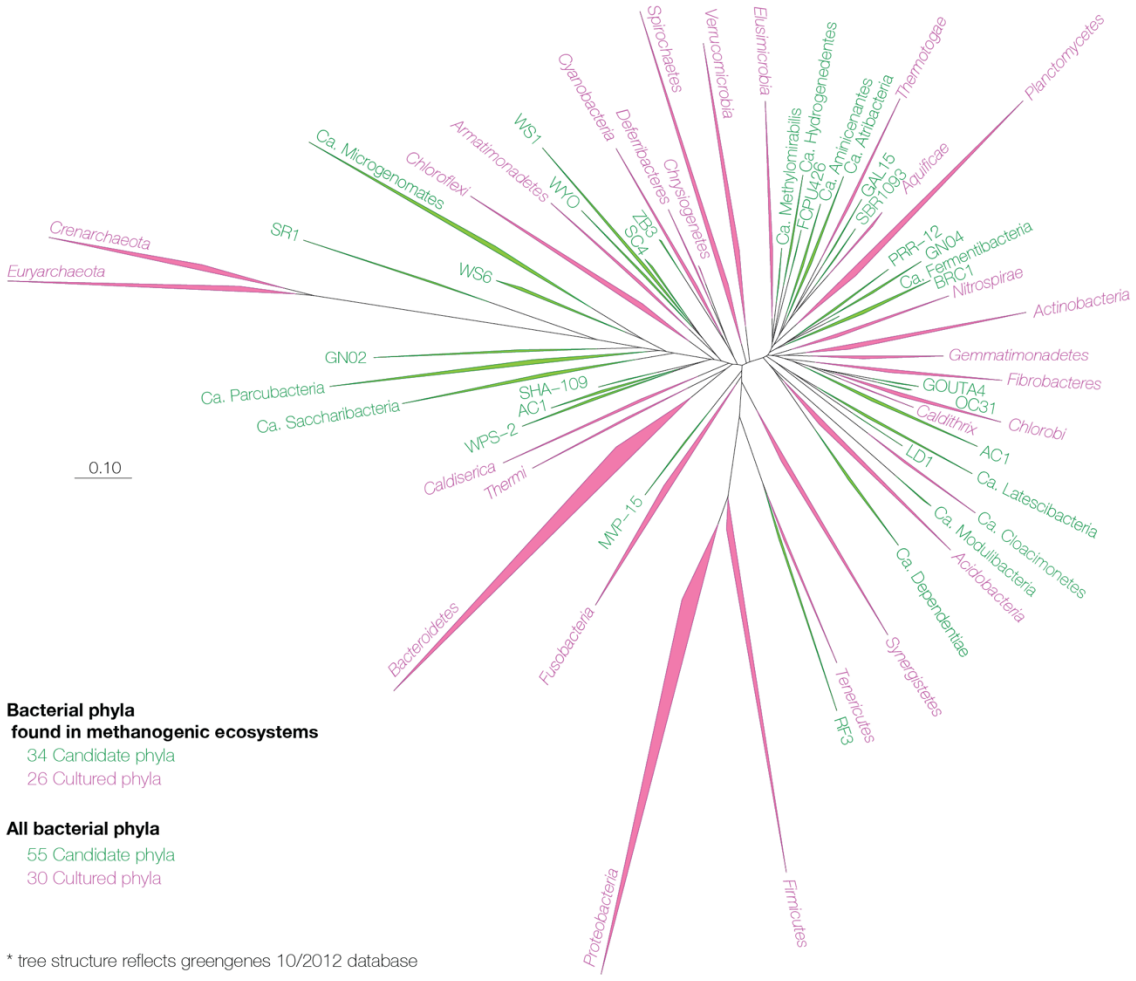

Figure 1.2. Phylum-level phylogenetic tree of all publically deposited 16S ribosomal RNA sequences associated with anaerobic environments. The tree structure is based on the Greengenes 10/2012 ARB database. It should be noted that several parts of the structure are not necessarily correct due to the massive size of the database.

For organisms respiring favorable electron acceptors (*e.g.*, aerobes and denitrifiers), typical physiological electron carriers used during substrate oxidation (*e.g.*, NAD(H) and quinone) can directly interact with the electron transport chain and generate energy; thus, as long as the catabolic pathway for a compound is present in the genome, one can fairly deduce that the organism likely catabolizes it (certainly not guaranteed to be true though!). On the other hand, under methanogenic conditions without favorable electron acceptors, identification of a catabolic pathway of a compound does not necessarily warrant inference of metabolic capacity to utilize the corresponding chemical as a substrate. As discussed above for syntrophs, typical physiological electron carriers cannot necessarily directly donate electrons into anaerobic electron disposal (*i.e.*, fermentation, hydrogen production, or formate generation); anaerobic metabolism is much more complex due to the necessity to employ specific enzymes that allow electron transduction from catabolism to electron disposal, including the aforementioned reverse electron transport and electron confurcation. In addition, references for annotation of genes involved in all such steps is rather limited due to difficulty in biochemical characterization of anaerobic enzymes, suggesting we ought to take caution in annotating the presence or absence of particular steps. Finally, the combined catabolism, electron transduction, and electron disposal must have a net positive yield of ATP and/or PMF to create a coherent energy-yielding pathway. Thus, complementing metagenomics with rigorous metabolic annotation following this logic is absolutely critical to predict the ecological contribution of uncultivated phyla in methanogenic environments. Of course, the same rationale applies to annotation of any novel syntroph genomes.

1.4 – Methanogenic wastewater treatment processes: ideal ecosystems for omics-based investigation

Another difficulty in studying uncultivated syntrophs and microbial dark matter phyla *in situ* is that their habitats tend to be challenging to access, temporally unstable, spatially heterogeneous, uncontrolled, and low-activity ecosystems (*e.g.*, groundwater and oil reservoir) or too diverse in some cases (*e.g.*, anaerobic soils). With low and unstable activity, it is challenging to obtain sufficient active biomass for metagenomics, metatranscriptomics, and microcosm-based experiments. In contrast, methanogenic wastewater treatment processes (MWT) often harbor similar syntrophs and candidate phyla and are also easily accessible, temporally stable, relatively homogeneous, well-controlled, high-activity, and have moderate diversity appropriate for omics-based investigations. More specifically, methanogenic bioreactors host half of the candidate phyla known to inhabit methanogenic ecosystems (Fig 1.3). In addition, we observe many uncharacterized lineages (family- or class-level) of cultured phyla affiliated with MWT, indicating that investigation of bioreactors could even expand our understanding of cultured phyla that are thought be relatively well-defined, especially *Euryarchaeota* and *Bacteroidetes*.

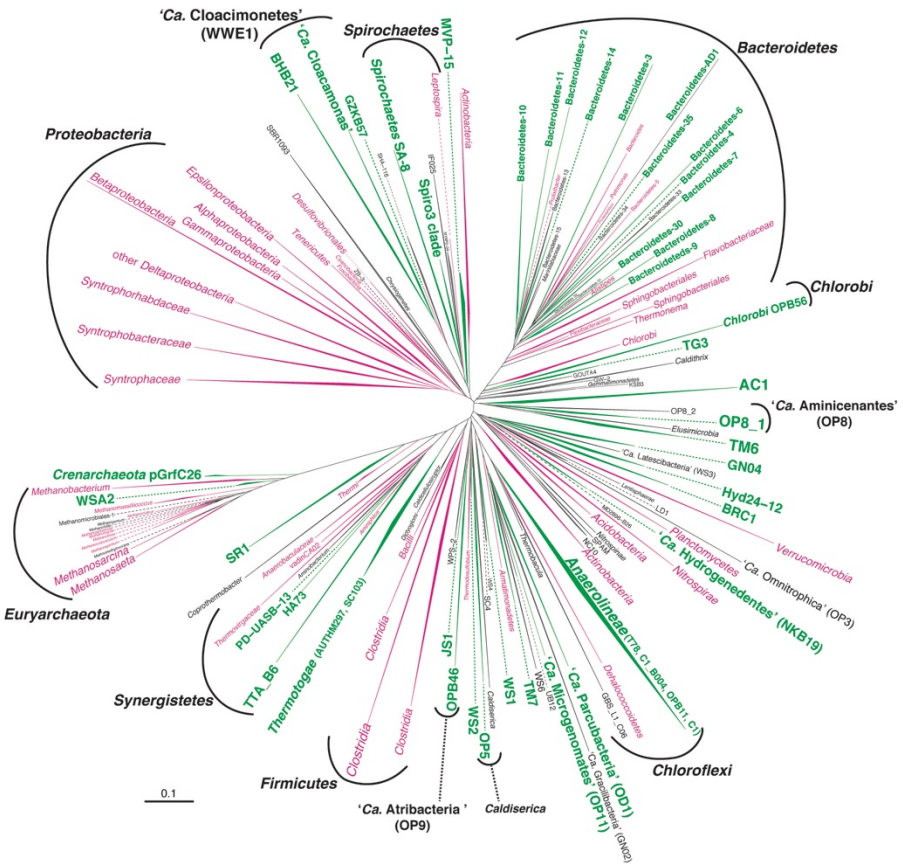

Figure 1.3. Phylogenetic tree of cultured (pink) and uncultured (green) clades associated with methanogenic bioreactors.

Besides characterizing uncultivated prokaryotic lineages, why study MWT? As the world's population and resource demand rise, our society continues to heavily impact Earth's biogeochemical cycles by producing copious amounts of waste. In 2011, the world produced more than 950,000 tons of municipal wastewater containing nutrients and toxic chemicals that can harm our environment (OECD - Organization for Economic Co-operation and Development) (*71*), which does not even account for the global industrial waste. As part of our environmental stewardship, it is paramount to treat both municipal and industrial wastewater. Although the traditional aerobic activated sludge process can rapidly treat wastewater, these systems require high-energy input and generate a large amount of waste sludge that are economically and environmentally taxing. As an alternative approach, MWT can treat wastewater without the costly aeration and excess sludge production (by up to 90%) (*72*). Moreover, this process generates CH_4 that can be recovered as a sustainable energy source, whereby accomplishing conversion of waste to energy (Fig. 1.4). These attractive characteristics have led to global implementation of this technology for treatment of both municipal and industrial waste. For municipal wastewater treatment, employing activated sludge

process and MWT in series has become the method of choice. Optimization of MWT is essential for reducing our society's environmental impact and developing sustainable energy sources; thus, it is our mission to elucidate the ecological and biochemical contribution of uncultivated microbial populations in such systems and provide novel insight into both uncharacterized prokaryotic lineages and also unlocking the full potential of MWT microbiota.

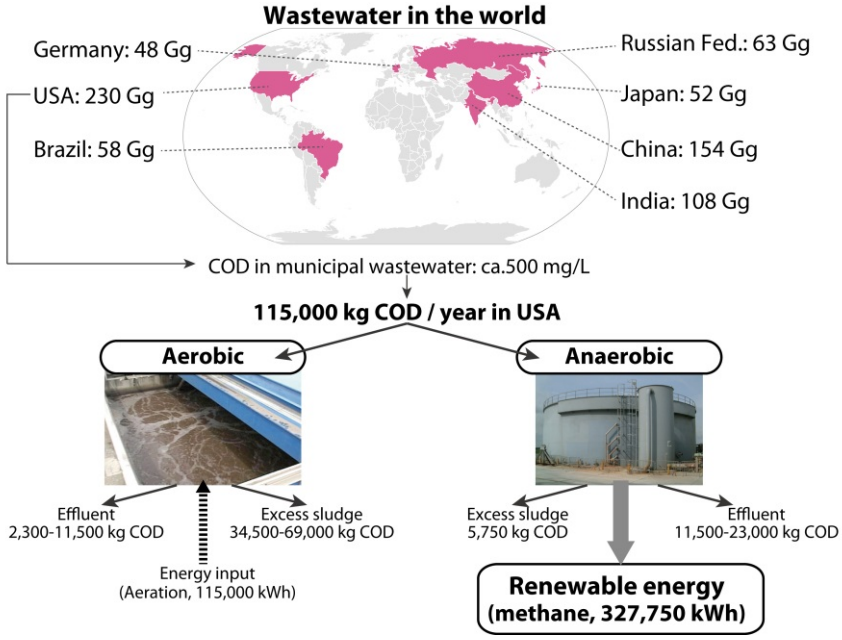

Figure 1.4. The top contributors for municipal waste and a comparison of aerobic and anaerobic processes. The conversion of waste mass to clean effluent (organics level represented as chemical oxygen demand – COD) and concurrent energy consumption and production are shown.

1.5 – Organization of this dissertation

This dissertation describes the omics-based elucidation of metabolic capacities and behavior of uncultivated syntrophs, methanogens, candidate phyla, *Bacteroidetes* affiliated with methanogenic bioreactors. Chapter 2 describes investigation of an isolated syntrophic metabolizer as foundational knowledge for characterizing syntrophic metabolism in complex ecosystems. Chapter 3 delves into characterization of the diverse catabolism and behavior of syntrophic aromatic compound degraders in bioreactors treating petrochemical waste. Chapter 4 reports the unique one-carbon metabolism of a novel methanogenic *Euryarchaeota* class WSA2 ubiquitous in MWT. Stepping beyond syntrophs and methanogens, Chapter 5 focuses on defining the ecological roles of uncultivated phyla supporting a bioreactor treating terephthalate. Chapter 6 expands the application of these studies' rigorous metabolic reconstruction to one of the most prevalent phyla across Earth, *Bacteroidetes*, to demonstrate the efficacy of such approaches in accurately predicting unique metabolic features with ecological impact.

CHAPTER 2

THE GENOME OF *SYNTROPHORHABDUS AROMATICIVORANS* STRAIN UI PROVIDES NEW INSIGHTS FOR SYNTROPHIC AROMATIC COMPOUND METABOLISM AND ELECTRON FLOW[1]

2.1 – Abstract

How aromatic compounds are degraded in various anaerobic ecosystems (e.g., groundwater, sediments, soils, and wastewater) is currently poorly understood. Under methanogenic conditions (i.e., groundwater and wastewater treatment), syntrophic metabolizers are known to play an important role. This study explored the draft genome of *Syntrophorhabdus aromaticivorans* strain UI, and identified the first syntrophic phenol-degrading phenylphosphate synthase (PpsAB) and phenylphosphate carboxylase (PpcABCD) and syntrophic terephthalate-degrading decarboxylase complexes. The strain UI genome also encodes benzoate degradation through hydration of the dienoyl-CoA intermediate as observed in *Geobacter metallireducens* and *Syntrophus aciditrophicus*. Strain UI possesses electron transfer flavoproteins, hydrogenases, and formate dehydrogenases essential for syntrophic metabolism. However, the biochemical mechanisms for electron transport between these H_2/formate-generating proteins and syntrophic substrate degradation remain unknown for many syntrophic metabolizers, including strain UI. Analysis of the strain UI genome revealed that heterodisulfide reductases (HdrABC), which are poorly understood electron transfer genes, may contribute to syntrophic H_2 and formate generation. The genome analysis further identified a putative ion-translocating ferredoxin:NADH oxidoreductase (IfoAB) that may interact with HdrABC and dissimilatory sulfite reductase gamma subunit (DsrC) to perform novel electron transfer mechanisms associated with syntrophic metabolism.

2.2 – Introduction

Aromatic compounds (AC) are often used as essential ingredients in industrial processes for the synthesis of various petrochemicals. During these industrial processes, AC are also generated as hazardous waste, which can lead to freshwater contamination through improper waste disposal, spills, and leaking storage tanks. Methanogenic consortia found in biological wastewater treatment and groundwater natural attenuation have been shown to effectively degrade AC (*73-75*). However, biodegradation of AC under

[1] This chapter appeared in its entirety in the published article: Nobu MK, Narihiro T, Hideyuki T, Qiu YL, Sekiguchi Y, Woyke T *et al* (2015). The genome of *Syntrophorhabdus aromaticivorans* strain UI provides new insights for syntrophic aromatic compound metabolism and electron flow. *Environ Microbiol* **17:** 4861-4872. doi: 10.1111/1462-2920.12444. As noted by *Environmental Microbiology*, authors retain the right to publish their paper in various medium/circumstances.

these conditions becomes thermodynamically unfavorable as the metabolic products (i.e., fatty acids and H_2) accumulate (*76*). To circumvent this issue, AC-degrading "syntrophic metabolizers" form partnerships with methanogens, which maintain the inhibitory metabolic byproducts at low concentrations to create thermodynamically favorable conditions for syntrophic substrate oxidation (*2, 13*). To degrade such syntrophic substrates, syntrophic metabolizers are reported to employ specific electron transfer mechanisms to help overcome the thermodynamic hurdle (*12*). However, much of the biochemical mechanisms involved in the syntrophic metabolism and microbial interaction still remains unclear. Thus, elucidating the metabolic strategies that syntrophic metabolizers employ to overcome thermodynamic limitations is critical to fully understanding anaerobic remediation of aromatic pollutants.

Several syntrophic AC metabolizers from *Pelotomaculum*, *Sporotomaculum*, *Syntrophorhabdus*, and *Syntrophus* of the classes *Clostridia* and *Deltaproteobacteria* have been isolated (*77-82*). Although previous studies have performed detailed characterization of *Syntrophus aciditrophicus* strain SB, the metabolism of other syntrophic AC-degraders remains unclear (*27, 83, 84*). The family *Syntrophorhabdaceae* is a highly environmentally relevant clade identified in a strikingly wide range of environments including mesophilic, hypermesophilic, and thermophilic bioreactors treating phenol and terephthalate (TA) (*69, 85-88*), a full-scale thermophilic hydrocarbon-degrading reactor (*75*), and hydrocarbon-contaminated groundwaters (*73, 74*) (Fig. 1a). *Syntrophorhabdus aromaticivorans* strain UI, a bioreactor-derived representative of this family, is a particularly versatile syntrophic metabolizer capable of degrading phenol, *p*-cresol, 4-hydroxybenzoate, isophthalate, and benzoate (*82, 88*). However, studying the metabolic and physiological traits of strain UI is difficult because it is extremely slow in growth (μ_{max} = 0.025 day^{-1} or doubling time = 20 days), and is an obligate syntrophic metabolizer that requires a partner for growth. Genomics analysis can be an effective alternative approach to characterize the potential physiology of slow-growing obligate syntrophic metabolizers. To do so, the genome of *S. aromaticivorans* strain UI was constructed through growing strain UI in co-culture with a known methanogen *Methanospirillum hungatei* strain JF-1, sequencing the composite co-culture "genome", and extracting the strain UI scaffolds through bioinformatics analysis (*89*).

In this study, the strain UI genome was further examined to advance our understanding of anaerobic AC degradation and syntrophic metabolism. Though previous studies on *S. aciditrophicus* strain SB have provided insight into benzoate metabolism (*27, 84, 90, 91*), metabolic pathways for syntrophic degradation of other AC (e.g., methylbenzoate and phenol, and phthalate) remain unclear. Likewise, the metabolic pathways necessary for syntrophic metabolizers to maintain a life style at the thermodynamic limit are poorly understood. Syntrophic metabolizers are known to utilize H^+ and HCO_3^- as electron acceptors to produce H_2 ($E^{0'}$ = -414 mV) and formate ($E^{0'}$ = -432 mV), respectively (*27*). However, using

NADH ($E^{0'}$ = -320 mV) as the primary electron carrier for those reductive reactions is thermodynamically inadequate, due to its relatively positive reduction potential. A recent study proposed that "reverse electron transport," energization of electrons on physiological electron carriers mediated by proton motive force, and "electron confurcation," oxidation of a high-energy electron carrier to drive electron transfer from a low-energy carrier to an unfavorable acceptor, are essential pathways for supporting syntrophic H_2/formate production (*12*). Still, genes that may perform such mechanisms have not been identified in many syntrophic metabolizers (*12*), suggesting the existence of other undiscovered pathways for electron flow. Elucidating such pathways can lead to better understanding of syntrophic AC metabolism. Strain UI is an ideal candidate for addressing these questions because it is an obligate syntrophic metabolizer and can degrade a wide-range of AC through syntrophic lifestyle.

2.3 – Methods

2.3.1 – Genomic DNA sequencing and annotation

This study analyzed the 3.76 Mbp *S. aromaticivorans* strain UI draft genome (GenBank accession: AJUN00000000.1) (*89*) composed of three linear scaffolds with an average G+C content of 52.0% and 3,691 genes, which were annotated using the Joint Genome Institute Integrated (JGI) Microbial Genomes (IMG) pipeline (*92*). As reported previously (*89*), genomic DNA was extracted and sequenced from a strain UI and *M. hungatei* strain JF-1 syntrophic co-culture using Illumina and 454 technologies (*93, 94*). The sequence data was assembled using Newbler and Phrap, curated using Illumina data with JGI Polisher (Alla Lapidus, unpublished), checked for mis-assemblies using gapResolution (Cliff Han, unpublished) and Dupfinisher (*95*), and further improved by closing gaps using polymerase chain reaction (PCR) and Bubble PCR (J-F Cheng, unpublished) with Consed (*96-98*) as previously described (*89*). Although this sequence pool originally contained both strain UI and JF-1 sequences, we confidently analyzed the published strain UI sequences as a pure draft genome because our previous study (*89*) rigorously identified and removed contaminant strain JF-1 sequences by performing best local alignment search tool (BLAST) analysis with a 100% identity cut-off against the strain JF-1 genome (Genbank accession: NC_007796.1) and also by further differentiating strain UI and JF-1 sequences based on read coverage (*99*), which was calculated by mapping raw Illumina and 454 reads to the contigs.

In order to improve and confirm the IMG-based annotation of the *S. aromaticivorans* strain UI genome, genes were individually analyzed outside of the IMG platform. BLAST and European Bioinformatics Institute (EBI) InterProScan were used to search for functional domains and characterize potential protein function (*100, 101*). The SABLE secondary structures and transmembrane domains prediction server was implemented to predict protein membrane association and secondary structure (*102*). As for non-coding

genes, clustered regularly interspaced short palindromic repeats (CRISPR) were identified and analyzed using the CRISPRs web server (*103*).

2.4 – Results and Discussion

2.4.1 – Genome comparison

Genome-wide BLAST of the *S. aromaticivorans* strain UI draft genome revealed top BLAST hits of genes to *Deltaproteobacteria* (51.5% of genes) and other syntroph-associated clades, including *Clostridia* (10.2%), *Nitrospiraceae* (2.7%), and *Synergistaceae* (0.6%) (Fig. 2.1B). Within *Deltaproteobacteria*, many genes relate to homologs in *Syntrophaceae* (16.9%), *Geobacteraceae* (12.7%), *Desulfobacteraceae* (10.0%), *Syntrophobacteraceae* (3.1%), and *Desulfovibrionaceae* (2.5%) homologs (Fig. 1b). Within *Clostridia*, many genes relate to homologs found in syntroph-associated families such as *Peptococcaceae* (4.3%), *Thermoanaerobacteraceae* (2.8%), *Clostridiaceae* (1.4%), and *Syntrophomonadaceae* (0.8%). Specifically, benzoate degradation genes relate to homologous genes encoded by AC-degrading syntrophic metabolizer families, *Peptococcaceae* and *Syntrophaceae* (Fig. 2.1B). Genes encoding syntrophic metabolism (electron carriers, hydrogenase, formate dehydrogenase, and cation translocation), pili/flagellar components, chemotaxis, and stress response had top BLAST hits to a wider phylogenetic range of syntrophic metabolizers (Fig. 2.1B). These observations suggest that strain UI may have experienced selective pressures similar to both closely and distantly related syntrophic metabolizers over its evolutionary history; however, based on this data alone, we cannot determine whether these genes were acquired through horizontal gene transfer between syntrophic metabolizer lineages or through independent convergent evolution.

Many bacterial genomes (47.7%) contain CRISPR regions as defense systems against the introduction of exogenous genetic materials (e.g., viral infection and horizontal gene transfer). Strain UI has four CRISPR regions (Table 2.1). The CRISPR spacers did not share significant homology with other microbial genomes, but the 70-spacer CRISPR repeat sequence shares high sequence homology (91%) with syntrophic metabolizers in *Deltaproteobacteria* (*S. aciditrophicus* strain SB and *Desulfovibrio vulgaris* strain Hildenborough) and moderate homology (79-82%) with *Clostridia* syntrophic metabolizers (*Pelotomaculum thermopropionicum* strain SI, *Syntrophomonas wolfei* strain Goettingen, and *Syntrophobotulus glycolicus* strain DSM 8271). The evolutionary and ecological implications of these observations remain unclear due to the limited understanding of CRISPR evolution.

2.4.2 – Central metabolism and general physiology

The strain UI genome encodes the Embden Meyerhoff pathway, tricarboxylic acid cycle, and pentose phosphate pathway for central metabolism and biosynthesis (Appendix A Table A.1). The genome

also possesses genes for complementary anaplerotic reactions (i.e., phosphoenol pyruvate carboxykinase and fructose 1,6-bisphosphatase). Genes for biosynthesis of amino acids and most co-factors are also present. Although strain UI is non-motile, genes encoding flagellum components, fibril-associated pilus (Flp), type IV pilus (Tfp) synthesis, and chemotaxis were identified (Appendix A Table A.1).

Core bacterial RNA polymerase subunits (RpoABCZ) with sigma factors including general housekeeping 70 factors (RpoD), heat shock 32 factors (RpoH), a flagella biogenesis factor (FliA), and a 54 factor (RpoN) were detected (Appendix A Table A.1). For coping with shifts in environmental conditions, strain UI possesses thioredoxins, catalases, rubrerythrins, and superoxide dismutases (Table 7.1). The strain UI genome also contains genes for 12 universal stress proteins (UspA), 6 heat shock proteins (e.g., HrcA and GrpE), and 4 cold shock proteins (Appendix A Table A.1).

2.4.3 – Degradation of aromatic compounds to Benzoyl-CoA

Anaerobic bacteria are known to degrade various AC to a central intermediate, benzoyl-CoA, for downstream reductive ring dearomatization (*104*). For benzoate degradation to benzoyl-CoA, strain UI possesses a benzoate-CoA ligase (SynarDRAFT_1026). The strain UI genome also harbors genes for degrading 4-hydroxybenzoate, phenol, and TA to benzoyl-CoA. For conversion of 4-hydroxybenzoate, this genome encodes putative 4-hydroxybenzoate CoA ligases (SynarDRAFT_1026 and 2749) and 4-hydroxybenzoyl-CoA reductase cassettes (HcrBCA: SynarDRAFT_1104-1102, HcrACB: 2331-2329) similar to *Thauera aromatica* strain K172, a known 4-hydroxybenzoate- and phenol-degrading denitrifier (*105*) (Fig. 2.2 and Appendix A Table A.3). For phenol catabolism, strain UI possesses a phenylphosphate synthase (PpsAB) (SynarDRAFT_2928-2927) and a phenylphosphate carboxylase complex (PpcBDAC) (SynarDRAFT_2921-2917). These complexes perform ATP-mediated phenol activation followed by *para*-carboxylation to yield 4-hydroxybenzoate for downstream degradation to benzoyl-CoA, as has been observed in *T. aromatica* (*106, 107*) (Fig. 2). Both the *pps* and *ppc* cassettes share high sequence similarities (>40%) and gene organizations with *T. aromatica* (*106, 107*) (Appendix A Table A.3). These observations suggest that strain UI invests scarce ATP into phenol activation rather than performing ATP-independent phenol degradation as proposed for *Sedimentibacter hydroxybenzoicus* strain JW/Z-1 (*108, 109*) (Fig. 2.2).

Previous studies proposed that *Pelotomaculum*-related syntrophic metabolizers might initiate TA degradation through decarboxylation to benzoate using candidate TA decarboxylase genes (PpcA-like and UbiD-like decarboxylases) (tadcc16349 and tadcc27178) identified from a TA-degrading bioreactor metagenome (*87, 110*). The strain UI genome possesses one aromatic decarboxylase (SynarDRAFT_0373) homologous to the PpcA-like TA decarboxylase (Appendix A Table A.3). In the strain UI and *Pelotomaculum* genomes, the PpcA-like decarboxylases associate with identical genes cassettes that encode

a PpcA-related aryl decarboxylase, PpcC-related aromatic decarboxylase small subunit, aryl CoA ligase, and BenE-related aryl transporter (SynarDRAFT_0373-0376 and tadcc16349-16346) (Appendix A Table A.2). These genes in strain UI and *Pelotomaculum* share much higher similarity to each other (61%, 44%, 51%, and 85% in order of gene organization) than to any top BLAST hit (40%, 34%, 44%, and 41% respectively). Given that these genes have exclusively high homology and that strain UI and *Pelotomaculum* spp. are the only known bacteria capable of degrading TA syntrophically, it is reasonable to speculate that these PpcAC-related and BenE-related genes encode a TA decarboxylase complex and TA transporter, respectively. The presence of an aryl CoA ligase gene in the operon suggests that syntrophic TA degradation may proceed through TA activation to terephthalyl-CoA and subsequent decarboxylation to benzoyl-CoA, rather than direct TA decarboxylation to benzoate (Appendix A Fig. A.2). For the latter case, strain UI may further metabolize benzoate to benzoyl-CoA using a benzoate-CoA ligase. Further biochemical investigation is necessary to determine the exact pathway for TA degradation (Appendix A Fig. A.2).

2.4.4 – Benzoyl-CoA metabolism

Anaerobic benzoyl-CoA degradation begins with highly endergonic reduction of benzoyl-CoA to a dienoyl-CoA intermediate (*111*). Amongst anaerobic AC-degrading organisms, *T. aromatica* and *Rhodopseudomonas palustris* are known to drive this reaction using ATP (*112, 113*), while it is unclear how *S. aciditrophicus* strain SB or *Geobacter metallireducens* strain GS-15 drive the reduction (*91, 111, 114, 115*). The strain UI genome encodes non-ATP-dependent benzoyl-CoA reductase (BCR) operons (SynarDRAFT_0938-0933 and SynarDRAFT_3490-3491) closely related to *S. aciditrophicus* strain SB and *G. metallireducens* strain GS-15 BamY, BamB-F, and BamGH (Fig. 2.2 and Appendix A Table A.3) (*27, 91*). Membrane potential and electron bifurcation have been proposed for this BCR as an alternative energy input (*91, 114, 115*). Likely, the BCR-associated heterodisulfide reductase (Hdr) and hydrogenase genes participate in this endergonic reaction, which is discussed in a latter section. To further degrade the dienoyl-CoA intermediate to fatty acids, H_2, and CO_2, strain UI encodes genes involving in the hydration of the dienoyl-CoA intermediate (SynarDRAFT_1236), ring oxidation/hydrolysis (SynarDRAFT_1235-1234), and final β-oxidation (SynarDRAFT_1515-1511 and SynarDRAFT_3464-3465) (Fig. 2.2 and Appendix A Table A.2) (*84*).

Besides H_2 and CO_2, syntrophic benzoate metabolism is known to yield acetate as a final product (i.e., *S. aciditrophicus*), and possibly butyrate as a transient product (i.e., *Sporotomaculum syntrophicum*) (*79*). Although syntrophic co-cultures of strain UI have only been observed to produce acetate (*82*), strain UI possesses genes for producing acetate and butyrate through respective β-oxidation and *Clostridium*-like butyrate dehydrogenase complex (butyryl-CoA dehydrogenase: SynarDRAFT_1017, electron transfer

flavoprotein subunits A/B: SynarDRAFT_3463-3462) (Appendix A Fig. 2.2 and Table 7.2). Strain UI could not use butyrate as a syntrophic growth substrate (*82*), so it can be speculated that strain UI may indeed utilize the butyryl-CoA dehydrogenase for butyrate production. Producing butyrate rather than two acetate sacrifices one ATP that could be generated through acetyl-CoA-driven substrate-level phosphorylation (Fig. 2.2). On the other hand, butyrate production can theoretically allow strain UI to efficiently oxidize NADH through *Clostridium*-like butyryl-CoA dehydrogenase electron bifurcation. Specifically, this complex performs exergonic electron transfer from NADH to crotonyl-CoA ($E^{0'}$ = -126 mV) to drive parallel endergonic NADH-oxidizing generation of reduced ferredoxin (Fd_{red}), a high-energy electron donor (midpoint potential = -453 mV) (*116*). This NADH sink would be particularly valuable when environmental H_2 accumulates and respiratory NAD^+ regeneration becomes increasingly thermodynamically difficult. In addition, generating Fd_{red} would allow strain UI to perform exergonic H^+ respiration (discussed in detail in the following section). However, further transcriptomic and/or proteomic investigation is necessary to characterize the speculated pathway.

2.4.5 – Energy conservation and electron flow

Anaerobic organisms that respire low-energy electron acceptors such as H^+ and CO_2 often encounter thermodynamic hurdles for electron flow (i.e., regeneration of oxidized electron carriers and disposal of electrons). Simply based on thermodynamic calculations, most physiological electron carriers (e.g., ETF_{red}, $FADH_2$, and NADH) have relatively positive reduction potentials that make them thermodynamically inadequate donors for syntrophic and methanogenic respiration (i.e., H^+ reduction to H_2 and CO_2 reduction to CH_4). To achieve these reductive reactions, organisms employ Fd_{red} as an alternative donor. However, in many cases, production of such a high-energy electron carrier requires energy conservation mechanisms, such as reverse electron transport and electron bifurcation. For H_2-producing organisms (e.g., syntrophic metabolizers), the *Rhodobacter* nitrogen fixation (Rnf) complex uses transmembrane proton motive force to drive endergonic reverse electron transport from NADH to Fd_{ox} (*27, 117, 118*) (Fig. 2.3). For H_2-oxidizing organisms (e.g., methanogens), electron bifurcating hydrogenases can perform exergonic H_2-oxidizing NAD^+ reduction to drive parallel endergonic H_2-oxidizing Fd_{red} generation (*119, 120*). Besides NAD^+, a thiol-disulfide redox pair (Coenzyme M-Coenzyme B) has been shown to serve as an exergonic electron acceptor for H_2-oxidizing electron bifurcation by a methanogen complex composed of a three-subunit heterodisulfide reductase (HdrABC) and hydrogenase (*116, 121, 122*). Syntrophic metabolizers and other H_2-producing organisms are thought to run electron bifurcation in the reverse direction for electron confurcation (*12, 116*). For example, a *Thermotoga maritima* electron-confurcating hydrogenase complex performs parallel exergonic Fd_{red} oxidation and endergonic NADH

oxidation to generate H_2 (*120, 123*) (Fig. 2.3). This reaction allows syntrophic metabolizers to harness the energy margin from exergonic Fd_{red} oxidation for regenerating NAD^+. However, energy conservation and electron flow mechanisms remain unclear for the majority of syntrophic metabolizers, including strain UI.

Based on the AC degradation genes identified in strain UI, its metabolism primarily yields reduced electron transfer flavoproteins (ETF) and NADH as reduced electron carriers. For using ETF as an electron carrier, the strain UI genome has two cassettes containing genes encoding ETF subunits (*fixAB*). One cassette encodes FixAB, ETF dehydrogenase (FixC) and ferredoxin (FixX) (SynarDRAFT_2457-2460) (Fig. 2.3). The *fixABCX* operon is found in many syntrophic metabolizer genomes and is thought to perform quinone-mediated reverse electron transport from ETF_{red} to a membrane-bound hydrogenase for H_2 generation (*35*). The other ETF-related cassette (SynarDRAFT_1013-1003) encodes FixAB with various electron-transfer proteins, Fd, hydrogenase, Fe-S oxidoreductase, HdrABC, and formate dehydrogenase (Appendix A Table A.2). This suggests that strain UI may use an intricate electron transfer scheme involving ETF with Fd and an uncharacterized thiol-disulfide redox pair for H_2/formate generation, which is detailed later in this section.

As for NADH, syntrophic metabolizers are thought to circumvent endergonic NADH-driven H_2/formate generation by transferring electrons from NADH to Fd (Fig. 2.3) (*12, 27, 124*). The strain UI genome encodes both Fd_{red}-dependent H_2 and formate production. For formate, strain UI possesses an Fd-dependent formate dehydrogenase (SynarDRAFT_1006) and corresponding activators (SynarDRAFT_0932-0931) (Fig. 2.3). These formate dehydrogenase genes are associated with the benzoyl-CoA reductase operon (*bamBCDEF*), suggesting that formate production may be linked to benzoyl-CoA metabolism (Table S2). For H_2 production, the strain UI genome harbors a *T. maritima*-like Fd/NADH-dependent electron-confurcating hydrogenase (SynarDRAFT_0747-0749) (Fig. 2.3). However, like many other syntrophic metabolizers, the strain UI genome does not encode the Rnf complex necessary for generating Fd_{red} from NADH, and an alternative reverse electron transport mechanism for producing Fd_{red} can be used.

Genomic comparison of strain UI and other syntrophic metabolizers revealed a gene cassette (SynarDRAFT_0715-0709) exclusively conserved amongst syntrophic metabolizers lacking the *rnf* operon, except *Desulfovibrio* spp., which possess both operons (Fig. 2.4B and Appendix A Table A.3). This cassette encodes HdrABC, hydrogenase subunits, and two putative oxidoreductases. These two neighboring oxidoreductases encode protein domains indicative of Fd:NAD(H) oxidoreductase (alphahelical ferredoxin, Fe-S cluster binding site, and NADH oxidoreductase) and transmembrane transport (7-strand β-barrel) activity (Fig. 2.4A). Interestingly, similar protein domains were also found in RnfC, a membrane-associated subunit essential for Rnf activity (*125, 126*). Furthermore, in previously published syntrophic metabolizer

transcriptomes studies (*127-129*), *D. vulgaris* strain Hildenborough was found to express this gene cassette and not the *rnf* operon when grown in a lactate-degrading syntrophic co-culture with a hydrogenotrophic methanogen. These observations strongly suggest that these genes (SynarDRAFT_0710-0709) encode a putative ion-translocating Fd:NADH oxidoreductase (IfoAB) involved in novel syntrophic reverse electron transport and perhaps electron transfer through HdrABC and hydrogenase.

In the strain UI genome, BCR subunits, ETF, formate dehydrogenase, hydrogenase, and IfoAB genes are found in gene cassettes encoding Hdr subunits (Appendix A Table A.2). This indicates that Hdr may participate in syntrophic metabolism, like *D. vulgaris* (*127*). Hdr are known to exchange electrons with thiol-disulfide redox pairs ($E^{0'} \approx$ -250 mV), such as CoM-CoB ($E^{0'}$ = -140 mV) in methanogens and possibly DsrC (dissimilarity sulfite reductase subunit C) in *D. vulgaris* (Fig. 2.5) (*127, 130*). Thus, strain UI and other syntrophic metabolizers may possess an analogous thiol-disulfide electron carrier. The BCR subunits, including an HdrA-homolog (BamE), may pair endergonic benzoyl-CoA reduction with exergonic disulfide carrier reduction through Fd_{red}-oxidizing electron bifurcation ($2Fd_{red}$ + R-S-S-R' + benzoyl-CoA = $2Fd_{ox}$ + R-SH + R'-SH + dienoyl-CoA) (Fig. 2.5). The existence of an uncharacterized BCR disulfide electron acceptor is plausible because the purified BCR could not utilize known physiological electron acceptors (*115*). Furthermore, to reoxidize the thiol generated from benzoyl-CoA reduction, strain UI, *S. aciditrophicus* strain SB, and *G. metallireducens* strain GS-15 all encode HdrABC genes. Specifically for strain UI, the IfoAB-associated HdrABC and hydrogenase subunits may perform endergonic thiol-oxidation and exergonic Fd_{red}-oxidation to accomplish hydrogenogenic electron confurcation mechanistically similar to the *Methanothermobacter* electron-bifurcating hydrogenase (R-S-S-R' + Fd_{red} + $2H^+$ = RSH + R'SH + Fd_{ox} + H_2) (Fig. 2.5). Interestingly, DsrC genes were found in the genomes of strain UI (SynarDRAFT_2648 and SynarDRAFT_1247) and *P. thermopropionicum* strain SI (PTH_2856), who neither possess DsrAB nor use sulfate/sulfite as electron acceptors. Therefore, further molecular and biochemical investigations are necessary to elucidate the roles of HdrABC with regard to each associated cassette, identify the Hdr-associated novel electron carrier, and clarify the function of the DsrC gene in non-sulfate-reducing organisms.

2.5 – Conclusion

In conclusion, examining the strain UI draft genome revealed essential insight into novel syntrophic AC degradation and reverse electron transport mechanisms. Strain UI possesses syntroph-associated genes responsible for 4-hydroxybenzoate, phenol, TA, and benzoate degradation. Like many other syntrophic metabolizers, strain UI lacks the *rnf*-operon for reverse electron transport to convert metabolically generated NADH to Fd_{red}. Exploring alternative energy conservation mechanisms, this study

discovered a gene cassette mostly found in Rnf-lacking syntrophic metabolizers that putatively encodes reverse electron transport from NADH to Fd, and evidence that HdrABC may be involved in syntrophic benzoate degradation and energy conservation of strain UI.

2.6 – Figures

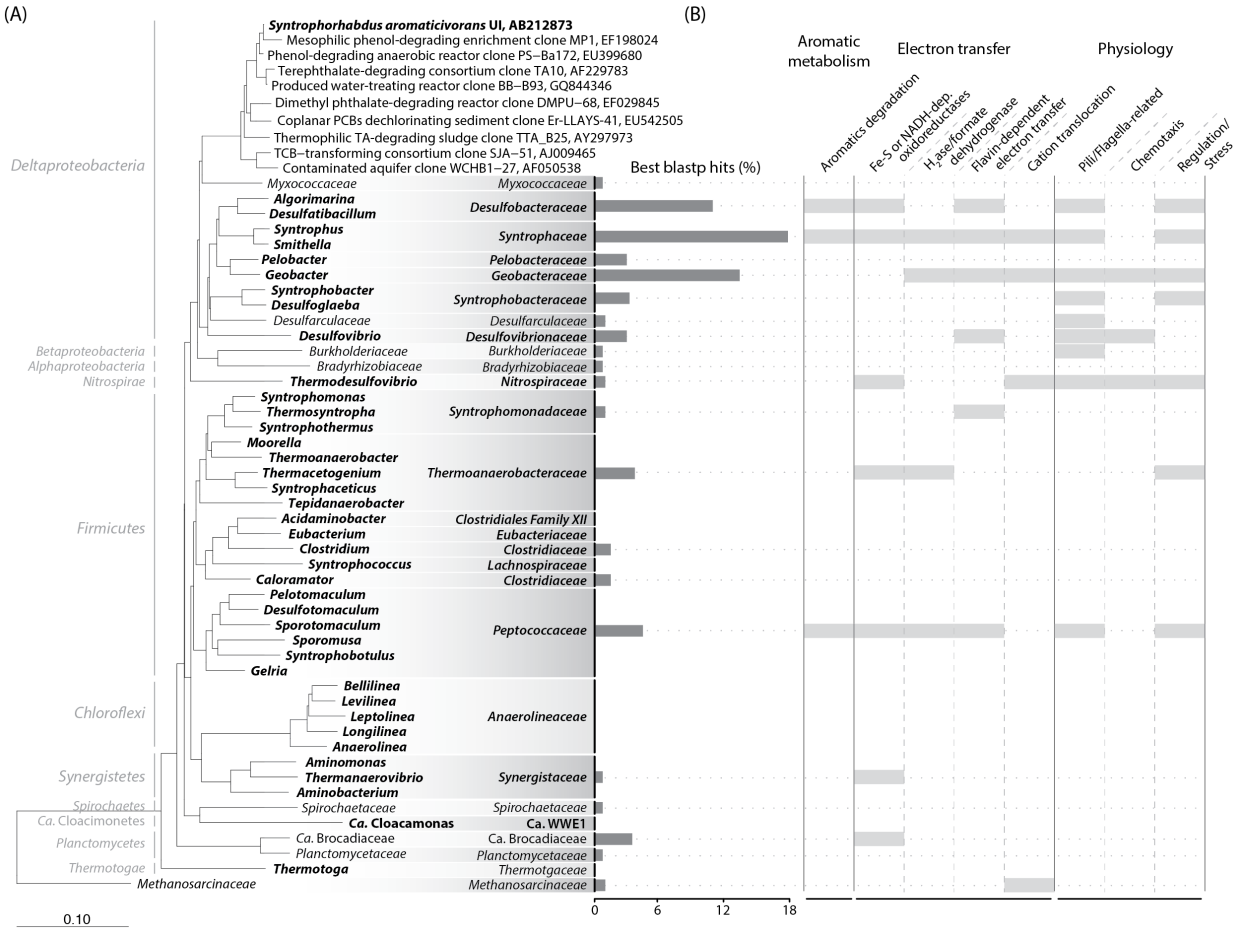

Figure 2.1. Relationship of *Syntrophorhabdus aromaticivorans* UI to other bacterial clades containing syntrophic metabolizers (a) phylogenetically and (b) genomically by best BLAST hits. Bacterial clades containing known or hypothesized syntrophic metabolizers are bolded. (b) The distribution of strain UI genes' best BLAST hits is shown as percentage of total strain UI genome gene count. Further, the right-hand table denotes that the best BLAST hits to a particular bacterial family include genes with function related to aromatic metabolism (AC degradation), electron transfer/flow (Fe-S or NADH-dependent oxidoreductases, hydrogenases, formate dehydrogenases, electron transfer flavoproteins, and cation translocation), and physiology (appendages, chemotaxis, regulation, and stress response) with a filled in gray cell.

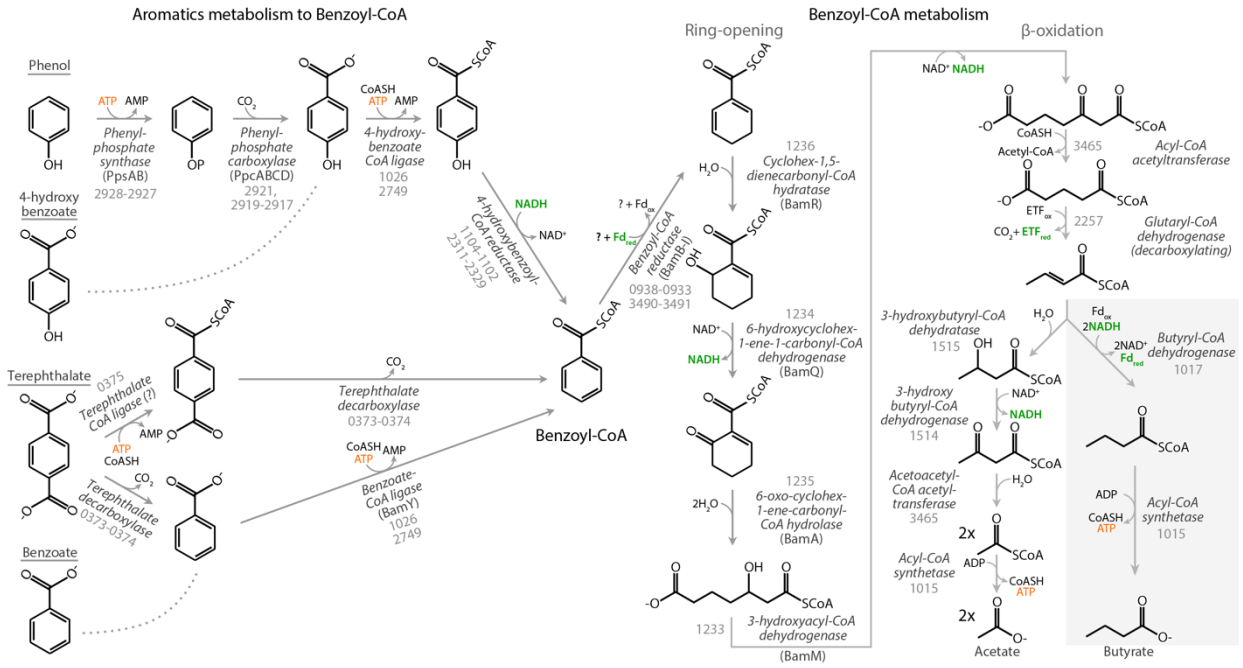

Figure 2.2. Strain UI aromatic compound metabolism to benzoyl-CoA and following benzoyl-CoA degradation pathway. Phenol, 4-hydroxybenzoate, terephthalate, and benzoate degradation to benzoyl-CoA is shown on the left. Terephthalate degradation has two potential pathways with terephthalyl-CoA or benzoate as an intermediate. Benzoyl-CoA degradation to acetate (white background) and possibly butyrate (gray background) are shown on the right side. Benzoyl-CoA degradation is split into reductive ring dearomatization/opening and β-oxidation. Each reaction is labeled with the abbreviated strain UI locus tag (e.g., "SynarDRAFT_0001" as "0001") and protein name (italicized) of the gene encoding the function.

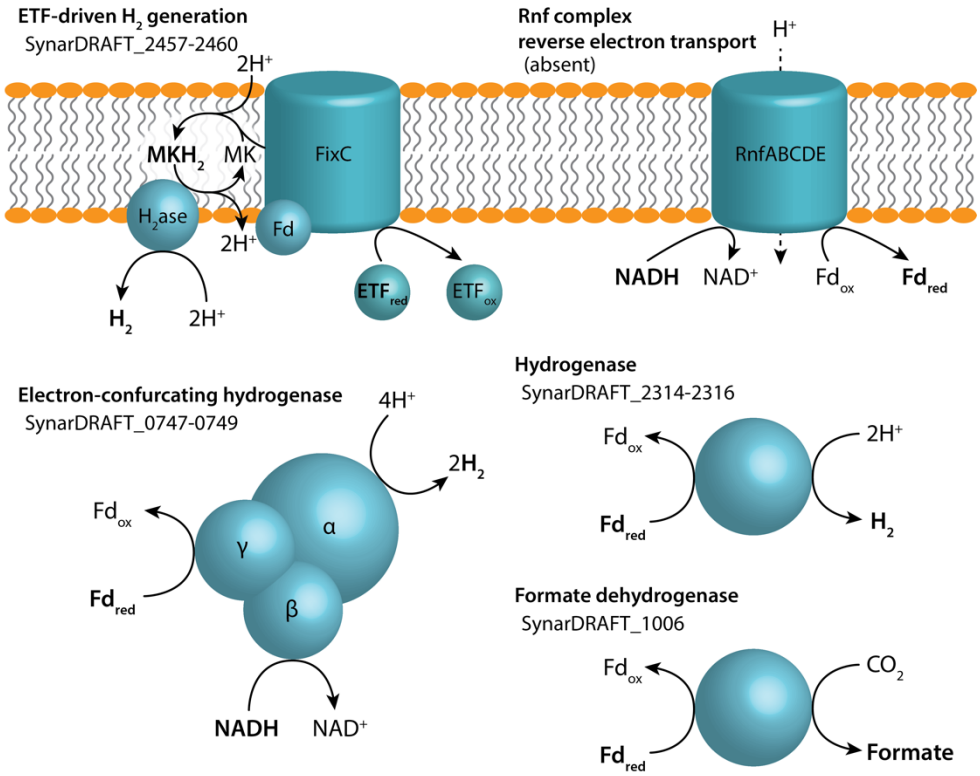

Figure 2.3. Known syntrophic pathways for oxidizing reduced electron carriers (NADH, ETF, Fd_{red}) and producing formate and H_2. ETF-driven H_2 generation transfers electrons from ETF_{red} to H^+, where menaquinone ($MK(H_2)$) serves as an intermediate electron carrier. The Rnf complex performs reverse electron transport from NADH to ferredoxin (Fd). The electron-confurcating hydrogenase concurrently performs Fd_{red} and NADH oxidation for H_2 generation, while formate dehydrogenase and hydrogenase only utilize Fd_{red} as an electron donor. The strain UI gene locus tags of each protein/complex are denoted.

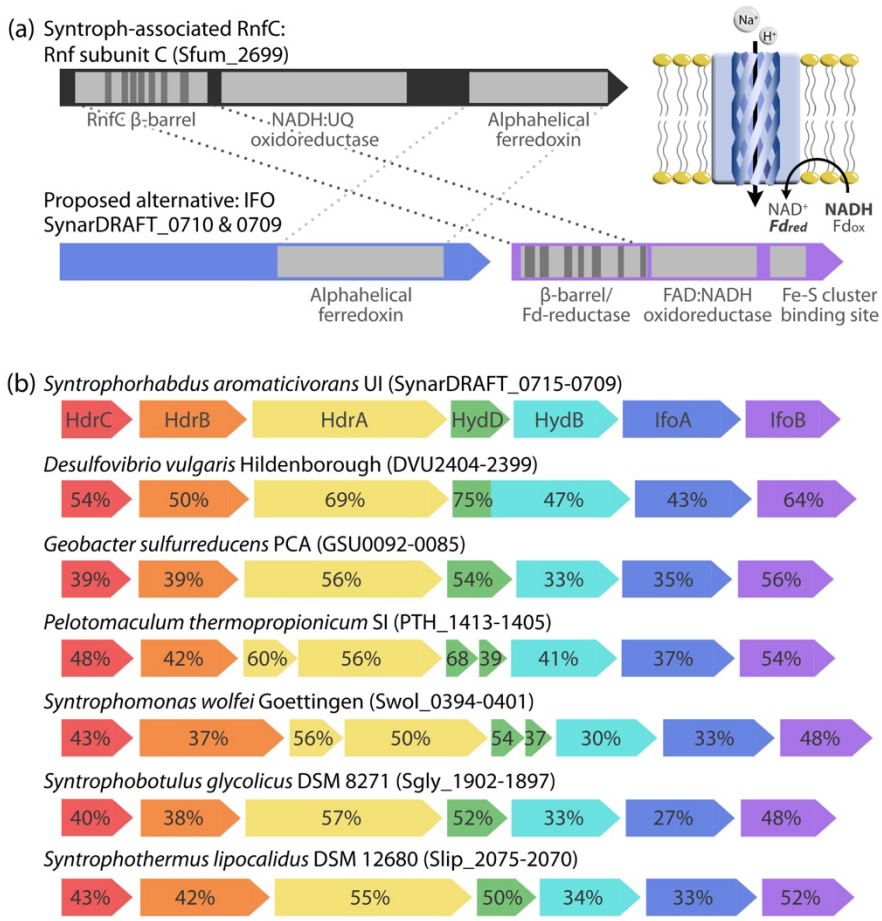

Figure 2.4. Comparison of (a) protein domains (gray rectangles) contained in *Syntrophobacter fumaroxidans* strain MPOB, *Rhodobacter* nitrogen fixation complex subunit RnfC and proposed ion-translocating ferredoxin oxidoreductase genes (IFO) identified in strain UI and (b) the IFO-associated gene cassettes found in many syntrophic metabolizers spanning various families. (a) The beta-strands identified in beta-barrel domains are shown (dark gray vertical bars). (b) The strain UI IFO-associated cassette encodes a heterodisulfide reductase complex (HdrCBA), hydrogenase subunits (HydDB), and proposed IFO subunits (IfoAB). Other syntrophic metabolizers' cassettes are denoted with peptide sequence percent similarity to the corresponding strain UI genes.

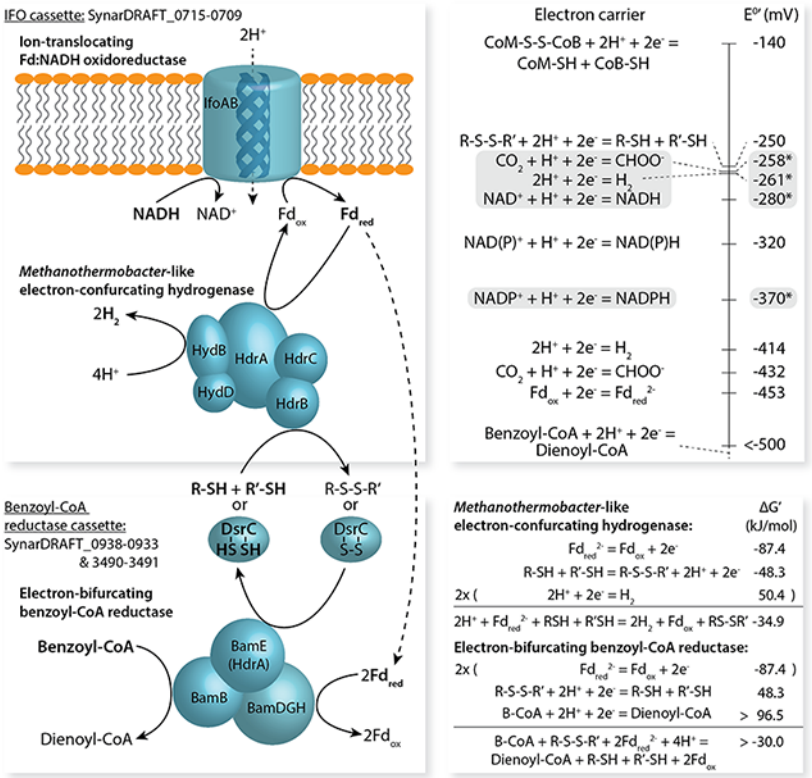

Figure 2.5. Hypothesized functions of heterodisulfide reductase subunits in strain UI syntrophic metabolism. The top-left panel shows the putative ion-translocating Fd:NADH oxidoreductase and *Methanothermobacter*-like electron-confurcating hydrogenase encoded in SynarDRAFT_0715-709. The bottom-left panel shows the benzoyl-CoA reductase complex encoded in SynarDRAFT_0938-0933 and 3490-3491. The electron-confurcating hydrogenase and benzoyl-CoA reductase complex may use a novel thiol-disulfide redox pair to conserve energy yielded from Fd_{red} oxidation and drive endergonic benzoyl-CoA reduction, for which thermodynamics calculations are shown in the bottom-right panel. DsrC may serve as the thiol-disulfide electron carrier. The top-right panel is an electron tower with syntroph- and energy-conservation- related redox pairs. *These E' values were calculated assuming non-standard conditions representing physiological concentrations of the respective redox pairs: 1μM formate, 1Pa hydrogen, 30:1 NAD^+:NADH, and 1:50 $NADP^+$:NADPH.

CHAPTER 3

THERMODYNAMICALLY DIVERSE SYNTROPHIC AROMATIC COMPOUND CATABOLISM

3.1 – Abstract

Specialized organotrophic *Bacteria* "syntrophs" and methanogenic *Archaea* "methanogens" form a unique metabolic interaction to accomplish cooperative mineralization of organic compounds to CH_4 and CO_2. Due to challenges in cultivation of syntrophs, mechanisms for how their organotrophic catabolism overcomes thermodynamic restrictions remains unclear. In this study, we investigate two communities hosting diverse syntrophic aromatic compound metabolizers (*Syntrophus*, *Syntrophorhabdus*, *Pelotomaculum*, and an uncultivated *Syntrophorhabdacaeae* member) to uncover their catabolic diversity and flexibility. Combined metagenomics and metatranscriptomics show that these syntrophs utilize unconventional alternative metabolic pathways producing butyrate, cyclohexanecarboxylate, and benzoate as catabolic byproducts and also diverse H_2 and formate generating pathways to facilitate interactions with partner methanogens. Based on thermodynamic calculations, these pathways that may enable syntrophs to combat thermodynamic restrictions. In addition, when fed with specific substrates (*i.e.*, benzoate, terephthalate, or trimellitate), each syntroph population expresses different pathways, suggesting ecological diversification among syntrophs. These findings suggest we may be drastically underestimating the biochemical capabilities, strategies, and diversity of syntrophic bacteria thriving at the thermodynamic limit.

3.2 – Introduction

As one of the most thermodynamically extreme modes of microbial respiration, many bacteria resort to using H^+ as an electron acceptor and produce H_2. Hydrogenogenesis rapidly becomes endergonic even with the slightest accumulation of H_2, so specialized "syntrophic" bacteria form a mutualistic interaction ("syntrophy") with H_2-oxidizing *Archaea* that reduce CO_2 to CH_4 (methanogens) to accomplish degradation of a wide variety of compounds (*131, 132*). Syntrophic metabolizers are essential components of the global carbon cycle's anaerobic sector; without them, complete mineralization of organic compounds would not be possible in a variety of natural and engineered environments (*2, 12, 133-135*). Thus, unraveling the rationale, diversity, and ecological impact of unique syntrophic metabolism is critical. However, due to challenges in cultivating these organisms, only a handful of syntrophic organisms have had genome sequencing, let alone thorough biochemical characterization, leaving the diversity and ecology of syntrophic metabolism largely unexplored.

Although syntrophs have been assumed to produce acetate, H_2, and formate as byproducts, a few studies discover deviant behavior – production of butyrate from propionate (*136-138*) and cyclohexanecarboxylate (CHC) from benzoate (*139*). We suspect that these deviant pathways make ecology revolving around syntrophy much more complex than anticipated. For instance, compared to conventional syntrophic propionate oxidation (*e.g.*, well-studied *Syntrophobacter*), a unique syntrophic propionate oxidizer (*i.e.*, *Smithella*) employs an atypical pathway (*136, 137*) to increase the thermodynamic favorability of propionate degradation (*i.e.*, lower $\Delta G^{o'}$), stoichiometrically halve H_2 generation (*138*), and theoretically catabolize propionate under environmental conditions that would otherwise shutdown the conventional pathway. Perhaps, metabolic diversity like this example may even provide an explanation for the redundancy of syntrophic metabolizers frequently observed in various ecosystems (*140-143*).

In this study, we specifically focus on the catabolic diversity of syntrophic aromatic acid degraders because they are phylogenetically diverse and often redundantly coexist in complex communities. Moreover, syntrophic aromatic compound degradation has been observed to generate butyrate and benzoate as byproducts (*79, 82*), but the mechanism and ecological relevance of such alternative metabolism remains to be uncovered. Thus, we employ metagenomics and metatranscriptomics to characterize the *in situ* catabolic behavior and diversity of such syntrophic metabolizers and deviant pathways during degradation of three model aromatic acids (benzoate – BA, 4-carboxybenzoate / terephthalate – TA, and 2,4-dicarboxybenzoate / trimellitate – TM) within complex petrochemical-degrading methanogenic microbial communities.

3.3 – Results and Discussion

We investigate two methanogenic communities (U1 and B) harboring diverse syntrophs: *Syntrophus* specialized in BA degradation (*81*), *Pelotomaculum* and *Syntrophorhabdus* capable of degrading diverse aromatic acids (*78, 82*), and a novel *Syntrophorhabdus*-related clade, tentatively referred to as "*Candidatus* Syntropharomatica." Metagenomics was used to recover bins for these organisms (Fig. 3.1; Appendix B Table B.1, Table B.2, and supplementary material) and predict their aromatic acid degradation capacity and complementary H_2- and formate- generating electron transduction pathways (Fig. 3.2; Appendix B Table B.3, and supplementary material). Through metatranscriptomics, we further estimate the activity of each syntrophic aromatic compound metabolizer during BA, TA, and TM amendment (Fig. 3.1; Appendix B Fig. B.1, Fig. B.2, Table B.4, and supplementary material), study the expression of pathways involved in syntrophic catabolism, and extrapolate each organisms' behavior during degradation of each substrate.

To determine which populations contribute to BA/TA/TM degradation, we identify organisms that (i) comprise at least 1% of the bacterial gene expression in the metatranscriptome and (ii) express genes involved in catabolism of the corresponding substrate at levels higher than 87.5% of other genes in their genome (*i.e.*, bin) (Appendix B Table B.4). Given this criteria, *Syntrophus* generally predominates BA degradation in both communities (Fig. 3.1 and Appendix B Table B.4). We also identify TA-degrading *Syntrophorhabdus* and *Pelotomaculum* populations that have both pronounced activity during TA amendment and high expression of a putative Na^+:TA symporter, TA--Coenzyme A ligase, and terephthalyl-CoA decarboxylase (Fig. 3.1, 3.2A, and 3.2B; Appendix B Table B.4 and supplementary material). Likewise, metatranscriptomics reveals TM-degrading *Syntrophorhabdus* and *Ca.* Syntropharomatica populations with particularly high activity during TM amendment and also high expression of putative TM degradation through a novel Na^+:TM symporter, TM--Coenzyme A ligase, and unique aromatic decarboxylase (Appendix B supplementary material). Besides identification of these syntrophic BA, TA, and TM degraders, we peculiarly find other populations with high activity during degradation of all substrates.

During TA and TM amendment, in addition to the TA/TM degraders, we detect activity of syntrophic aromatic compound metabolizers that cannot degrade these substrates (Fig. 3.1). Each syntrophic TA/TM degrader expresses catabolism to conventional end-products (*i.e.*, acetate, H_2, formate, and CO_2; Appendix B Table B.5, B.6, and B.8), which leaves no ecological niche for the other populations. However, these seemingly irrelevant populations display high expression of BA degradation (*i.e.*, in top 12.5% of genes; Appendix B Table B.5), suggesting that the syntrophic TA/TM degraders generate BA as an additional product (*e.g.*, $TA^{2-} + H_2O \rightarrow BA^- + HCO_3^-$) and transfer it to BA-catabolizing syntrophs for further degradation. Although BA could not be detected in the TA/TM amendments, we believe this is due to the high *in situ* activity of syntrophic BA degraders (Fig. S1). As a mechanism for BA secretion, many TA and TM degraders express genes that implicate dethiolation of the benzoyl-CoA intermediate to BA through a BA--CoA ligase (BamY) and BA secretion through a putative ABC-type BA transporter (BtrABCX) conserved among anaerobic aromatic compound degraders (Fig. 2, Table S3, Table S6, Table S7, and supplementary material). Although BamY and BtrABCX can also be used for BA uptake, the TA- and TM-amended microcosms did not contain any absorbable BA. Moreover, several TA (PtaU1.Bin065) and TM (PtaU1.Bin034 and PtaB.Bin047) degraders express these enzymes at levels in top 12.5 or 25% of their genes (Fig. 3.3A and 3.3B; Appendix B Table B.5, B.6, and B.7), confirming that BA generation is deliberate and that these organisms convert TA/TM to both acetate and BA.

Secreting BA forfeits energy that would otherwise be available from complete conventional oxidation (Fig. 3.2A), so why perform such metabolism? The key differences are that BA secretion is

exergonic (*e.g.*, $\Delta G^{\circ\prime}$=-26.1 kJ mol TA^{-1}) and does not produce H_2, while conventional oxidation is endergonic (*e.g.*, $\Delta G^{\circ\prime}$=44.5 kJ mol TA^{-1}) and generates H_2. Given the thermodynamic favorability of TA/TM decarboxylation to BA, cellular BA can accumulate to much higher concentrations than cellular TA/TM (*e.g.*, $7.7x10^5$ BA:TA at equilibrium). The high BA concentration can theoretically drive ATP synthesis through BA secretion by the corresponding ABC transporter (BtrABCX); on the other hand, the low cellular TA/TM concentration allows low-energy TA/TM uptake via Na^+:aromatic acid symporters. Thus, BA secretion can theoretically net a positive energy yield (*e.g.*, ~0.7 mol ATP mol TA^{-1}) by taking advantage of this concentration and energy differential. Though the energy yield is lower than conventional oxidation (*e.g.*, ~1.9 mol ATP mol TA^{-1}), diving TA/TM catabolism to acetate and BA can decrease H_2 production, whereby partially alleviating H_2-imposed thermodynamic inhibition. In other words, these syntrophic TA/TM degraders may funnel a portion of TA/TM into BA secretion to sacrifice ATP yield for improved catabolic thermodynamics (*e.g.*, 50% of TA to BA reduces $\Delta G^{\circ\prime}$ by 34.8 kJ mol TA^{-1}; Fig. 4C). In theory, flexibly adjusting the two pathways would allow the syntrophic aromatic compound degraders to balance energy yield and thermodynamics to optimize their metabolism depending on the environmental conditions (*e.g.*, high H_2 concentration) and diversify among populations with seemingly redundant function.

Besides BA secretion, we also observe syntrophic metabolizers in community U1 and B expressing pathways related to butyrate (Appendix B Table B.3 and B.5). *Pelotomaculum*-, *Syntrophorhabdus*-, and *Syntrophorhabdaceae*-related bins encode an energy-conserving butyryl-CoA dehydrogenase complex that can redirect crotonyl-CoA generated during aromatic acid degradation to reductive butyrate fermentation (*e.g.*, $TA^{2-} + 5H_2O \rightarrow Butyrate^- + Acetate^- + H_2 + 2H^+ + 2HCO_3^-$) (Fig. 3.2A and Appendix B Table B.3). Indeed, we observe expression of this pathway across all treatments with particularly high levels under BA and TA amendment in community U1 and B respectively (Fig. 3.3A and 3.3B; Appendix B Table B.5) and also detect butyrate at low concentrations (1.4~2.6 mM) during exponential phase. Interestingly, only *Pelotomaculum* populations display pronounced expression of butyrate fermentation, suggesting that *Pelotomaculum* may tend to specialize in syntrophic aromatic compound degradation to butyrate. Similar to BA secretion, compared to conventional oxidation, this pathway is more exergonic (*e.g.*, $\Delta G^{\circ\prime}$=-4.2 kJ mol TA^{-1}) and produces less H_2 at the cost of lower ATP (*e.g.*, ~1.3 mol ATP mol TA^{-1}); thus, partially funneling aromatic compound degradation into butyrate fermentation may be an alternative approach to modulating the thermodynamics of syntrophic catabolism (Fig. 3.4A).

Although we only observe the alternative BA- and butyrate-generating pathways in *Pelotomaculum*, *Syntrophorhabdus*, and *Ca*. Syntropharomatica, we identify one *Syntrophus* population that uniquely encodes and highly expresses CHC fermentation under BA amendment (Fig. 3.3B; Appendix B Table B.3,

B.5, and B.8). As described in a previous study, *Syntrophus aciditrophicus* can disproportionate the dienoyl-CoA intermediate to acetate and CHC to accomplish BA catabolism (Fig. 3.2) (*144*). Following the same trend, pairing conventional oxidation with CHC fermentation can modulate H_2 generation and thermodynamics of BA catabolism (Fig. 3.4B). In total, we discover that co-existing syntrophs utilize diverse alternative metabolic pathways (*i.e.*, BA, butyrate, and CHC production) that can sacrifice ATP yield to decrease H_2 generation and increase the thermodynamic favorability of syntrophic aromatic compound degradation. Notably, such pathways would theoretically allow syntrophs to segregate into distinct thermodynamic niches, similar to the aforementioned differences between two types of syntrophic propionate degradation. Moreover, phylogenetic diversity in the utilization of alternative pathways implicates that catabolism modulation strategies may differ between phylogenetic groups.

To complement aromatic compound catabolism, syntrophic metabolizers are known to use H^+ and CO_2 as electron acceptors to dispose catabolism-derived electrons, whereby generating H_2 and formate respectively. Each studied syntrophic population encodes multiple hydrogenases and formate dehydrogenases dependent on different electron carriers (*i.e.*, quinol, NADH, NADPH, and ferredoxin) (Fig. 3.2B; Appendix B Table B.3). In addition, we observe association of these enzymes with different phylogenetic lineages (Fig. 3.2B). For H_2 generation, all syntrophic metabolizers encode trimeric electron-confurcating FeFe hydrogenases while only *Pelotomaculum* possess the tetrameric electron-confurcating hydrogenase (Fig. 3.2B; Appendix B Table B.3). Interestingly, various populations may also perform novel quinol-oxidizing H_2 generation through a "hybrid" hydrogenase composed of a periplasmic FeFe hydrogenase associated with NiFe hydrogenase membrane subunits (HyaAC). For formate dehydrogenases, many non-*Syntrophus*-related aromatic compound degraders encode a putative NADPH-dependent electron confurcating complex, *Syntrophorhabdaceae* members harbor a putative NADH-dependent electron confurcating complex, and various populations across the four studied lineages possess NADH- and quinol-oxidizing formate dehydrogenases (Appendix B Table B.3).

The differences in hydrogenase/formate dehydrogenase partner electron carrier reduction potentials theoretically result in varying thermodynamic tolerance of H_2 or formate accumulation (*i.e.*, concentration at which $\Delta G^{\circ\prime}$ becomes positive; Fig. 3.4D). Thus, each population likely encodes multiple hydrogenases and formate dehydrogenases to adjust H_2 and formate generation depending on the immediate environmental conditions (*e.g.*, H_2 concentration). We observe little consistency between syntrophic metabolizers regardless of phylogeny or aromatic compound degradation capacity (Fig. 2B). Like aromatic compound degradation to differing byproducts, variation in H_2- and formate-generating mechanisms between syntrophic metabolizers implicates that each population has distinct strategies to modulate the thermodynamics of syntrophic catabolism.

Pairing the metabolic diversity of the "thermodynamic escape" pathways for modulating aromatic compound metabolism thermodynamics and H_2/formate generation mechanisms creates an unexpected level of variation among syntrophs. Close inspection of each syntrophic metabolizer's gene expression reveals that concurrently active syntrophic populations utilize different aromatic compound metabolism, hydrogenases, and formate dehydrogenases (Fig. 3.3A and 3.3B). In community B, the two TA-degrading *Pelotomaculum* (PtaB.Bin013 and Bin104) both encode BA production and two types of electron-confurcating hydrogenases (Fig. 3.2B); however, only PtaB.Bin013 secretes BA and produces H_2 through a tetrameric hydrogenase and PtaB.Bin104 does not express BA production and uses the trimeric hydrogenase (Fig. 3.3B; Appendix B Table B.6). In addition, although three *Pelotomaculum* populations encoding butyrate fermentation collaboratively degrade TA in community B, only the TA-degrading populations express butyrate fermentation. Likewise, all populations in community U1 and B express different combinations of BA secretion, butyrate fermentation, CHC fermentation, hydrogenases, and formate dehydrogenases regardless of BA, TA, or TM amendment (Fig. 3.3A and 3.3B). Therefore, seemingly redundant syntrophic populations can actually occupy different ecological niches.

We also find that the same syntrophic population alters their thermodynamic behavior depending on the treatment. A community U1 *Pelotomaculum* (PtaU1.Bin065) capable of degrading both BA and TA only expresses butyrate fermentation when catabolizing BA (Fig. 3.3A). Similarly, a BA- and TM-degrading *Ca*. Syntropharomatica (PtaU1.Bin034) only expresses butyrate fermentation under BA amendment. Interestingly, neither population changes their expression pattern of hydrogenases and formate dehydrogenases between BA and TA/TM degradation. On the other hand, a *Syntrophus* population (PtaB.Bin001) in community B adjusts its H_2 and formate generation mechanisms across all treatments even though it is consistently degrading BA (Fig. 3.3B). Thus, syntrophic metabolizers clearly have distinct approaches to adapting to different environmental and ecological conditions. In conclusion, through comparing the behavior of concurrently active syntrophic metabolizers and syntrophic metabolizers across different conditions, we discover that these organisms can take advantage of diverse metabolic pathways to segregate into different thermodynamic niches from competitors targeting the same substrate and also to adjust syntrophic catabolism according to changing conditions (*e.g.*, substrate availability).

Syntrophs are a fascinating group of organisms that thrive at the thermodynamic limit. Although these organisms were presumed to have strict thermodynamic restrictions, we discover that they have various metabolic strategies at their disposal to adjust such boundaries. Such unexpected biochemical prowess can serve as the cornerstone to the survival and diversification syntrophs, perhaps allowing them to partition into different niches and explaining the frequently observed redundancy of syntrophs (*140-143*). These diverse niches may also allow syntrophs to adapt to natural gradients and heterogeneity found

in many ecosystems, including the studied communities (Appendix B Fig. B.3), and even influence the resistance and resilience of methanogenic ecosystems (*145, 146*). Thus, this study provides a valuable step forward in comprehending syntrophic metabolism at the thermodynamic limit and its diversity. Future efforts into omics-, cultivation-, and biochemistry- based approaches are necessary to continue unraveling such enamoring anaerobic metabolism and ecology.

3.4 – Materials and Methods

This study investigated the microbial communities of methanogenic, mesophilic (37°C) laboratory- (U1) and full- (B) scale granular upflow anaerobic sludge blanket reactors treating purified terephthalic acid process wastewater, which contains a mixture of TA, BA, TM, 4-methylbenzoate, and acetate as the major constituents (Appendix B Table B.9 and supplementary information) (*147-149*). Metagenomic DNA was extracted from three granular sludge samples at separate times points in intervals of at least one month for U1 and at different sludge bed depths for B using the FastDNA SPIN kit for Soil DNA Extraction (MP Biomedicals, Santa Ana, CA, US). The extracted DNA was sequenced, assembled, and annotated as described previously (*150*). Draft genomes were constructed through combining binning results from MaxBin 2.0 (*52, 56*) and MetaBAT (*55*), checking for contamination using CheckM (*151*), and manual curation as described previously (*143, 152*). For the substrate amendment experiments, granular sludge samples were inoculated as 1.5% (w/v) into anaerobic medium containing 12 mM BA, TA, and TM. The anaerobic medium was prepared as previously described with a decreased $Na_2S•9H_2O$ concentration (0.03 g L^{-1}) (*153*). Aromatic compound concentrations were monitored using high-pressure liquid chromatography with a Agilent Eclipse XDB C18 column (ID x L = 4.6 mm x 250 mm), an acetonitrile/water (1:1 v/v) mobile phase, average pressure of 700 psi, and 0.6 mL min^{-1} flow rate. TA and TM were measured using an absorption wavelength of 239 nm and BA using 230 nm. CH_4 was quantified using gas chromatography as described previously (*153*).

For RNA extraction, granules were collected when aromatic compound degradation and CH_4 generation reached exponential phase while flushing N_2 gas and immediately frozen at -80°C. More specifically, for the experiments using lab-scale bioreactor granules, biomass was collected 1.1, 3.2, 7.9 days after inoculation for BA-, TA-, and TM-fed treatments respectively. For full-scale granules, biomass was collected 2.1, 5.8, and 20 days after inoculation respectively. Metatranscriptomic RNA was prepared by stabilizing the RNA with pH 5.1 buffer and acid-phenol:chloroform, pulverizing the biomass using 2 mL FastPrep Lysing Matrix (MP Biomedicals) with a bead beater, extracting RNA through acid phenol:chloroform extraction (*154*), and treated with DNase to remove residual DNA using TURBO DNA-free kit (Ambion, Foster City, CA, USA) and RNeasy Mini kit and RNase-free DNase set (Qiagen, Hilden,

Germany). Ribosomal RNA was removed from the DNase-treated RNA using the Ribo-Zero Magnetic Bacteria kit (Illumina, San Diego, CA, USA). A random hexamer and SuperScript II (Life Technologies, Grand Island, NY, USA) were used to perform first-strand synthesis. Double stranded cDNA were ligated to indexed adaptors and amplified through 10 PCR cycles with the Kapa HiFi polymerase (Kapa Biosystems, Wilmington, MA). After determining the size distribution using an Agilent bioanalyzer DNA7500 DNA chip (Agilent Technologies, Santa Clara, CA) and dilution to 10 nM, the sample libraries were pooled evenly for sequencing on an Illumina HiSeq2500 with version 4 sequencing reagents. The metatranscriptomic sequences were trimmed using Trimmomatic v0.30 with a quality cutoff of 30, sliding window of 6 bp, and minimum length cutoff of 50 bp (*155*) and mapped to metagenomic bins using the Burrows Wheeler Aligner (*156*). The gene expression levels were calculated separately for individual bins as reads per kilobase transcript per million reads mapped to the bin (RPKMB). In the heat map illustration, the gene expression level is further normalized to the median gene expression level of each bin (RPKMB-NM).

Thermodynamic calculations were conducted using Gibbs free energy of formation from Thauer *et al.* (*117*) using the following concentrations: 1 mM acetate (Ac), 1 mM BA, 1 mM TA, 10 μM butyrate, 1 mM CHC, 50 mM HCO_3^-, and 50 kPa CH_4. For calculations of butyrate fermentation, a relatively low butyrate concentration (10 μM) was used since the bioreactor influent contains no butyrate, but does contain TA, TM, BA, and acetate. As for CHC fermentation, the CHC concentration is set relatively high (1 mM) as the isolate known to perform this metabolism (*Syntrophus aciditrophicus* strain SB) produces it at concentrations greater than 1 mM (*144*). As for thermodynamic calculations for H_2 and formate generation, half reaction values were used from Schuchmann and Muller (*157*) and the H^+ gradient across the membrane was assumed to be ΔpH of approximately 3.1 based on an energy requirement of 15 kJ per mol H^+ extruded (*133*). Similarly, ATP yield for each pathway was calculated based on the assumption that 3.3 mol H^+ are extruded out of the cytosol per mol ATP hydrolyzed by ATP synthase.

Sludge granules were washed and fixed with paraformaldehyde immediately after being sampled from the reactor, as previously reported (*69*). Good shaped granules with the diameter of 2 cm were selected from the fixed ones. Then the phosphate buffered saline (PBS) and ethanol in the granules were replaced with optimal cutting temperature (OCT) compound step-by-step. Briefly, the granules were rinsed in 0.1 M PBS and ethanol mixture (2:1 and 4:1 in volume, respectively) for 5 min, then 0.1 M PBS for 5min, followed by 0.1 M PBS and OCT mixture (4:1 and 1:1 in volume, respectively) for 1 hr, finally immersed in 0.1 M PBS and OCT (1:1) overnight. The granule embedded OCT block was cooled and sectioned to 12 μm slices on a Leica CM3050 S cryostat (Leica Biosystems, Wetzlar, Germany) and transferred to poly-L-Lysine coated microscope slides (Sigma-Aldrich, St. Louis, MO, USA). The slides were heated at 65 ˚C for 2 hours

to allow tight adherence of cells. The OCT compound on the slides was washed by dipping them into molecular grade water for 2 min and air dried at room temperature.

In situ hybridization was conducted according to protocols previously described (Sekiguchi et al, 1999). Fluorescence labeled probes targeting *Deltaproteobacteria* (Cy3-labeled Delta495a, b, and c probe mixture) (*158*), total bacteria (FITC-labeled EUB338 probe mixture) (*159, 160*) and total archaea (Cy5-labeled Arc915) (*161*) were mixed with hybridization buffer (0.9 M NaCl, 1% SDS, 100 mM Tris-HCl, pH 7.2) to a final concentration of 10 ng/μL for each probe. Optimized formamide amount of 35% was used to assure the hybridization stringency. For each slide, 100 μL mixture was applied on the granule slices until all cells are immersed. Slides were put in a lid-closed chamber with hybridization buffer and hybridized at 46 ˚C over night. For washing, slides were briefly rinsed in the same hybridization buffer and then incubated for 30 min at 48 ˚C. Then slides were gently rinsed in milliQ ultra-pure water to wash off the residual buffer and probe. Anti-fade reagent (Life Science, USA) was added to the slides after air dry. Finally, fluorescence photos were taken at randomly picked areas by a confocal laser scan microscope (Zeiss LSM 700).

3.5 – Figures

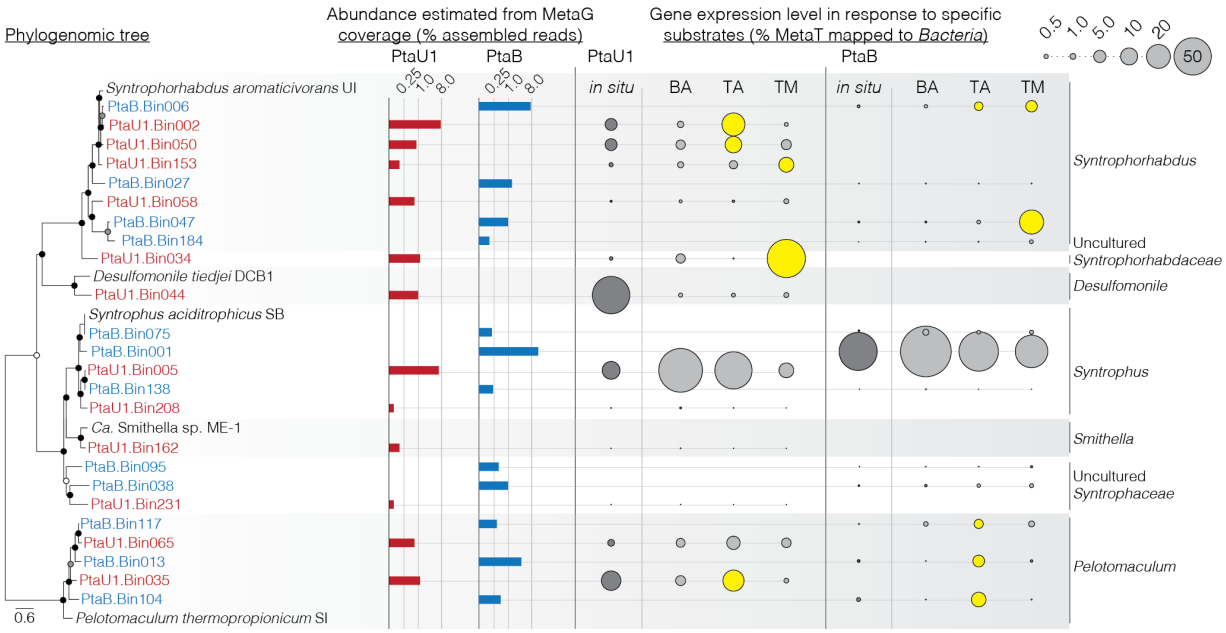

Figure 3.1. Metagenomics-based abundance and metatranscriptomics-based activity of syntrophs. The phylogeny of the syntroph-related populations identified in communities U1 and B are shown with their corresponding abundance estimated from their genomes' metagenomic coverage calculated as the percentage of metagenomic reads mapped to each genome relative to the total reads mapped to all constructed bacterial draft genomes (both *Bacteria* and *Archaea*). The estimated activity of these syntrophs in response to specific substrates (benzoate – BA, terephthalate – TA, and trimellitate – TM) are shown as the percentage of metatranscriptomic reads mapped to each genome relative to total reads mapped to all constructed bacterial draft genomes. For TA and TM amendment, the circle representing activity level is marked yellow if the population's estimated activity increased at least three-fold relative to the BA treatment. Syntroph populations with notable abundance and activity (>1% for both estimated abundance and activity) are assigned population names and symbols (colored circles with abbreviation of population name) for the purposes of this study.

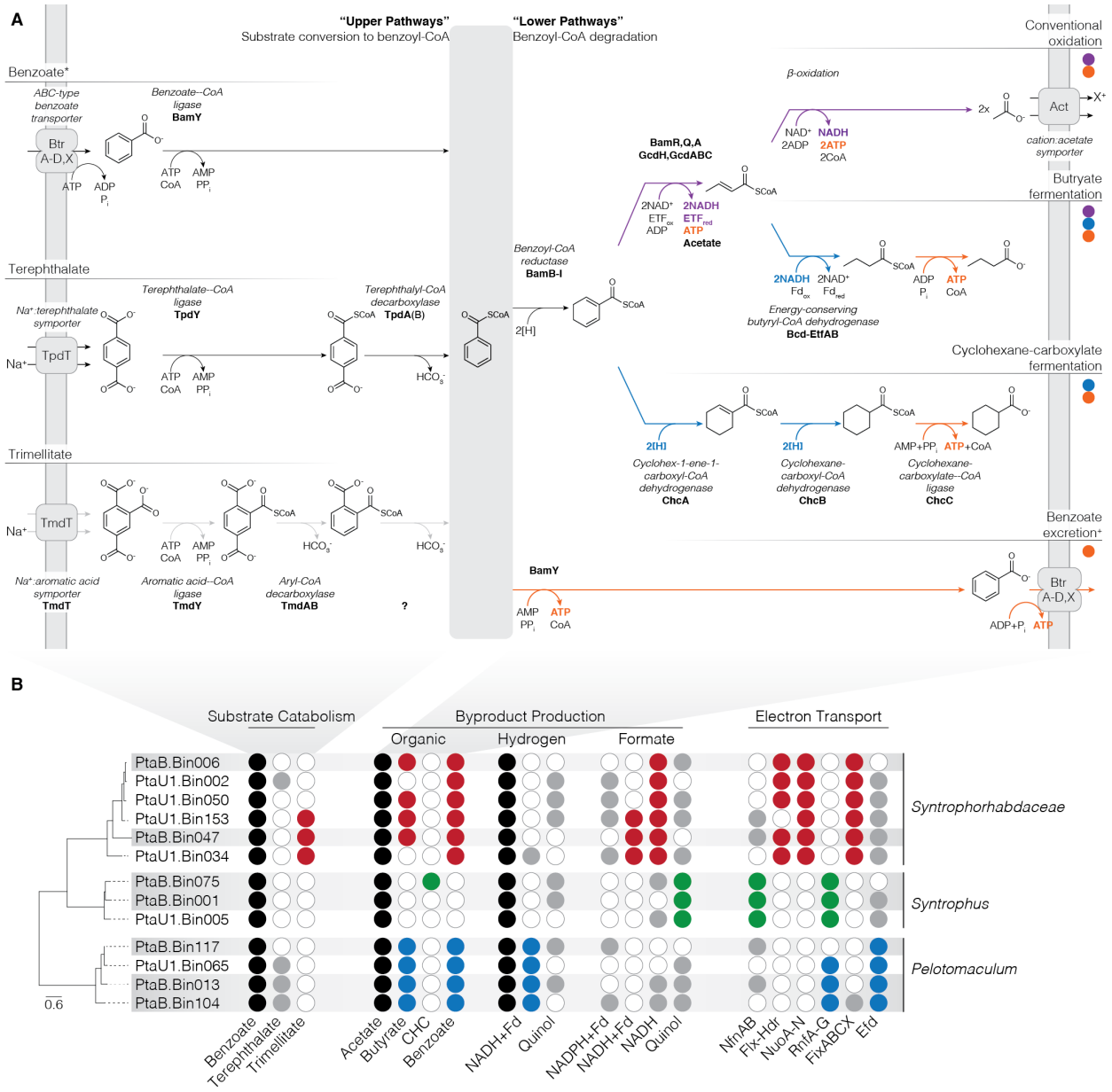

Figure 3.2. Catabolic pathways for aromatic compound metabolism and distribution of catabolic pathways among the studied syntrophs. (A) For "upper" pathways of aromatic compound conversion to benzoyl-CoA, benzoate, terephthalate, and trimellitate degradation are shown. In addition, the syntrophs encoding these pathways are shown using symbols defined in Figure 1. For benzoate degradation (*), the *Pelotomaculum* benzoate transporter was not found since a benzoate transporter has yet to be identified in gram-positive anaerobes. For "lower" pathways of benzoyl-CoA degradation, conventional oxidation to acetate, butyrate fermentation, cyclohexanecarboxylate (CHC) fermentation, and benzoate excretion are shown. For benzoate excretion (+), *Syntrophus* are not shown even though they encode BamY and BtrABCDX since the only aromatic compound they can degrade is benzoate and benzoate excretion during such metabolism would not generate energy. These lower pathways and their metabolic steps are colored based on whether they contain oxidative (purple), ATP-generating (orange), or reductive (blue) steps. These metabolic pathways use nicotinamide adenine dinucleotide (NAD), electron transfer flavoprotein (ETF), and ferredoxin (Fd) as electron carriers and other cofactors such as adenosine

Figure 3.2. (cont.) mono/di/triphosphate (AMP/ADP/ATP) and coenzyme A (CoA). Enzyme abbreviations and their corresponding genes are elaborated in Table B.3. (B) For each syntroph, we show presence (indicated by filled circle) of genes encoding pathways for substrate catabolism; production of organic byproducts (acetate, butyrate, CHC, and benzoate), H_2 (using various electron donors), and formate (same as H_2); and energy conservation pathways (see Table B.3). Filled circles are colored if the pathway is present in all syntrophs (black), all *Syntrophorhabdaceae* (red), all *Syntrophus* (green), all *Pelotomaculum* (blue), and several syntrophs with no phylogenetic consistency (gray).

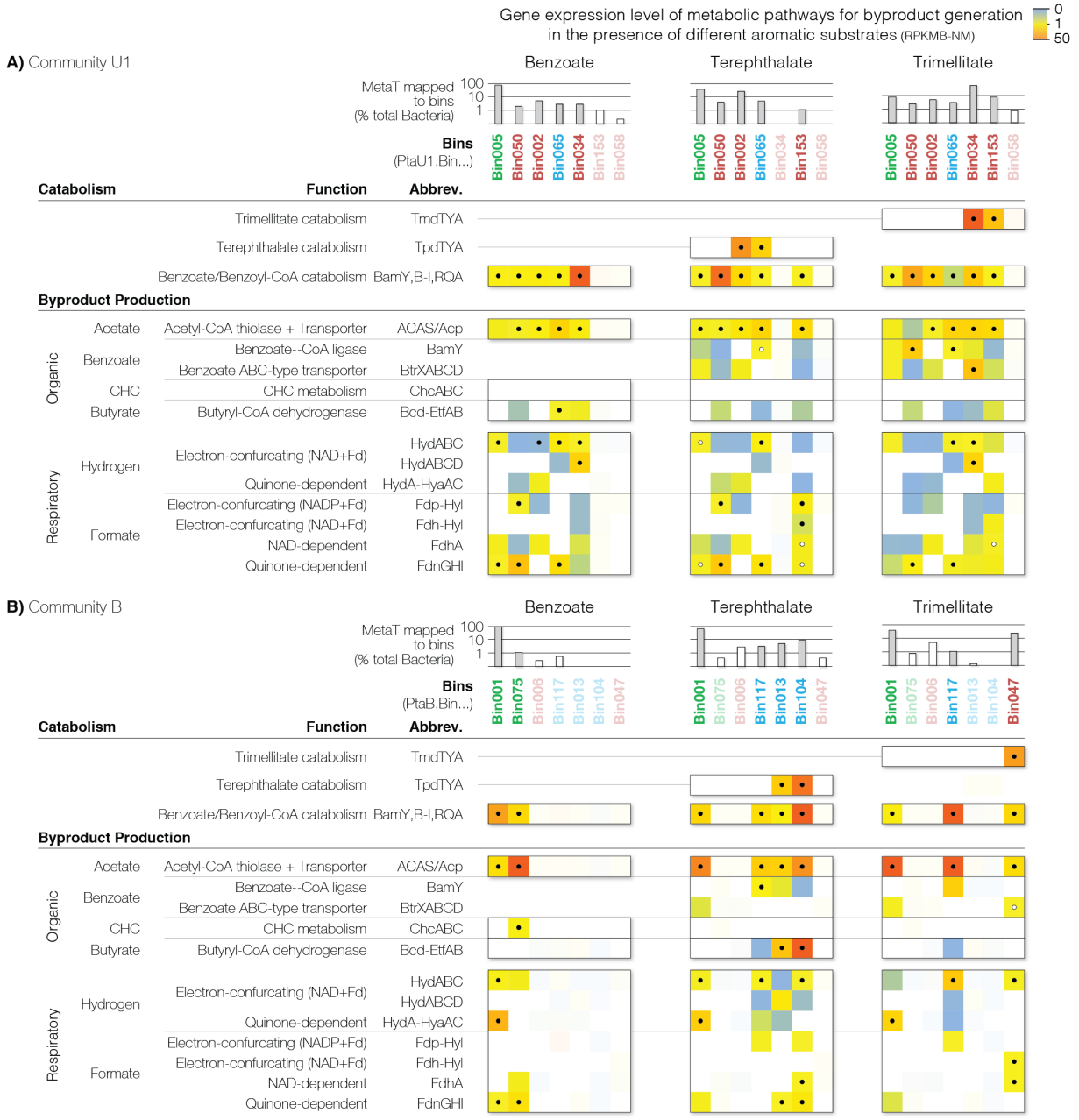

Figure 3.3. Gene expression of thermodynamic escape routes and H_2 and formate generation mechanisms for syntrophic aromatic compound metabolism. (A, upper) For each bin, the percentage of the metatranscriptomics reads mapped to the bin out of the metatranscriptome mapped to all bacterial bins is shown. Those expressing activity greater than 1% and genes involved in aromatic compound metabolism at expression levels higher than the bin's top octile (>87.5% of other genes in same genome) are marked grey. Those that do not meet the criteria are marker white. (A, lower) For syntrophs (*Syntrophus* – green, *Syntrophorhabdaceae* – red, *Pelotomaculum* – blue) in community U1 that meet the criteria, a heat map of genes expression levels of benzoate, terephthalate, and trimellitate catabolism and acetate, benzoate, and cyclohexanecarboxylate (CHC) production and H_2 and formate generation mechanisms in response to benzoate (left panel), terephthalate (middle), and trimellitate (right) amendment is shown. Different H_2/formate generation mechanisms are categorized based on their electron donor: nicotinamide adenine dinucleotide (NAD), NAD-phosphate (NADP), quinone, and

Figure 3.3. (cont.) ferredoxin (Fd). The gene expression levels are calculated as reads per kilobase of transcript per million reads mapped to individual bins (RPKMB) normalized to the median gene expression for the corresponding bin (RPKMB-NM) averaged from duplicate treatments. For each syntrophic population, pathways that contain genes with expression levels higher than 87.5% or 75% of other genes in the bin are marked with black and white dots respectively. (B, upper and lower) The same data is presented for community B.

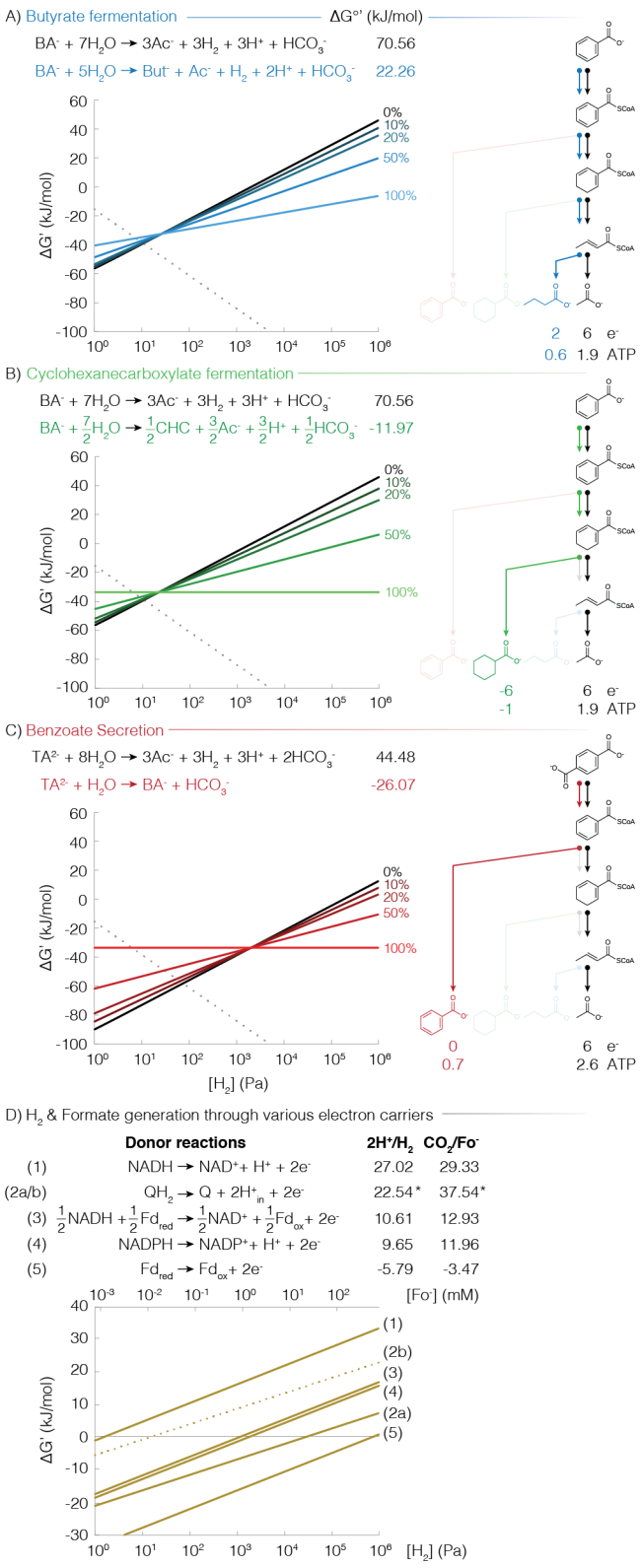

Figure 3.4. Thermodynamics of "escape" pathways for syntrophic aromatic compound degradation and H_2 and formate generation mechanisms. The thermodynamics plots of Gibbs free energy ΔG' in kJ mol^{-1} *v.* H_2 concentration in Pa for syntrophic metabolism are shown with hydrogenotrophic methanogenesis as a reference (gray). (A) In the plot, conventional oxidation of benzoate (BA) to acetate (Ac), H_2, and HCO_3^- (black line) is shown with alternative BA degradation to butyrate, Ac, H_2, and HCO_3^- (lightest blue). Additional curves are shown for when 10, 20, and 50% of the BA is degraded through the alternative pathway (*e.g.*, 10% implies 90% of BA degraded through conventional oxidation and 10% through alternative pathway). On the right side of the plot, the theoretical adenosine triphosphate (ATP) yield, electron yield, and ΔG°' are shown for each reaction. (B) Similarly, conventional oxidation of terephthalate (TA) is shown against an alternative pathway degrading TA to BA. (C) Conventional oxidation of (BA) is shown against an alternative pathway degradation BA to cyclohexanecarboxylate (CHC). These were calculated with 1 mM acetate (Ac), 1 mM BA, 1 mM TA, 10 µM butyrate, 1 mM CHC, 50 mM HCO_3^-, and 50 kPa CH_4. (D) ΔG°' are shown for H_2 and formate generation through oxidation of a variety of electron carriers: nicotinamide adenine dinucleotide (NAD), NAD-phosphate (NADP), quinone, and ferredoxin (Fd). In the plot, ΔG°' is shown with H_2 concentration (Pa) on the bottom horizontal axis and formate concentration (mM) on the top horizontal axis. The cellular conditions assume physiological concentrations of NAD(P) species. (*) For calculating the quinol-oxidizing H_2 and formate generation that occurs at the membrane, the pH directly outside of the cytosol was calculated as 4.4 based on an energy demand of 15kJ per mol of H^+ extruded to the cytosol. This results in different curves for H_2 and formate, so the formate-generating curve is shown as a dotted line.

CHAPTER 4

CHASING THE ELUSIVE *EURYARCHAEOTA* CLASS WSA2: GENOMES REVEAL A UNIQUELY FASTIDIOUS METHYL-REDUCING METHANOGEN[2]

4.1 – Abstract

The ecophysiology of one candidate methanogen class WSA2 (or Arc I) remains largely uncharacterized, despite the long history of research on *Euryarchaeota* methanogenesis. To expand our understanding of methanogen diversity and evolution, we metagenomically recover eight draft genomes for four WSA2 populations. Taxonomic analyses indicate that WSA2 is a distinct class from other *Euryarchaeota*. None of genomes harbor pathways for CO_2-reducing and aceticlastic methanogenesis, but all possess H_2 and CO oxidation and energy conservation through H_2-oxidizing electron confurcation and internal H_2 cycling. As the only discernible methanogenic outlet, they consistently encode a methylated thiol coenzyme M methyltransferase. Although incomplete, all draft genomes point to the proposition that WSA2 is the first discovered methanogen restricted to methanogenesis through methylated thiol reduction. In addition, the genomes lack pathways for carbon fixation, nitrogen fixation, and biosynthesis of many amino acids. Acetate, malonate, and propionate may serve as carbon sources. Using methylated thiol reduction, WSA2 may not only bridge the carbon and sulfur cycles in eutrophic methanogenic environments, but also potentially compete with CO_2-reducing methanogens and even sulfate reducers. These findings reveal a remarkably unique methanogen "*Candidatus* Methanofastidiosum methylthiophilus" as the first insight into the sixth class of methanogens "*Candidatus* Methanofastidiosa."

4.2 – Introduction

As an essential component of the global carbon cycle, methanogenic archaea produce approximately one billion tons of CH_4 per year (*1*). These "methanogens", uniquely affiliated to the archaeal phylum *Euryarchaeota*, facilitate anaerobic conversion of H_2, formate, acetate, methyl compounds, and simple alcohols to CH_4 and CO_2. In a variety of anaerobic ecosystems, including both natural and engineered, fermenters, syntrophs, and acetogens metabolically interact with methanogens to accomplish holistic mineralization of organic compounds (*143*). Methanogens also date back to 3.5 billion years ago based on isotopic signatures of Paleoarchean methane bubbles locked in the Dresser formation (*162*). Thus,

[2] This chapter appeared in its entirety in the published article: Nobu MK, Narihiro T, Kuroda K, Mei R, Liu WT (2016). Chasing the elusive *Euryarchaeota* class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. *ISME J* **10:** 2478-2487. (doi: 10.1038/ismej.2016.33). The journal does not require authors of original research papers to assign copyright of their published contributions.

researchers have made continuous efforts to characterize these ecologically, evolutionarily, and metabolically important organisms using cultivation-, biochemistry-, and genomics- based tools.

Recent cultivation- and metagenomics- based studies on *Euryarchaeota* methanogens discover novel methanogenic lineages, including permafrost-associated "*Ca.* Methanoflorentaceae" (also known as rice cluster II) (*163*) and gut-associated *Thermoplasmata* order *Methanomassiliicoccales* (formerly "*Ca.* Methanoplasmatales") (*164-166*). However, one potentially class-level *Euryarchaeota* clade thought to be capable of methanogenesis, WSA2 (or Arc 1), (*167, 168*) remains uncharacterized, despite identification through 16S rRNA sequencing more than 15 years ago (*169*). Moreover, members of this clade have been observed in a wide range of natural and engineered environments (*e.g.*, freshwater and marine sediments, contaminated groundwater, and bioreactors) (*170-173*). To fill this gap in our understanding of methanogen phylogeny and WSA2's potential roles in anaerobic biogeochemical cycles, we construct the first WSA2 genomes through metagenomics of methanogenic bioreactors treating wastewater. Using these genomes, we characterize the physiology and methanogenic potential of the sixth class of methanogens and ninth class of *Euryarchaeota*, "*Candidatus* Methanofastidiosa."

4.3 – Results and Discussion

4.3.1 – Phylogeny of WSA2 based on 16S rRNA, mcrA, and phylogenomics

Metagenomics collectively generated genomes for four WSA2 populations from a full-scale anaerobic digester treating a mixture of waste activated sludge and primary sludge in Urbana, Illinois, USA (population ADurb1213_Bin02801) and lab-scale (U1lsi0214_Bin055 and U1lsi0214_Bin089) and full-scale (B15fssc0709_Meth_Bin003) methanogenic bioreactors treating purified terephthalate wastewater (Appendix C Table C.1). For the Urbana anaerobic digester, we sequenced metagenomes in one-month intervals over three months. As for the full-scale reactor, microbial community samples were taken at separate sampling ports along the depth of the reactor sludge bed. Only one high quality bin was recovered for each WSA2 population identified in the lab-scale reactor. In total, the eight draft genomes reconstructed from seven distinct sequencing efforts range between 1.50-1.99Mb in size and 32.9-35.4% G+C content with estimated completeness of 85.0-92.5% based on 40 universal marker genes (*174, 175*). Draft genomes from the same environment (*i.e.*, digester and full-scale reactor) with at least 85% overlap of predicted genes with a 99% similarity cutoff were combined, generating population pangenomes for populations ADurb1213_Bin02801 and B15fssc0709_Meth_Bin003. 16S rRNA gene classification, McrA (methyl-coenzyme M reductase alpha subunit) phylogeny, and phylogenomics clearly indicate that these organisms belong to the uncharacterized *Euryarchaeota* clade WSA2 falling between super-classes *Methanomicrobia-Halobacteria-Thermoplasmata* and *Methanobacteria-Methanococci* (Fig. 4.1A; Appendix C Fig. C.1 and

C.2). WSA2 is distinct from related *Euryarchaeota* classes and uncultivated clades based on phylogenetic tree topology with high bootstrap support, regardless of the reference 16S rRNA alignments used (Greengenes 2013 and Silva v123). Moreover, the median sequence similarities of WSA2 to neighboring classes (76.63-81.7%) are lower than or comparable to the similarity between *Methanobacteria* and *Methanococci* (81.5%), confirming that WSA2 is indeed a class-level clade in *Euryarchaeota* (Fig. 1). Based on tree topology and this median sequence similarity criterion, uncultivated *Euryarchaeota* clades pMC1 and MSBL1 may form another class (tentatively termed pMC1-MSBL1) distinct from WSA2 with phylogenetic radiation between WSA2 and superclass *Methanomicrobia-Halobacteria-Thermoplasmata*.

Similarly, genome content also suggests distant relationship with known methanogens and other *Euryarchaeota* as only 25-28% and 50-58% of genes in each WSA2 genome respectively relate to methanogens and *Archaea* based on top blastp hits (Fig. 4.2C). The remaining *Archaea*-related genes hit *Thermococci* (12-16%), *Archaeoglobi* (3-4%), and other clades outside of *Euryarchaeota* (11-13%). A notable amount of genes relate to bacterial clades (28-35%) as also previously observed for other *Euryarchaeota* (*176*). However, these WSA2 genes have low average amino acid similarity to genes of known organisms (44-49%) (Appendix C Fig. C.3). The core genome of all four WSA2 genomes (defined by >80% amino acid similarity conserved between all genomes) contains 842 genes, which has a similar phylogenetic gene distribution (Fig. 4.2C).

To further define WSA2's physiology and relationship with other *Euryarchaeota* beyond this taxonomic analysis, we compare the core genomes of each clade (at genus and class level). Genera of similar phylogeny have pronounced core genome overlap; however, the WSA2 core genome has low overlap with both methanogenic (6-15%) and non-methanogenic (2-15%) *Euryarchaeota* genera (Fig. 4.2D). Moreover, even though distantly related methanogenic genera in *Methanomicrobia*, *Methanobacteria*, and *Methanococci* share many core genes (152-257 genes average), WSA2 shares much fewer genes with these methanogens (69-111 genes average) (Fig. 4.2B). The genes shared between WSA2 with *Euryarchaeota* genera primarily consist of housekeeping genes, including ribosomal proteins, biosynthesis genes, and *Archaea*-specific genes (*e.g.*, histone, thermosome, and ATP synthase). WSA2 and methanogenic clades all share genes encoding methanogenesis marker proteins and methyl-Coenzyme M (CoM) reductase complex (Mcr), responsible for the terminal step of methanogenesis (Appendix C Table C.2). Interestingly, WSA2 shares the methylmalonyl-CoA pathway with non-methanogenic genera, which no known methanogen genomes harbor. These findings of Mcr, unique genetic composition, and unusual methylmalonyl-CoA pathway suggest that WSA2 may perform methanogenesis and yet have distinct capabilities from typical methanogens.

4.3.2 – Uniquely restricted methanogenic metabolism

Even though the WSA2 genomes harbor Mcr, all eight draft genomes from the seven independent sequencing efforts lack conventional CO_2 reduction to CH_4 and acetyl-CoA synthase pathway, indicating that neither CO_2 nor acetate can serve as substrates. Moreover, while many methanogens missing these pathways respire methanol and methylamines (*e.g., Methanomethylovorans* and *Methanosphaera*), we could not identify corresponding methyltransferses essential for such metabolism in any of the WSA2 draft genomes. Strikingly, the WSA2 draft genomes consistently encode methylated thiol-CoM methyltransferase (Mts) homologs for funneling C1 compounds into the reductive arm of methanogenesis (Appendix C Table C.2). Thus, WSA2 may only be capable of demethylation of methylated thiols for CH_4 generation, which would be the first example of such restricted methanogenic catabolism. Moreover, the lack of the methanogenic C1 pathway indicates that WSA2 likely use methylated thiols as electron acceptors rather than donor. Methanol and methylated amine metabolism are observed in methanogens associated with sediments or gut environments as plants produce methanol and methylated amines derive from eukaryotic lipids and osmoregulators; thus, it may be logical that the studied anaerobic digester WSA2 populations lack such metabolism. Perhaps other WSA2 class members associated with marine environments or sediments may possess catabolic capacity beyond methylated thiols.

Mts substantially vary in specificity to methylated thiol substrates, such as methanethiol (MeSH), dimethylsulfide (DMS), 3-methylmercaptopropionate (MMPA), and 3-mercaptopropionate (MPA) (*177, 178*). Thus, accurate annotation of WSA2's methanogenic capacity necessitates systemic determination of the specific function of the two Mts homologs conserved amongst WSA2 genomes. Phylogenetic analysis of all publically available methanogen Mts-related methyltransferase sequences reveals clusters related to (a) methyl-metabolizing *Methanosarcinales* and *Methanomassiliicoccales*, (b) non-methyl-metabolizing methanogens, or (c) both (Fig. 4.3). We suspect that clusters found in non-methyl-metabolizing methanogens are not catabolic and rather serve physiological roles in methanogens. Clusters related to methyl-metabolizing methanogens were categorized into protein families, of which three have representatives with biochemical or transcriptomic characterization. A recent genetic and transcriptomic study on *Methanosarcina acetivorans* C2A shows evidence that MtsD, MtsF, and MtsH (family IV) respectively degrade DMS, MeSH, and both and MptA (family VII) is critical for MMPA metabolism (*178-181*). *M. barkeri* MtsA (family VIII) has also been demonstrated to facilitate methyl transfer from DMS, MMPA, and MPA, also MeSH to a much lesser extent (*182, 183*). In WSA2, we identify an Mts fusion protein encoding a methyltransferase and corrinoid protein, both of which relate to family VIII MtsA and MtsB respectively (Appendix C Fig. C.4); thus, WSA2 may utilize this MtsA and Mcr to perform methanogenesis through reduction of multi-carbon methylated thiols.

Many uncharacterized Mts homologs fall into five other families affiliated with methyl-metabolizing methanogens (I, II, III, V, and VI). While these groups have no experimentally studied representatives, their exclusive affiliation with methyl-degrading methanogens suggests that these Mts homologs relate to methanogenic methyl metabolism. Family I is unlikely to support MeSH, DMS, or MMPA metabolism as *Methanosarcina acetivorans* C2A does not express MA0847 during growth on these substrates (*178, 181*). Thus, the metabolic function of family I-related WSA2 Mts-like protein (MtlA) is unclear. Interestingly, MtlA associates with a corrinoid-CoM methyltransferase (MtlB), corrinoid protein (MtlC), and methyltransferase reductive activase (MtlD) (Appendix C Table C.2). Notably, MtlC is phylogenetically distinct from other known Mts corrinoid proteins (Appendix C Fig. C.4). While known Mts can single-handedly catalyze the two-stage methyl transfer from substrate to CoM, methanol- and methylamine- methyltransferases require two subunits (*183*); therefore, MtlAB may have substrates and biochemical properties distinct from known Mts. However, the presence of MtlA in other methyl-degrading methanogens warrants further experimental investigation to expand our understanding of diversity in methanogenic methyl metabolism.

To complement the reductive methylated thiol metabolism, the WSA2 genomes also possess H_2- and CO- oxidizing pathways for donating reducing power into methanogenesis (Appendix C Table C.2). For H_2 metabolism, WSA2 can perform (i) electron-bifurcating H_2 oxidation using a cytosolic methylviologen-reducing hydrogenase (MvhADG) complexed with a heterodisulfide reductase (HdrABC) and (ii) proton-pumping H_2 generation mediated by a membrane-bound energy-converting hydrogenase (EhbA-Q) (Fig. 4.4). WSA2 may pair these two pathways to accomplish energy-conserving internal H_2 cycling similar to *Methanosphaera stadtmaniae* and *Methanomassiliicoccales* spp. (*1, 184*) analogous to the *Methanosarcina* cytochrome-dependent H_2 cycling (*185*). Though WSA2 also encodes F_{420}-reducing hydrogenase (FrhAB), none of the draft genomes encode other F_{420}-oxidizing enzymes related to methanogenesis so the physiology function of F_{420} remains unclear. The WSA2 genomes lack genes encoding cytochromes and methanophenazine synthesis, suggesting that they perform cytochrome-independent methanogenesis like all non-*Methanosarcinales* methanogens (*1*).

4.3.3 – Physiology

Archaea are known to utilize a variety of autotrophic carbon fixation pathways (*186*), yet methanogenic *Euryarchaeota* specifically possess reductive acetyl-CoA pathway (rACP), pentose phosphate pathway (rPPP; or Calvin-Benson cycle), and/or tricarboxylic acid cycle (rTCA) (Fig. 4.2A). *Methanosarcinales*, *Methanocellales*, "*Ca.* Methanoflorens," *Methanobacteriales*, *Methanococcales*, and *Methanopyri* members encode rACP; *Methanomicrobiales* members either possess both rACP and rPPP or

only rPPP; and only several phylogenetically scattered species encode rTCA. Interestingly, the *Methanomicrobiales* members only encoding rPPP all require exogenous organic carbon for growth, suggesting that the rPPP ribulose-1,5-bisphosphate carboxylase/oxygenase in these organisms participates in adenosine monophosphate metabolism rather than carbon fixation (*187*). Likewise, WSA2 may require organic carbon sources for growth (*i.e.*, heterotrophic) as they lack any discernable carbon fixation pathway (Fig. 4.2A). In agreement, known methanogens missing such pathways require acetate or complex exogenous nutrients (*e.g.*, yeast extract or rumen fluid) for growth (*188-192*). In addition, unlike most methanogens, the WSA2 genomes are also deficient in nitrogen fixation genes (*i.e.*, Nif) though they do encode non-nitrogen-fixing NifD- and NifH- like genes (NflDH) (Fig. 4.2A) (*193*). Thus, the draft genomes suggest that WSA2 may be incapable of both autotrophy and nitrogen fixation, a rarity among known methanogens. This genomic analysis also revealed that several gut-associated *Methanobacteriales* (*Methanobrevibater* sp. JH1 and *Methanosphaera stadtmanae* DSM3091) also cannot fix carbon and nitrogen.

In order to determine the carbon source requirement for WSA2, we genomically evaluate their biosynthetic capacity. All WSA2 genomes have genes for gluconeogenesis and TCA cycle, but only two genomes encode a complete PPP (Appendix C Table C.3). The presence of TCA cycle and malonate decarboxylase suggests they can utilize acetate as a carbon source and also generate acetate from malonate. They encode a pyruvate:ferredoxin oxidoreductase (Por) and pyruvate carboxylase (Pyc) for acetate-CoA assimilation. Similarly, heterotrophic methanogens encode Por with either Pyc or oxaloacetate decarboxlyase for acetate assimilation. Unique from other methanogens, WSA2 encode methylmalonyl-CoA pathway that integrate propionate into the TCA cycle. TCA-based propionate assimilation could either split into the oxidative and reductive arms or strictly go through oxidative TCA given acetate or malonate is available as an acetyl-CoA source. Thus, WSA2 may utilize acetate, malonate, or propionate with CO_2 as carbon sources. However, all genomes lack complete biosynthesis pathways for glycine, homocysteince, homoserine, methionine, proline, threonine, and tryptophan and some genomes lack full pathways for isoleucine, leucine, phenylalanine, and tyrosine, suggesting severe auxotrophy (Appendix C Table C.3). To complement this, they encode amino acid and peptide transporters.

We identify pathways for synthesis of methanogenesis cofactors: Coenzyme B, F_{420}, F_{430}, and cobamide (partial) (Appendix C Table C.3). The WSA2 genomes also consistently harbor F_{420} modification genes. Interestingly, they also possess F_{420} alpha-L-glutamate ligase (CofF) that have only been observed in *Methanosarcinales* and *Methanococcales* previously (*194*). On the other hand, the WSA2 genomes are missing CoM synthesis pathways. Similarly, several phylogenetically scattered methanogen isolates also lack such genes. All of these methanogens are either dependent on rumen fluid for growth or strongly stimulated

by CoM supplementation. Though whether WSA2 and these methanogens possess an unidentified CoM synthesis pathway is unclear, WSA2 is likely adapted to availability of exogenous CoM produced by other methanogens *in situ*. This suggests that WSA2 can only inhabit ecosystems with other active methanogens. As for methanogens that do encode CoM biosynthesis, *Methanomicrobia* and *Methanomassiliicoccales* employ the *Methanosarcina*-type pathway while *Methanobacteria*, *Methanococci*, and *Methanopyri* employ *Methanocaldococcus*-type (*195-197*).

Taxonomic classification, core genome comparison, unique methylated thiol-specific methanogenic metabolism, and obligate heterotrophy all agree that WSA2 is a novel *Euryarchaeota* class with distinct features from other methanogens. Their heterotrophic, auxotrophic, and ammonia-dependent nature is unique and yet appropriate for the organic-rich anaerobic environments they are often associated with (*i.e.*, wastewater treatment sludge and marine sediments) where fermentation and methanogenesis produce the required carbon sources (acetate, propionate, and CO_2), amino acids, ammonia (*198, 199*), and possibly free CoM. Moreover, fermentative degradation of methionine and methoxylated compounds may yield sufficient DMS to support methanogenesis (*200-203*). Thus, we suspect that this novel archaeon is adapted to performing methanogenesis in eutrophic anaerobic environments and bridging the carbon and sulfur cycles. In addition, the existence of a dedicated H_2-oxidizing methylated thiol reducing methanogen can significantly change our thermodynamic understanding of anaerobic ecosystems. Typical CO_2-reducing methanogenesis ($4H_2 + HCO_3^- \rightarrow CH_4 + 3H_2O$: $\Delta G°' = -135.56$ kJ mol CH_4^{-1}) rapidly becomes thermodynamically unfavorable as H_2 concentration decreases due to the high stoichiometry of input H_2 to generated CH_4 (4:1) in contrast with the low stoichiometry (1:1) of WSA2-like methanogenesis (*e.g.*, H_2 + DMS $\rightarrow CH_4 + CH_3SH$: $\Delta G°' = -161.14$ kJ mol CH_4^{-1}). For example, at 10 Pa H_2, CO_2 reduction would only yield -18.48 kJ mol CH_4^{-1} while DMS reduction -140.04 kJ mol CH_4^{-1} (50 kPa CH_4, 100 μM thiols, 50 mM bicarbonate). In theory, WSA2-like methyl-reduction could thrive under low H_2 concentrations and maintain H_2 lower than CO_2-reducing methanogenesis to support H_2-producers (*i.e.*, syntrophs and fermenters) and also compete against sulfate reducers ($4H_2 + SO_4^{2-} + H^+ \rightarrow HS^- + 4H_2O$: $\Delta G°' = -152.2$ kJ mol^{-1}). Thus, WSA2 may play an essential and overlooked ecological role for syntrophy in methanogenic environments (*e.g.*, anaerobic digestion) and competitive methanogenesis in sulfate-reducing environments (*e.g.*, marine sediments).

Besides these unique features, WSA2 has signatures of cytochrome-independent methanogenesis (HdrABC-MvhDGA energy conservation), *Methanosarcinales* and *Methanococcales* F_{420} modification, and *Methanosphaera* and *Methanomassiliicoccales* internal H_2-cycling. In addition, characterization of WSA2 expands the phylogenetic range of methanogen heterotrophy. In conclusion, we present the first detailed description of the sixth methanogen class WSA2 as a significant step forward in understanding methanogen

phylogeny, metabolic diversity, and contribution to global biogeochemical cycles with provisional class and species assignment of "*Candidatus* Methanofastidiosa" class. nov. (Me.tha.no.fas.tid.i.o'sa. N.L. pref. *methano*-, pertaining to methane; N. L. f. adj. *fastidiosa*, highly critical; referring to the nutritional fastidiousness of the organism, particularly on primary isolation) and "*Candidatus* Methanofastidiosum methylthiophilus" gen. nov. sp. nov. (Me.tha.no.fas.tid.i.o'sum. N.L. pref. *methano*-, pertaining to methane; N. L. neut. adj. *fastidiosum*, highly critical; referring to the nutritional fastidiousness of the organism, particularly on primary isolation / me.thyl.thi.o'phi.lus. M. L. n. *methyl*, the methyl group; thi.o′phi.lus. Gr. n. *thion* sulfur; Gr. adj. *phylos* loving; N.L. neut. adj. *methylthiophilus* methyl- and sulfur- loving).

4.4 – Methods

4.4.1 – Metagenome sequencing, assembly, and binning

We collected samples from three anaerobic wastewater treatment samples: one full-scale anaerobic digester (reactor ADurb) and full- and lab- scale bioreactors treating wastewater from purified terephthalate process (Bfssc and U1lsi, respectively). For ADurb and U1lsi, three samples were taken at different time points, October, November, and December 2013 and February, April, and May 2014 correspondingly. For Bfssc, samples were taken through sampling ports at different depths of the reactor along the sludge bed. Metagenomic DNA was extracted using FastDNA SPIN Kit for Soil kit (MP Biomedicals, Santa Ana, CA, USA). Sequencing libraries were prepared with Kapa Library Preparation kit (Kapa Biosystems, Wilmington, MA, USA) with a genomic DNA fragment size ranging between 300-750bp. These libraries were sequenced on HiSeq2500 with TruSeq SBS Rapid Sequencing kit (Illumina, San Diego, CA, USA), generating paired-end reads up to 165bp each. The generated reads were trimmed using Trimmomatic v0.30 with a quality cutoff of 30, sliding window of 6bp, and minimum length cutoff of 75bp (*155*); digitally normalized and partition using the khmer package (*204, 205*); and assembled using SPAdes v.3.5.0 (*206*). The assembled contigs from three metagenomes corresponding to the same reactor were binned comparatively using MaxBin2.0 (*52*). Genes were then predicted using Prodigal v2.5 (*207*) and annotated using Prokka (*208*). We manually curated these bins and their annotations as described in our previous metagenomic study (*143*). For interpretation of physiology and metabolism, WSA2 draft genomes recovered from different samples of the same environment (*i.e.*, reactor ADurb time points or Bfssc depths) were collectively analyzed as a pangenome.

4.4.2 - Phylogenetic analysis

Genes annotated as small subunit ribosomal RNA (16S rRNA) gene, methyl Coenzyme M reductase alpha subunit (mcrA) and methylated thiol coenzyme M methyltransferase (Mts) were extracted for

phylogenetic analysis. The 16S rRNA genes were aligned against the Greengenes and Silva v123 databases using PYNAST and SINA respectively (*209-212*). Phylogenetic trees for 16S rRNA genes were constructed in ARB using the neighbor-joining algorithm with 5,000 bootstrap replications (*213*). The mcrA and Mts genes from WSA2 and publically available methanogen genomes were aligned and put into a phylogenetic tree using ClustalW in the MEGA package (*214*). A representative genome for each *Euryarchaeota* genus was used to construct a phylogenomic tree including WSA2 genomes using PhyloPhlAn (*57*). To further compare the phylogenetic relationship of these *Euryarchaeota*, core genomes were predicted for each genus and family with more than one genome publically available by identifying genes with >50% amino acid similarity and >80% alignment coverage across all genomes of that clade using the BLAST+ package (*215*), based on criteria used in a core genome prediction software (*216*). Core genome overlap between genera and families were determined using the same criteria.

4.5 - Figures

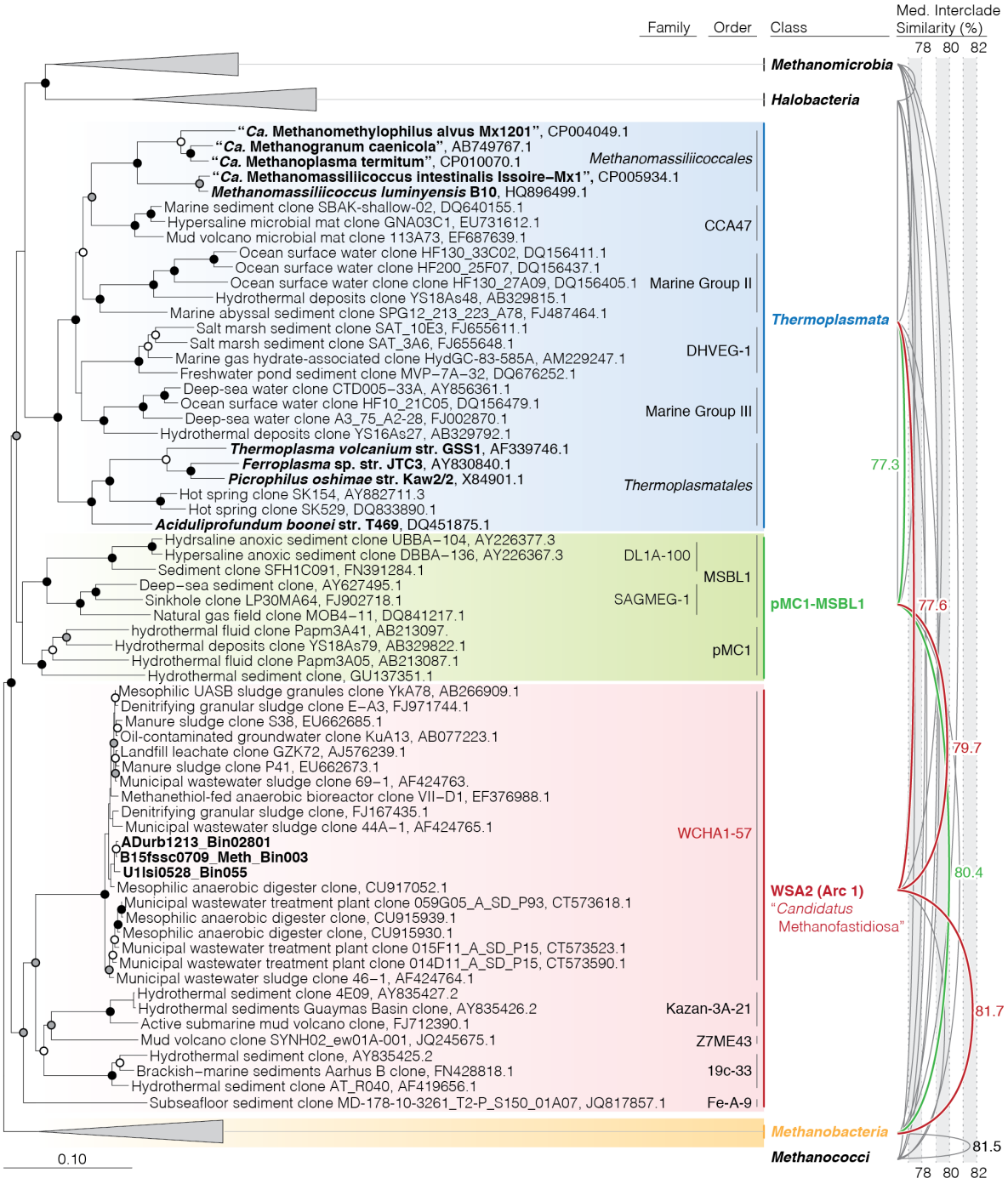

Figure 4.1. 16S rRNA-based phylogeny of WSA2 (red) compared to closely related *Euryarchaeota* classes and clusters (pMC1 and MSBL1). The tree was constructed using ARB neighbor-joining algorithm and GreenGenes 16S rRNA gene database with 5000 bootstrap replications, sequences at least 1200bp in length, and *Methanobacteria* as the outgroup. Bootstrap values greater than 90% (black), 75% (gray), and 50% (white) are indicated. The median sequence similarity of the *Euryarchaeota* classes and clusters (WSA2, pMC1, and MSBL1) are shown as arcs connecting the cluster labels with horizontal heights indicating similarity (right).

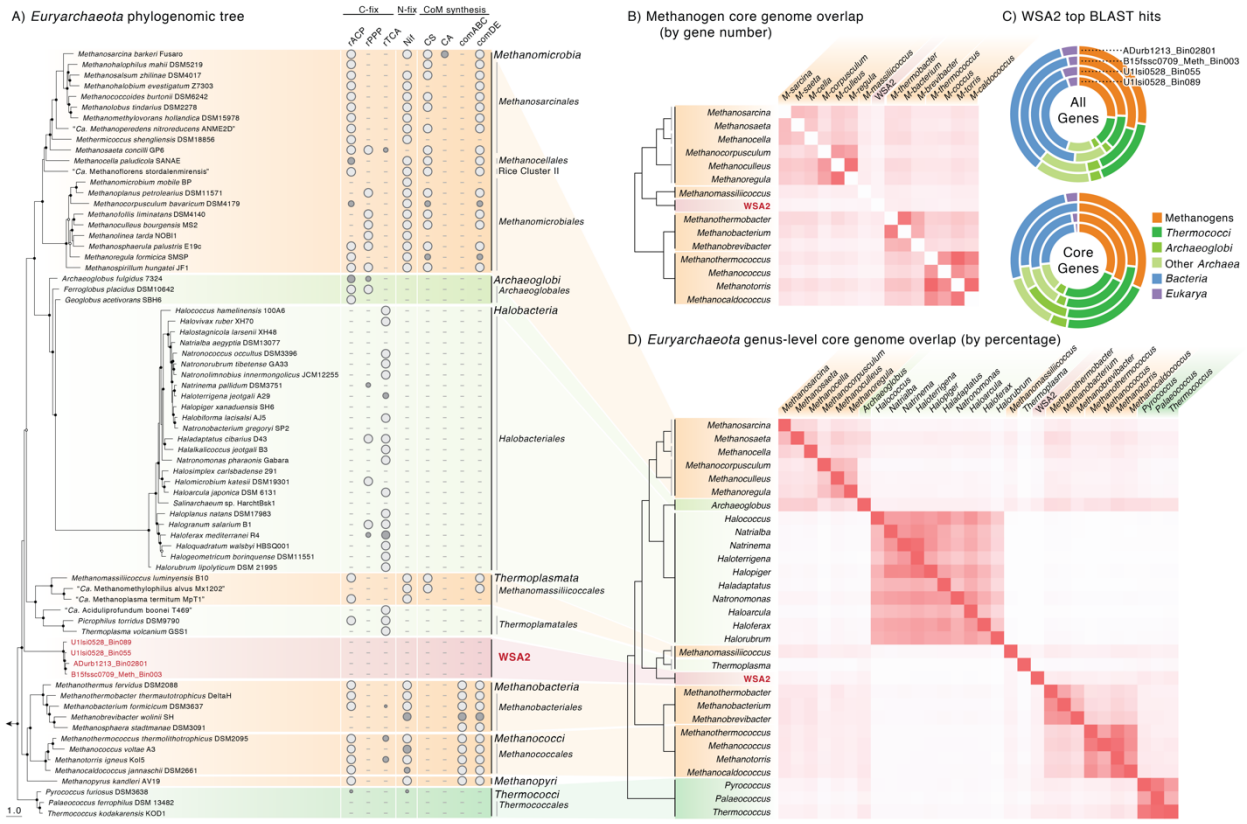

Figure 4.2. (A) *Euryarchaeota* phylogenomics tree of representative genera with *Crenarchaeota* as the outgroup. Nodes are marked for bootstrap values greater than 75% (filled) and 50% (open). For each genus, one representative species is shown on the tree. Corresponding to each genus, the presence of the gene / pathway in all (white) or only some (grey, size proportional to percentage) of the genomes of that genus are shown. Genes / pathways are shown for carbon fixation (reductive acetyl-CoA pathway – rACP; reductive pentose phosphate pathway – rPPP; and reductive tricarboxylic acid cycle – rTCA), nitrogen fixation (Nif), type I coenzyme M (CoM) synthesis (comABC), type II CoM synthesis (cysteate synthase – CS; cysteate aminotransferase – CA), and final CoM synthesis step (sulfopyruvate decarboxylase – comDE). (B) The overlap of core genomes between methanogenic genera, including WSA2. The intensity of the red color corresponds to the number of overlapping genes. (C) The phylogenetic distribution of WSA2 genes in the entire genome and only core genomes. All four genomes are shown as individual rings. (D) The overlap of core genomes between all *Euryarchaeota* genera. The intensity of the red color corresponds to the percent overlap of core genes in reference to the genome on the vertical axis.

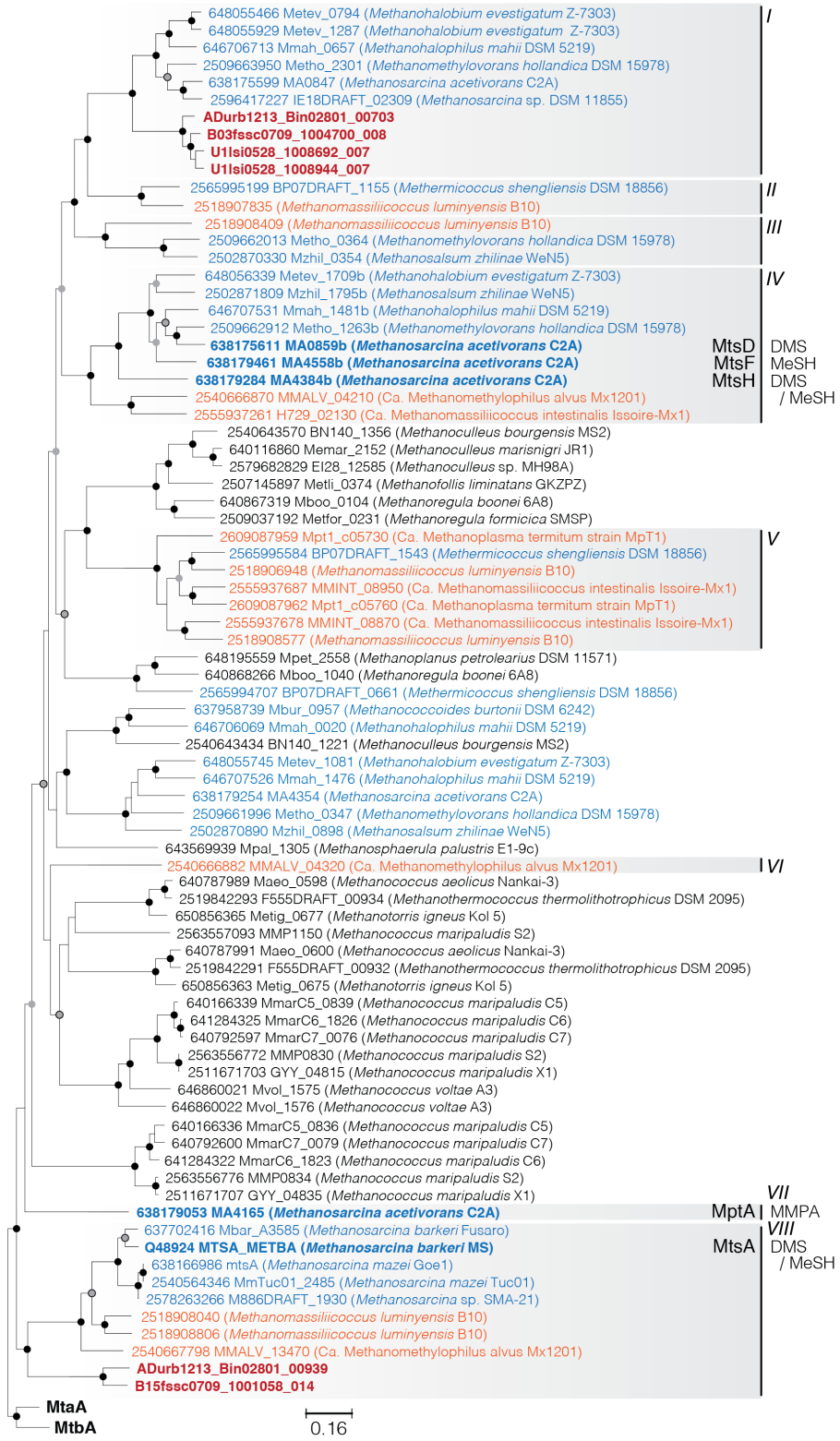

Figure 4.3. Phylogeny of methylated thiol Coenzyme M methyltransferase (Mts) homologs with methanol- (MtaA) and methylamine- (MtbA) specific methyltransferases as the outgroup. The tree was constructed using Clustalw with default parameters for alignment and neighbor-joining clustering method for tree construction with 5000 bootstrapped replications. Family classifications are assigned to clusters only containing Mts homologs from methyl-metabolizing methanogens (*Methanosarcinales* – blue, *Methanomassiliicoccales* – orange) and/or WSA2 (red and bolded). Only three families contain representatives biochemically or transcriptomically characterized in *Methanosarcina acetivorans* and *M. barkeri* (bolded). For those Mts, their predicted substrates are noted. Nodes are marked for bootstrap values greater than 90% (black), 75% (gray with outline), and 50% (gray).

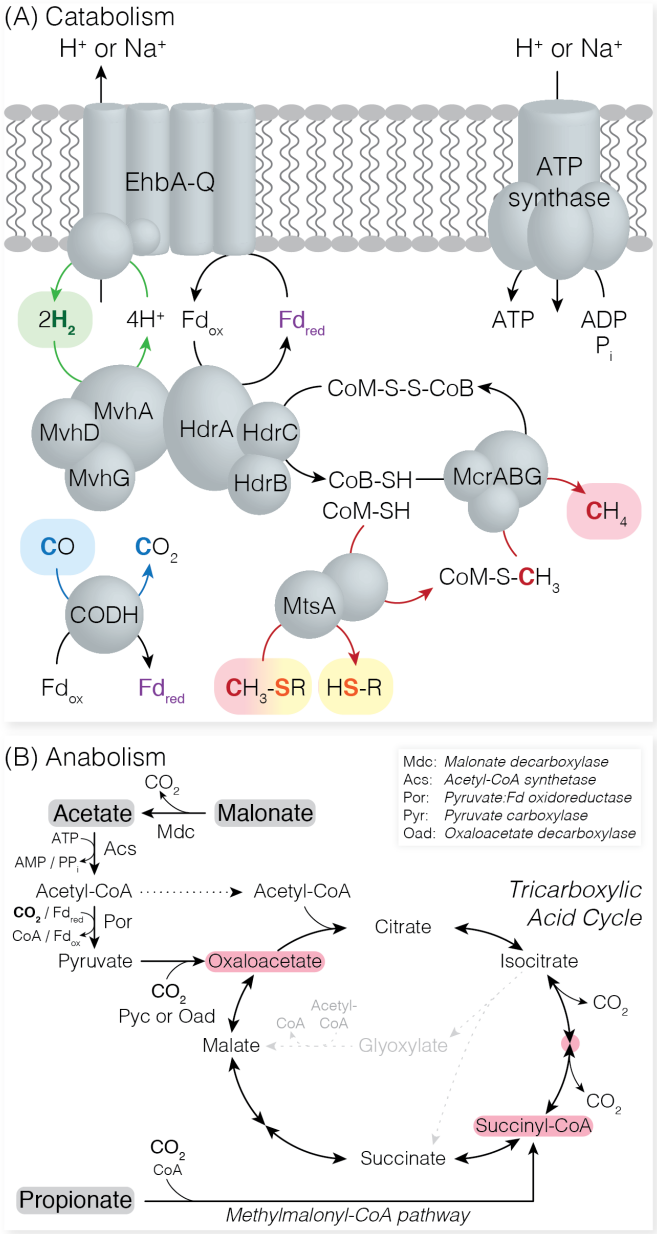

Figure 4.4. WSA2 (A) catabolism and (B) anabolism. (A) WSA2 has genes for H_2 oxidation through electron-bifurcating hydrogenase (HdrABC-MvhDGA) and H_2 cycling by energy-converting hydrogenase (EhbA-Q); CO oxidation by carbon monoxide dehydrogenase (CODH); and methylated thiol reduction and methanogenesis by methylated thiol Coenzyme M methyltransferase corrinoid fusion protein (MtsA) and methyl coenzyme M reductase (McrABG). The proton motive force (or cation gradient) generated by Ehb can support ATP production by ATP synthase. (B) Malonate decarboxylase and acetyl-CoA synthetase can convert malonate and acetate into acetyl-CoA for downstream co-assimilation with CO_2 (bolded) through pyruvate:ferredoxin oxidoreductase, pyruvate carboxylase, and tricarboxylic acid (TCA) cycle. As identified for other heterotrophic methanogens, WSA2 does not encode the glyoxylate shunt for acetate assimilation (gray with dotted line). The methylmalonyl-CoA pathway can facilitate co-assimilation of propionate and CO_2 also into the TCA cycle. WSA2 can use key TCA cycle intermediates (pink) as building blocks for biosynthesis.

CHAPTER 5

MICROBIAL DARK MATTER ECOGENOMICS REVEALS COMPLEX SYNERGISTIC NETWORKS IN A METHANOGENIC BIOREACTOR[3]

5.1 – Abstract

Ecogenomic investigation of a methanogenic bioreactor degrading terephthalate (TA) allowed elucidation of complex synergistic networks of syntrophs, methanogens, and uncultivated microorganisms, including those from candidate phyla with no cultivated representatives. Our previous metagenomic investigation proposed that *Pelotomaculum* and methanogens may interact with uncultivated organisms to degrade TA; however, many members of the community remained unaddressed due to past technological limitations. In further pursuit, this study employed state-of-the-art omics tools to generate draft genomes and transcriptomes for uncultivated organisms spanning 15 phyla and reports the first genomic insight into candidate phyla Atribacteria, Hydrogenedentes, and Marinimicrobia in methanogenic environments. Metabolic reconstruction revealed that these organisms perform fermentative, syntrophic, and acetogenic catabolism facilitated by energy conservation revolving around H_2 metabolism. Several of these organisms could degrade TA catabolism byproducts (acetate, butyrate, and H_2) and syntrophically support *Pelotomaculum*. Other taxa could scavenge anabolic products (protein and lipids) presumably derived from detrital biomass produced by the TA-degrading community. The protein scavengers expressed complementary metabolic pathways indicating syntrophic and fermentative step-wise protein degradation through amino acids, branched-chain fatty acids, and propionate. Thus, the uncultivated organisms may interact to form an intricate syntrophy-supported food web with *Pelotomaculum* and methanogens to metabolize catabolic byproducts and detritus, whereby facilitating holistic TA mineralization to CO_2 and CH_4.

5.2 – Introduction

In methanogenic environments, fermenters, syntrophs, and methanogens interact to facilitate anaerobic degradation of organic compounds to CH_4 and CO_2, an essential component of anaerobic biotechnology and natural global carbon cycle (*2*). Fermenters and syntrophs degrade organic compounds to methanogen-utilizable substrates (e.g., acetate and H_2) and methanogens further convert these

[3] This chapter appeared in its entirety in the published article: Nobu MK, Narihiro T, Rinke C, Kamagata Y, Tringe SG, Woyke T *et al* (2015). Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J* **9:** 1710-1722. (doi: 10.1038/ismej.2014.256). The journal does not require authors of original research papers to assign copyright of their published contributions.

compounds to CH_4 and CO_2. Syntrophs specifically form obligate mutualistic interactions with methanogens to metabolize organic compounds whose degradation otherwise rapidly becomes thermodynamically unfavorable ($\Delta G>0$) as byproducts accumulate (*76, 132*). Despite such low-energy conditions, engineered an natural methanogenic environments harbor a wide diversity of uncharacterized microorganisms from known phyla and candidate phyla without cultivated representatives (microbial dark matter (*18, 19, 87, 217-223*). Their presence indicates many undiscovered microbial niches and interactions. Understanding the roles and interactions of the ubiquitous uncultivated microorganisms is necessary to unravel the black-box microbial ecology of anaerobic biotechnology and the global carbon flux.

An excellent example of a methanogenic community rich in uncultivated taxa is a hypermesophilic bioreactor (46-50°C) treating terephthalate (TA), a major waste product of plastics manufacturing (*87*). *Pelotomaculum* and *Syntrophorhabdus* are known to syntrophically metabolize TA (*69, 78, 82, 88*). In this reactor specifically, *Pelotomaculum* syntrophically interacts with methanogens and predominates TA degradation (*85*). Our recent metagenomic and proteomic studies proposed that *Pelotomaculum* may produce acetate, H_2, and butyrate as a byproduct and that uncultivated taxa (*Caldiserica* and "*Ca.* Cloacimonetes", formerly OP5 and WWE1) may further metabolize the H_2 and butyrate (*87, 110*). This provided evidence that non-methanogen community members may serve as syntrophic partners (secondary degraders) to metabolize TA degradation byproducts alongside aceticlastic and hydrogenotrophic methanogens. However, the sequencing technology of the time limited investigation of the other uncultivated community members' ecological roles. Moreover, elucidation of their ecological roles is paramount as many of these taxa are consistently found across methanogenic bioprocesses and their contribution to anaerobic biotechnology remains unclear (*224*).

Although the persistent uncultivated organisms in methanogenic bioprocesses have long been suspected to contribute to anaerobic carbon flow, their fastidiousness to has hindered exploration. In the TA-degrading community, some non-methanogenic taxa may also support TA degradation by metabolizing *Pelotomaculum* byproducts (i.e., acetate, butyrate, or H_2). Besides such catabolic byproducts, other uncultivated taxa may scavenge TA metabolism anabolic products (*e.g.*, protein and lipids) bound in detrital biomass, which is estimated to account for up to 10% of the TA-derived carbon (*14*). Based on the known metabolic diversity of methanogenic environments (*2, 132*), uncultivated secondary degraders and scavengers must perform fermentative, syntrophic, or acetogenic metabolism. To investigate the ecological roles and metabolic capabilities of the uncultivated taxa, this study used an 'ecogenomics' approach synthesizing single-cell genomics (*225*), metagenomics (*39*), metatranscriptomics (*226*), and metabolic reconstruction assisted by comparative genomics of acetogen and syntroph energy conservation pathways (*12, 27, 35, 227, 228*). In doing so, we provide the first look into the genomes and metabolism of many

uncultivated taxa and propose how these organisms metabolically interact to achieve holistic carbon flux from TA to CH_4 and CO_2 in the methanogenic reactor.

5.3 – Materials and Methods

5.3.1 – Metagenome, single-cell amplified genomes, and metatranscriptome

Samples for metagenomic DNA, single-cell sorting, and metatranscriptomic RNA were taken from biofilm carriers (*i.e.*, porous ceramic media for biofilm attachment) in a hypermesophilic (46-50°C) bioreactor degrading TA at >80% efficiency after 1373 days of operation. The 1.2L hybrid reactor was packed with tubular ceramic carrier (Siporax) and fed TA as the sole energy source (3.6 g TA L^{-1} day^{-1}) in an anaerobic mineral culture medium (*85, 87*). For single-cell collection, biofilms were separated from the carrier, dispersed through sonication, and sorted through flow cytometry. After DNA extraction and whole genome amplification of the sorted cells, cells of target clades were screened through 16S-rRNA-gene-targeted polymerase chain reaction and amplicon sequencing (*217*). For target clades, amplified genomic DNA was sequenced to generate single amplified genomes (SAGs) (Appendix D Table D.2). Metagenomic DNA was extracted from dispersed biofilm using a cetyltrimethylammonium bromide-based protocol (Joint Genome Institute). Paired-end reads (100bp, $3.77x10^8$ reads) sequenced using a HiSeq-2000 sequencer (Illumina) were trimmed and assembled as detailed in the Supplementary Note.

For constructing a metatranscriptome, total RNA was extracted from the TA-degrading reactor using mirVana RNA isolation kit (Ambion, USA) and purified by removing DNA through the Turbo DNA-free kit (Ambion, USA). Ribosomal RNA was also removed from the total RNA sample using a RiboMinus bacteria transcriptome isolation kit (Invitrogen, CA) and the enriched messenger RNA fraction was treated with an mRNA-Seq kit for downstream RNA-Seq analysis (Illumina Inc., CA). Only fragments between 200 and 400bp were used for final sequencing with a Genome Analyzer IIx using a single lane (Illumina Inc., CA). Metatranscriptomic sequences (available on MG-RAST under 4447821.3) were mapped to the metagenome and SAGs using blastn with a cut-off of 100% sequence identity.

5.3.2 – Phylogenetic binning

Four methods were compared for phylogenetically annotating metagenomic scaffolds and, ultimately, constructing accurate TA-degrading community member draft genomes: PhylopythiaS, ClaMS, Metawatt version 1.7, and BLAST (*50, 51, 215, 229*). Metagenomic bins generated from these k-mer- and homology- based methods (detailed in Appendix D Supplementary Information) were compared to attain accurate classification of contigs containing novel genes that BLAST cannot address. For comparison of the constructed bins, Multibase 2014 (Numberical Dynamics) was used to perform principal component

analysis (PCA) based on k-mer frequency. Tetramer frequencies were calculated for scaffolds longer than 1000 bp. Scaffolds longer than 10 kb were split into fragments with a minimum length of 5000 bp. Within each tetramer, the mean tetramer frequency was adjusted to zero and the other values were shifted accordingly. The frequencies were normalized to the maximum observed frequency across all tetramers. The PCA results were visualized as a three-dimensional scatter plot with principal component 1, principal component 2, and scaffold read coverage as axes using TOPCAT version 4.0-1 (*230*).

5.4 – Results and Discussion

5.4.1 – Single-cell amplified genomes and metagenome

After 1373 days of operation with stable TA degradation (Appendix D Fig. D.1), biofilms were sampled from carriers (*i.e.*, biofilm attachment media) for downstream metagenomics and flow-cytometry-based cell sorting. After screening sorted cells through 16S rRNA sequencing from whole-genome-amplified DNA, 32 cells were selected for single amplified genome (SAG) sequencing, as described previously (*217*). This generated 32 single amplified genomes (SAGs) to provide genomic insight into uncultivated clades, support phylogenetically annotating ("binning") metagenomic contigs, and improve genome coverage. Shotgun-sequencing of the TA-degrading community and assembly generated a metagenome of 205310 contigs totaling to 217Mb sequence with an N50 of 8452bp. EMIRGE-based construction of 16S rRNA gene sequences from the raw metagenomic reads revealed that the TA-degrading community is 39% *Bacteria* and 61% *Archaea* (Appendix D Table D.1). Within the bacterial community (Supplementary Fig. S1), *Pelotomaculum*- (16.8% of bacterial community) and *Syntrophorhabdus*- (6.1%) related members were identified as syntrophs most likely contributing to TA degradation (Fig. 2). In the archaeal community, both aceticlastic (*Methanosaeta*, 51.0% of archaeal community) and hydrogenotrophic (39.4% *Methanolinea* and 6.3% *Methanospirillum*) methanogens were identified to syntrophically support TA degradation through respective acetate and H_2 degradation. Besides these primary TA degraders and methanogens, the community contained "*Ca.* Marinimicrobia" (SAR406; 19.9%), *Planctomycetes* (15.8%), non-*Syntrophorhabdaceae Deltaproteobacteria* (5.2%), *Synergistetes* (8.5%), *Thermotogae* (5.4%), *Spirochaetes* (3.0%), "*Ca.* Cloacimonetes" (WWE1; 3.3%), *Chloroflexi* (2.6%), *Caldiserica* (1.9%), *Betaproteobacteria* (0.7%), "*Ca.* Aminicenantes" (OP8; 0.9%), "*Ca.* Hydrogenedentes" (NKB19; 0.8%), "*Ca.* Atribacteria" (OP9; 0.8%), candidate division AC1 (0.7%), and "*Ca.* Acetothermia" (OP1; 0.6%) at abundance >0.5% of the bacterial community (Appendix D Fig. D.1 and Table D.1). Thus, we detect the same clades as our previous study (*87*) even though the community structure slightly differs possibly due to both increase in TA loading rate and discrepancy between 16S rRNA gene-based PCR – dependent (*i.e.*, pyrosequencing) and –independent (*i.e.*, EMIRGE) analyses.

To generate high quality draft genomes for these organisms, four binning methods (BLAST, PhylopythiaS, Metawatt, and ClaMS) were compared. Although BLAST was precise, it only binned contigs that overlapped with reference genome sequences, which are incomplete SAGs for many clades. To maximize genome coverage for target clades, k-mer frequency binning methods can annotate non-overlapping contigs. Accuracy of these methods was measured by evaluating the agreement with BLAST binning results. PhylopythiaS and Metawatt agreed better with BLAST than ClaMS (Appendix D Table D.3). Although PhylopythiaS generally had higher BLAST-extrapolated accuracy, its results were complemented with results from Metawatt for taxa poorly binned by PhylopythiaS. Manual curation of these binning results (Fig. 5.1A) using essential single copy genes, k-mer frequency distribution, read coverage, and BLAST generated 26 high quality draft genomes spanning over 13 phyla (Fig. 5.1B, Fig. 5.2, and Appendix D Table D.2).

Combining SAGs and bins derived from the metagenome generated 35 draft genomes and six pangenomes (compiled bins and SAGs with >98% similarity of overlapping contigs) averaging 70% genome completeness and spanning 24 putative genera over 15 phyla, which provided coverage for approximately 90% of the TA-degrading bacterial community (Fig. 5.2 and Appendix D Table D.2). (Pan)genomes with at least 40% genome completeness (es as described in Supplementary Note) and metatranscriptomics-based gene expression profiles available were analyzed (12 SAGs, 23 non-SAG draft genomes, and six pangenomes) (Appendix D Table D.2). As an improvement to our previous study, we constructed multiple population-level genomes for methanogens and TA-degrading syntrophs: *Methanolinea* (3 genomes), *Methanosaeta* (4), *Methanospirillum* (2), *Pelotomaculum* (3), *Syntrophorhabdaceae* (3) (Fig. 5.2 and Appendix D Table D.1 and D.2). This study also constructed much more complete genomes of *Pseudothermotoga*, *Mesotoga*, *Caldiserica* 31q06, and Cloacimonetes. Moreover, this effort generated genomes for taxa that our previous study could not address, including uncultivated phyla Atribacteria, Hydrogenedentes, Marinimicrobia, and WS1; poorly understood phyla *Armatimonadetes* and *Chloroflexi*; and other uncultivated phylogenetic branches of novel *Caldisericia* sister clade, *Syntrophus*-related clade, *Spirochaetes* SA-8, *Synergistetes* TTA-B6, *Planctomycetales*-related clade, and *Phycisphaerae*-related cluster WPS-1 (*18, 231, 232*). Of these clades, pangenomes were constructed for Atribacteria, Cloacimonetes, Hydrogenedentes, Marinimicrobia, *Chloroflexi*, and *Syntrophus*-related clade. Population abundance was estimated for each clade (Appendix D Supplementary Information, Table D.1 and Table D.2).

5.4.2 – Metatranscriptomics

The 25 million read (2.0 Gb) metatranscriptome contained 78.6% rRNA and 21.4% putative mRNA reads. Blastn mapped 16.6% and 2.1% of the mRNA reads to the assembled metagenome and SAGs

respectively using a 100% sequence similarity cut-off. The final draft genomes (combined SAGs and bins) accounted for 74.4% of these mapped reads. Methanogens accounted for the majority (75.6%) of the mapped data (64.4% *Methanolinea*, 5.7% *Methanospirillum*, and 5.5% *Methanosaeta*) presumably due to the significantly larger net energy yield from methanogenesis as compared to TA degradation, especially considering that 3 mol acetate and H_2 are generated from 1 mol TA. Amongst the bacterial clades, most reads were mapped to *Pelotomaculum* (31.7%), Marinimicrobia (22.8%), *Mesotoga* (11.3%), Hydrogenedentes (10.1%), Cloacimonetes (6.9%), *Syntrophorhabdaceae* (4.1%), *Phycisphaerae*-related cluster WPS-1 (3.8%), *Syntrophus*-related clade (3.1%), and *Planctomycetales* (2.1%). Atribacteria, *Chloroflexi*, *Caldiserica*, *Spirochaetes*, *Synergistetes*, and *Pseudothermotoga* members each only accounted for 0.1-0.7% of the mapped sequences; however, this does not necessarily preclude their contribution for complete TA degradation.

5.4.3 – Redefining syntroph and acetogen energy conservation: a comparative genomics approach

To rigorously determine the catabolic capacity of target organisms through ecogenomics, this study capitalizes on the complementarity of catabolism and energy conservation unique to methanogenic environments. Metabolism under methanogenic conditions faces two types of obstacles: thermodynamically unfavorable electron disposal and energy acquisition from limited energy margins (*117, 133, 233*). While H^+ and CO_2 are the only available exogenous electron acceptors (*e.g.*, syntrophy), coupling re-oxidation of the general physiological electron carrier, NADH ($E^{0'}$ = -320 mV), with respective reduction of these compounds to H_2 and formate ($E^{0'}$ = -414 mV and -420 mV) is thermodynamically unfavorable. To drive such endergonic reactions, anaerobic organisms are thought to employ energy from proton motive force (reverse electron transport, RET), oxidation of a low-potential donor (electron confurcation), and reduction of a high-potential acceptor (electron bifurcation), often taking advantage of a high-energy electron carrier, ferredoxin (Fd, $E^{0'}$ = -453 mV). Conversely, for catabolic H_2- and formate- oxidation (*i.e.*, homoacetogenesis and methanogenesis), exergonic NAD^+ reduction would sacrifice marginal energy yield while Fd_{ox} reduction is endergonic. To resolve this issue, acetogens and methanogens are thought to employ similar energy conservation strategies. As such, metabolism under methanogenic conditions necessitates complementation of substrate oxidation with electron balance and energy conservations and, thus, we genomically explore the metabolic capacity of target organisms based on this principle.

While fermenters and methanogens have been extensively studied across a wide phylogenetic range, insight into acetogen and syntroph energy conservation still remains limited. In order to better understand energy-conserving mechanisms employed for acetogenesis and syntrophic metabolism, we survey all publically available homoacetogen and syntroph genomes (Appendix D Supplementary Information).

Many syntrophs appear to rely on the capacity to perform RET-driven energy-conserving H_2 generation through electron-confurcating hydrogenase (ECHyd) in combination with Fd_{red}-generating Rnf (NADH:Fd oxidoreductase) or Hdr-Ifo (heterodisulfide-reductase-associated ion-translocating Fd:NADH oxidoreductase) (*12, 228*). Specifically, the syntroph-characteristic co-occurrence of Hdr-Ifo and ECHyd can serve as a good indicator for syntrophic capacity. Similarly, co-occurrence of the electron-transfer-flavoprotein-oxidizing hydrogenase (FixABCX) (*35, 234*) and carboxylate catabolism with ECHyd and Rnf or Hdr-Ifo can further implicate ability to degrade carboxylates syntrophically. As for syntrophs who do not encode Rnf and Hdr-Ifo or ECHyd, this survey reveals other RET and electron bi(con)furcation mechanisms that syntrophs may employ. For acetogenesis, we identify that acetogens require (*235*) and encode at least one energy-conserving electron transfer between physiological electron carriers (Rnf, NAD(P)H transhydrogenase, and/or NADH-dependent Fd_{red}:$NADP^+$ oxidoreductase) (*227, 236-238*); often possess a putative clostridial sensory hydrogenase (*239*); and utilize a wide variety of hydrogenases and formate dehydrogenases. In addition, we identify five genotypes of acetogen MetF (methylene-tetrahydrofolate reductase) essential for driving the homoacetogenic Wood-Ljungdahl pathway, but also postulate that undiscovered MetF homologs and genotypes must exist. Finally, in order to characterize syntrophic or acetogenic metabolism, it is paramount that we detect no electron disposal pathways for fermentative or respiration (*i.e.,* non-H^+/CO_2 electron acceptor). Using these findings as a foundation for genomic and metabolic properties of syntroph and acetogen, we genomically characterize syntrophic and acetogenic capacity of target taxa found in the TA-degrading reactor based on complementation between catabolic pathways and the observed energy conservation strategies.

5.4.4 – Syntrophic TA-degradation by *Pelotomaculum* and *Syntrophorhabdus* with methanogens

Organisms of the genus *Pelotomaculum* and family *Syntrophorhabdaceae* are thought to be responsible for syntrophic TA degradation in reactor systems (*87, 110*). We identify diverse *Pelotomaculum* (bins TAPelo1–TAPelo4) predominate the community, accounting for 16.8% and 31.7% of the bacterial community and metratranscriptome. These *Pelotomaculum* indeed encode and express genes for syntrophic energy conservation (Hdr-Ifo and ECHyd) and a previously observed pathway for TA degradation to acetate, CO_2, and H_2 ((*27, 87*), Fig. 5.2, and Appendix D Table D.4). In addition, we newly identify expression of a clostridial electron-bifurcating butyryl-CoA dehydrogenase in TAPelo3 that may facilitate the previously hypothesized *Sporotomaculum*-like energy-conserving butyrate generation from aromatic compound degradation (*79, 110, 116, 228*), albeit refuting the involvement of previously identified non-energy-conserving acyl-CoA dehydrogenase (*87, 110*). Although butyrate-fermenting TA degradation is thermodynamically more favorable, it sacrifices precious ATP from substrate-level phosphorylation. Our

current understanding of syntrophic TA degradation precludes determination of the energetic feasibility of butyrate fermentation, but it is plausible that butyrate serves as a supplementary electron sink during high H_2 partial pressure. While members of *Syntrophorhabdaceae* (bins TASrha1, TASrha2, and TASrha3) also encode and express syntrophic TA metabolism through Hdr-Ifo and ECHyd (Fig. 5.2 and Appendix D Table D.4), they comprise a smaller fraction of the bacterial community (6.1%) and metatranscriptome (4.1%). Thus, they may compete with *Pelotomaculum* in syntrophic TA degradation, but have a lesser contribution to total TA removal.

Supporting this TA degradation, we identify the methanogenic partners: aceticlastic *Methanosaeta* (30.9% of total community; 5.5% of metatranscriptome) and hydrogenotrophic *Methanolinea* (23.9%; 64.4%) and *Methanospirillum* (3.8%; 5.7%). We confirm that these organisms indeed express acetate- and H_2- degrading methanogensis pathways (Appendix D Table D.5). Although *Methanolinea* is less abundant than *Methanosaeta*, *Methanolinea* constitutes the majority of methanogen metatranscriptome and, perhaps, a larger proportion of the methanogenic activity. This discrepancy is in agreement with both the higher energy yield from hydrogenotrophic methanogenesis and also suggests high importance of H_2 oxidation in driving syntrophic TA degradation.

5.4.5 – *Thermotogae*: unconventional syntrophic acetate degraders

Thermotogae, a phylum often found in wastewater treatment ecosystems (*224*), may be responsible for syntrophic acetate degradation, as observed for *Pseudothermotoga lettingae* strain TMO (*240*); however, the metabolic pathway remains unclear due to absence of acetyl-CoA synthase / CO dehydrogenase required for conventional syntrophic acetate catabolism through the Wood-Ljungdahl pathway (*34*). Interestingly, the TA-degrading community *Mesotoga* (bin TAMoga) encodes a potential alternative pathway mediated by the glycine cleavage system and tetrahydrofolate pathway (Appendix D Supplementary Information, Fig. D.2, and Table D.6), and H_2-generating energy conservation complementary to syntrophic carboxylate degradation (*i.e.*, FixABCX, Rnf, and ECHyd) (Fig. 5.2). Further, strain TMO and another community *Thermotogae* member (*Pseudothermotoga*, bin TAPoga) and also encode the novel acetate degradation pathway with Rnf and ECHyd. Correspondingly, although the community *Thermotogae* member was previously inferred to oxidize butyrate based identification of a butyryl-CoA dehydrogenase (*87*), we did not detect a complete butyrate degradation pathway. Thus, the *Mesotoga* and *Pseudothermotoga* members may syntrophically oxidize acetate through a previously uncharacterized acetate-oxidizing pathway. While *Pseudothermotoga* has very low activity, the high *Mesotoga* activity (11.3% of the bacterial metatranscriptome or 2.8% of total) is comparable to aceticlastic

Methanosaeta, which suggests that *Mesotoga* may specifically play a critical role in TA degradation by catabolizing acetate, one of the major byproducts of TA metabolism.

5.4.6 – Novel *Syntrophaceae* member: a versatile syntrophic fatty acid degrader

Our previous study (*87*) suspected that the *Syntrophus*-related clade (93.8% 16S rRNA similarity to *S. aciditrophicus* strain SB), composing 1.2% of the bacterial community and 3.1% total metatranscriptome, may also contribute to aromatic compound degradation as characteristic of *Syntrophus* (*81*). While its genome lacks the benzoate degradation pathway, we identify expression of Hdr-Ifo, ECHyd, and FixABCX (Fig. 5.2, Fig. 5.3, and Appendix D Table D.2), which suggests the capacity to syntrophically degrade carboxylates. In agreement, this organism expresses a butyrate degradation pathway similar to *Syntrophomonas*, which also depends on identical energy conservation pathways ((*35, 228*) and Appendix D Table D.7). In addition, this clade expresses newly postulated syntrophic branched-chain fatty acid (*i.e.*, 2-methyl-butyrate, isovalerate, and isobutyrate) degradation pathways that are consistent with previous cultivation-based studies (*241-243*) and share high homology with other organisms thought to degrade branched-chain fatty acids (Appendix D Supplementary Information, Fig. D.5, and Table D.7). Therefore, this *Syntrophus*-related clade likely performs syntrophic degradation of butyrate and branched-chain fatty acids (discussed later). We propose that this clade may support TA degradation through syntrophically metabolizing TA-derived butyrate. In agreement, its transcript-level is much lower than *Pelotomaculum* and yet comparable to *Syntrophorhabdaceae* despite low abundance in the community.

5.4.7 – An uncultivated *Chloroflexi* subphylum I member: a novel homoacetogen

While recent studies cultivate many *Chloroflexi* subphylum I representatives (*231*), much of this phylogenetic cluster remains to be characterized. For example, the TA-degrading community *Chloroflexi* is distantly related from the closest relatives, *Anaerolinea thermophila* (84.0% rRNA similarity) and *Caldilinea aerophila* (84.8%). Surprisingly, the pangenome of this *Chloroflexi* subphylum I member uniquely encodes the complete Wood-Ljungdahl pathway required for homoacetogenesis (Appendix D Table D.8). Furthermore, we identify Rnf, NfnAB (NADH-dependent Fd_{red}:$NADP^+$ oxidoreductase), and HfsABC (putative sensory hydrogenase) that are conserved amongst many acetogens and are paramount for acetogenic metabolism (Appendix D Supplementary Information and Table D.12). The *Chloroflexi* MetF, a critical gene for acetogenesis, is most likely a new type of MetF as it only shares low homology with type III acetogen MetF (<35% amino acid similarity) and lacks MetV necessary for type III MetF function; however, discovery of a functionally novel MetF is reasonable because we have yet to discover MetF homologs for several known acetogens and characterize an acetogenic *Chloroflexi*. Thus, this *Chloroflexi*

may be capable of H_2-oxidizing homoacetogenesis, whereby supporting *Pelotomaculum* TA degradation in parallel with other H_2-oxidizers (*i.e.*, *Methanolinea* and *Methanospirillum*). However, further experimental investigation is necessary as low genome coverage and gene expression levels limit definitive analysis. Although our previous study proposed that *Caldiserica* cluster 31q06 may perform acetogenic metabolism, reevaluation of the genes reveals that this is not likely (Appendix D Supplementary Information).

5.4.8 – "*Ca.* Hydrogenedentes" (formerly NKB19): a lipolytic glycerol degrader

Hydrogenedentes is often associated with methanogenic environments (*18, 224*), but its ecological role has remained enigmatic. While the genomic technology available for our previous study was insufficient for metagenomically capturing Hydrogenedentes, implementing modern sequencing platforms allowed construction of a Hydrogenedentes pangenome and transcriptome. These data reveal expression of extracellular lipases, type II secretion systems, and the Sec system for extracellular hydrolysis of triacylglycerols to glycerol and long-chain fatty acids. This organism also expresses genes for syntrophically oxidizing glycerol to acetate, CO_2, 2 NADH, and Fd_{red} and rather ECHyd than fermentative pathways for following electron disposal through H_2-generation ((*244*); Fig. 5.2; Appendix D Fig. D.3 and Table D.9). As another NADH re-oxidation mechanism, we postulate that Hydrogenedentes performs novel NADH-oxidizing H_2 generation mediated by Na^+-translocating NADH:quinone oxidoreductase, cytochrome bc fusion protein, and periplasmic Fe-hydrogenase (Appendix D Supplementary Information). In agreement, this organism expresses many Na^+-transporting complexes and, thus, may rely on a transmembrane Na^+ gradient metabolically and physiologically. We also identify expression of a novel *Methanothermobacter* electron-bifurcating hydrogenase-like gene cassette encoding MvhADG hydrogenase subunits and HdrABC ((*122*); Appendix D Table D.9); however its function remains unclear. Although the complete energy conservation scheme requires further investigation, we infer that Hydrogenedentes syntrophically degrades glycerol and lipids derived from detrital biomass. Despite the low Hydrogenedentes population abundance (0.8%), the strikingly high expression level (10.1% of bacterial transcriptome) suggests that Hydrogenedentes lipolysis and glycerol degradation is an important component of this TA-degrading community carbon flux. Although another clade, *Planctomycetales*-related clade (bin TAPire), also expresses similar lipolytic non-fermentative glycerol degradation (Appendix D Table D.9), its low genome completeness prohibits accurate interpretation.

5.4.9 – "*Ca.* Marinimicrobia" (formerly SAR406): a proteolytic amino acid degrader

The ecological role of this typically marine-associated phylum, Marinimicrobia, in methanogenic ecosystems has remained unclear (*224*); however, modern sequencing technology has allowed us to peer

into its genome, which could not be addressed in our previous study. We discover expression of extracellular proteases and secretory pathways for extracellular proteolysis (Appendix D Table D.10). For degrading the proteolysis-derived monomers, this organism expresses pathways for catabolizing amino acids that are strictly syntrophic (*i.e*, Ile/Leu/Val/Pro) and also fermentable (*i.e*., Gly/Ser/Thr/Glu/Asp/Asn) ((*133*); Appendix D Table D.10). More specifically, as observed in other syntrophs (*245-248*), Ile/Leu/Val and Pro are catabolized to branched chain fatty acids and propionate respectively, along with H_2 and CO_2 (Appendix D Fig. D.4 and Table D.10). To support H_2-generating electron disposal, Marinimicrobia expresses Hdr-Ifo, Rnf, and ECHyd (Fig. 5.2). As for the fermentable amino acids, Gly/Ser are degraded to H_2, CO_2, and acetate, while Thr/Glu/Asp/Asn are degraded to either H_2, CO_2, and acetate or propionate. Interestingly, Marinimicrobia also utilizes the uncharacterized bacterial MvhADG-HdrABC found in Hydrogenedentes, suggesting that this complex may facilitate a novel energy conservation pathway found in uncultivated taxa. Based on these findings, we propose that Marinimicrobia syntrophically and fermentatively degrades amino acids through proteolysis of protein bound in the community detritus. Similar to Hydrogenedentes, Marinimicrobia also dictates a large portion of the bacterial metatranscriptome (22.8%), suggesting that Marinimicrobia proteolysis and amino acid degradation is also a significant component of the TA-degrading community carbon flux. Although to a lesser extent, the member of *Phycisphaerae*-related uncultivated cluster WPS-1 (bin TAPhyc, 3.8% of metatranscriptome) may also contribute to proteolytic syntrophic (Ile/Leu/Val) and fermentative (Glu) amino acid degradation mediated by Hdr-Ifo and ECHyd, based on the expressed pathways (Fig. 5.2; Appendix D Table D.10). It is currently unclear why these proteolytic organisms are highly abundant and active in the community. *Caldisericia* sister clade (bin TACald), *Aminiphilus*-related clade (bin TAAmin), *Synergistetes* cluster TTA-B6 (SAG D21), and *Spirochaetes* clade SA-8 (TASpir) members also harbor proteolytic AA metabolism pathways, albeit with low transcriptomic coverage (Appendix D Table D.11).

5.4.10 – "*Ca.* Atribacteria" and "*Ca.* Cloacimonetes": syntrophic propionate degraders

This ecogenomic effort also generated a pangenome for Atribacteria, another poorly characterized candidate phylum associated with methanogenic environments (*224*) that our previous study could not address. This Atribacteria encodes propionate metabolism and specifically expresses methylmalonyl-CoA pathway genes with high homology (52-71%) to those found in *Pelotomaculum thermopropionicum* strain SI, a representative thermophilic propionate-degrading syntroph ((*29*); Appendix D Table D.11). For energy-conserving electron disposal, Atribacteria expresses an electron-bifurcating formate dehydrogenase (*249*) and Fd_{red}-dependent energy-conserving membrane-bound hydrogenase (Fig. 2 and Fig. 3), suggesting that both formate and H_2 generation may take part in propionate degradation as observed in

Syntrophobacter (*250*). Moreover, we identify that electron-bifurcating formate dehydrogenase is also encoded by *Pelobacter carbinolicus* (Pcar_1846-1843), a syntroph dependent on formate transfer (*251, 252*). While propionate metabolism can be used in the reverse direction fermentatively, we did not detect any pathways for catabolizing fermentative substrates. Similarly, Cloacimonetes also only expresses propionate metabolism and *Pelotomaculum*-like methylmalonyl-CoA pathway (51-57% similarity) and energy-conserving H_2 production through Rnf and ECHyd (Fig. 5.2 and Fig. 5.3). Thus, Atribacteria and Cloacimonetes may perform syntrophic propionate metabolism. Compiling the amino acid, branched chain fatty acid, and propionate degradation genes expressed by the uncultivated organisms forms a comprehensive pathway metabolizing amino acids to methanogen-utilizable substrates. Based on this striking metabolic complementarity in degradation of these substrates, we speculate that chaining syntrophic interactions (secondary syntrophy) and substrate exchange may facilitate proteinaceous detritus metabolism. However, further co-culture-based studies are necessary to investigate the nature and motivation behind such compartmentalization of syntrophic and fermentative degradation in methanogenic environments.

5.5 – Conclusion

Implementation of cutting-edge ecogenomics and insights into anaerobic metabolism allowed a much more robust investigation of the TA degrading community microbial dark matter hidden behind the binary *Pelotomaculum*-methanogen syntrophic degradation. This study not only provided novel genomic and metabolic insights into several uncultivated phyla (Hydrogenedentes, Marinimicrobia, Atribacteria, and Cloacimonetes) and clades (within *Thermotogae*, *Syntrophaceae*, *Chloroflexi*, *Planctomycetes,* and *Caldiserica*) associated with methanogenic environments, but also shed light on their potential ecological roles and interactions. While several uncultivated taxa may support TA degradation as secondary degraders metabolizing *Pelotomaculum* catabolism byproducts (acetate, butyrate, and H_2), many others may contribute to scavenging detritus accounting for up to 10% of the degraded TA (*14*) through macromolecule hydrolysis, fermentation, and chaining syntrophic degradation of glycerol, amino acids, and branched-chain fatty acids (Fig. 4). Therefore, the holistic carbon flux from TA to CH_4 and CO_2 may require primary TA degraders, secondary degraders, detritus scavengers, and methanogens to form "syntrophic networks" beyond the conventionally studied binary syntrophy. Although full validation of the proposed carbon flow model further requires effortful cultivation-based studies tracking the specific niches in the TA-degrading community, we believe that metabolic characterization of the uncultivated taxa provides a valuable step forward in improving our insight into the potential roles of these organisms in methanogenic environments and ecological synergy (*i.e.*, secondary syntrophy) facilitating degradation at the thermodynamic limit. Last,

we propose *Candidatus* species designations (Supplementary Note) to reflect the newly discovered behavior of those elusive organisms.

5.6 – Figures

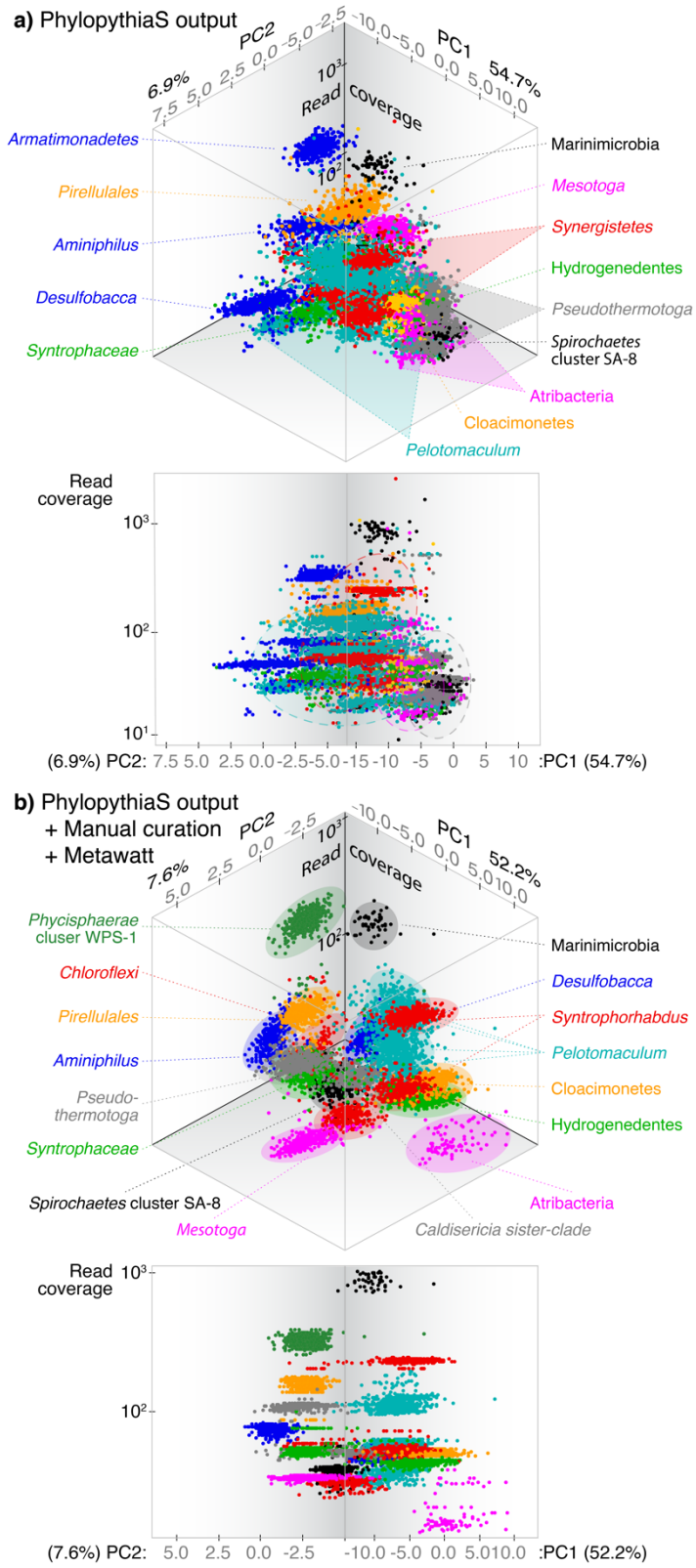

Figure 5.1. Comparison of PhylopythiaS and manually curated metagenome binning results. Principal component analysis on tetramer frequency of metagenomic contigs (>1kb) binned through (a) PhylopythiaS and (b) combined PhylopythiaS, Metawatt, and manual curation are shown with taxonomic bins differentiated by color. Angled (top) and front-face (bottom) views are shown for plots with principal components (PC1 and PC2) and contig read coverage on the horizontal and vertical axes respectively. For the raw PhylopythiaS output, several bins had high read coverage ranges (highlighted with dotted outlines) indicating poor bin quality. Relative to (a) the raw output, (b) manual curation allowed clearer separation of metagenomic bins. Some bins were manually taxonomic reclassified based on 16S rRNA gene phylogeny and BLAST.

Figure 5.2. TA reactor biofilm microbial community bin and SAG phylogeny and key anaerobic energy conservation pathways. The 16S rRNA gene-based phylogenetic tree (bootstrap 1000: >90% black node, >70% gray node with black outline, and >50% gray node) contains sequences from single-cell genomes, analyzed bins (bolded, blue if methanogenic, and red or turquoise based on identified energy conservation pathways detailed below), and related isolates and clones (gray) with phyla distinguished by alternating background color. Classification of target taxa and their proposed *Candidatus* names are shown to the right of the tree. For each taxon, the genome completeness of SAG (grey), binned draft genome (black), and pangenome (black with *) are shown as pie charts. The following right-hand columns indicate the presence (circle) or absence (blank) of specific genes related to general (turquoise) and syntroph-associated (red)

Figure 5.2. (cont.) energy conservation. General energy conservation includes electron-confurcating hydrogenase (ECHyd), *Rhodobacter* nitrogen fixation complex (Rnf), and electron-bifurcating formate dehydrogenase, membrane-bound hydrogenase (Mbh) with or without accessory formate dehydrogenase H (FdhH), while syntroph-associated pathways are heterodisulfide reductase-associated putative ion-translocating ferredoxin:NADH oxidoreductase (Hdr-Ifo) and electron-transfer-flavoprotein-oxidizing hydrogenase (FixABCX). The Hdr-Ifo column indicates presence of an Hdr-Ifo gene cassette (red circle), non-adjacent Ifo and Hdr genes (light red circle with red outline), and only Hdr genes (empty red circle).

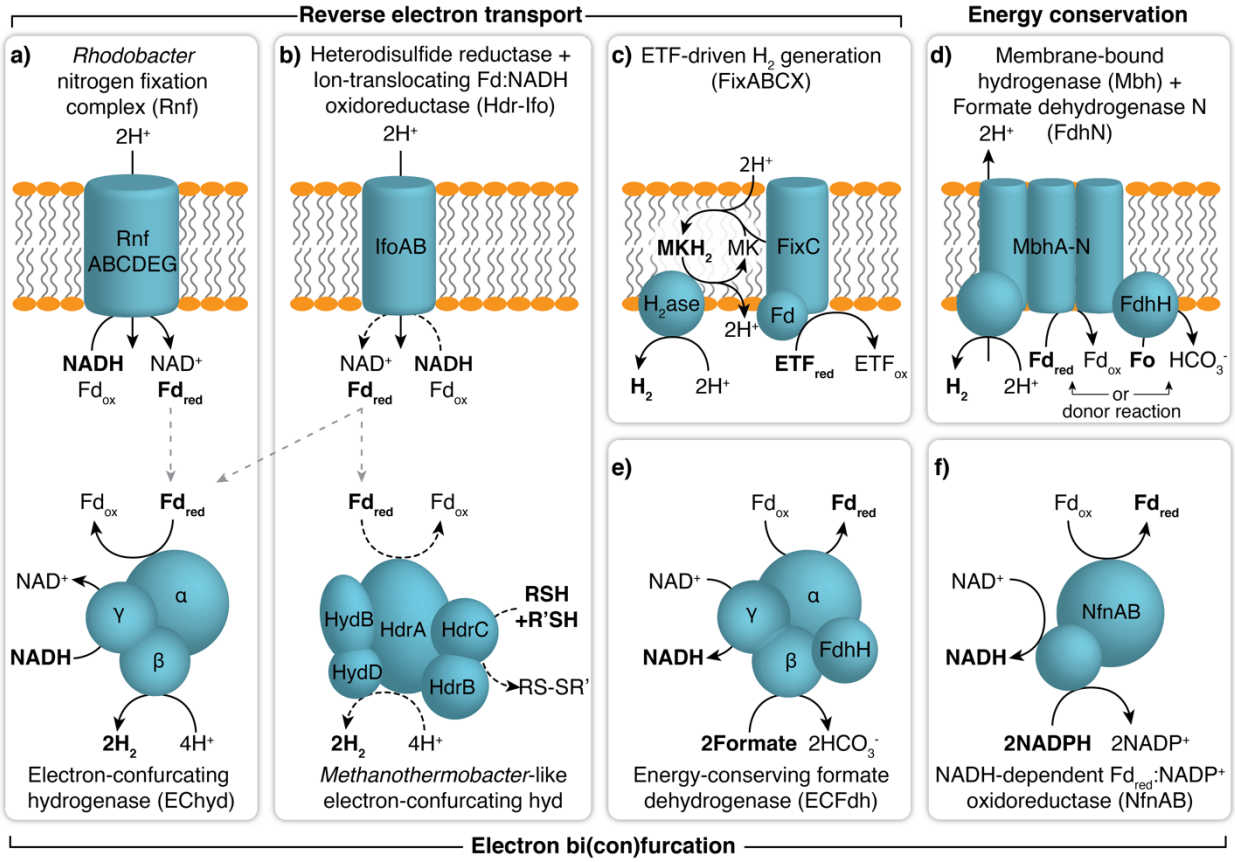

Figure 5.3. Reverse electron transport and electron confurcation mechanisms relevant to acetogen, fermenter, and syntroph energy conservation. (a) The *Rhodobacter* nitrogen fixation complex (Rnf) reverse electron transport (RET) generates reduced ferredoxin (Fd_{red}) for electron-confurcating hydrogenase (ECHyd) H_2 generation. (b) The heterodisulfide reductase (Hdr)-associated putative ion-translocating ferredoxin:NADH oxidoreductase (Hdr-Ifo) is thought to also perform RET-driven Fd_{red} generation to facilitate energy-conserving H_2 production through either ECHyd or putative *Methanothermobacter*-like electron-confurcating hydrogenase associated with the Hdr-Ifo cassette. (c) The electron-transfer-flavoprotein (ETF)-oxidizing hydrogenase complex (FixABCX) takes advantage of quinol oxidation-driven H_2 production. (d) Membrane-bound hydrogenases (Mbh) can conserve energy by extruding protons when performing Fd_{red}-oxidizing H_2 generation. This complex can be modulated by a formate dehydrogenase H (FdhH) to use formate (Fo) as an alternative electron donor. (e) An electron-bifurcating formate dehydrogenase (EBFdh) can facilitate energy-conserving formate metabolism; however, its reversibility remains unclear. (f) The NADH-dependent Fd_{red}:$NADP^+$ oxidoreductase is an electron bi(con)furcating enzyme found in organisms that utilize NADP(H) as an electron carrier (*i.e.*, acetogenesis and acetate-degrading syntrophy).

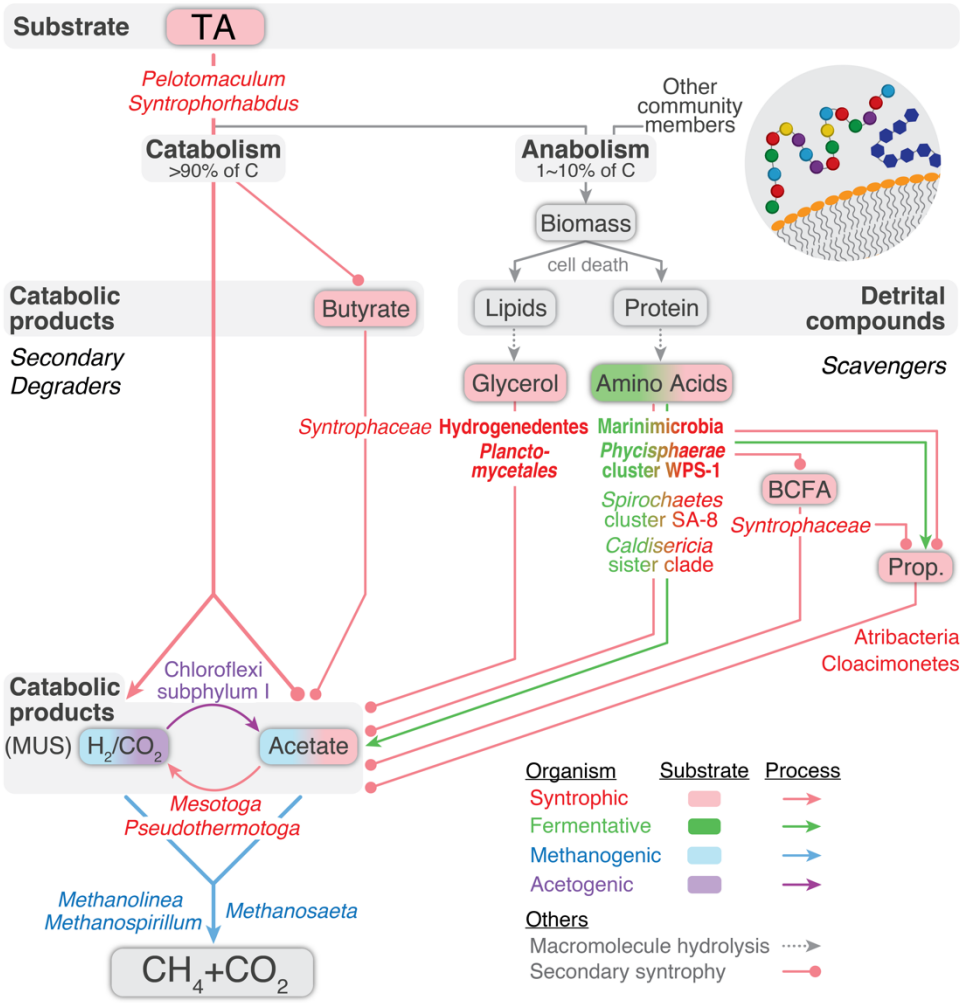

Figure 5.4. Holistic carbon flux from terephthalate (TA) to CH_4 and CO_2. TA degradation generates catabolic byproducts and detrital compounds. Catabolic byproducts are primarily methanogen-utilizable substrates (MUS) (i.e., acetate and H_2) and also include butyrate. Detrital compounds consist of biological macromolecules such as lipids and protein. Syntrophy- (red), fermentation- (green), methanogenesis- (blue), and acetogenesis- (purple) related taxa (name), substrates (box), and pathways (arrow) are shown. Taxa and substrates related to two processes are indicated by both colors. Syntrophic, fermentative, and acetogenic secondary degraders and scavengers interact to metabolize these compounds to acetate and H_2, which are finally mineralized to CH_4 and CO_2 by methanogens (blue). In particular, syntroph-to-syntroph substrate transfer (secondary syntrophy, circle-headed line) may play an important role in completing degradation of TA and protein. For scavenging detrital compounds, exoenzyme-producing organisms (bolded) are necessary to hydrolyze macromolecules (dotted gray arrow).

CHAPTER 6

DELVING INTO THE VERSATILE METABOLIC PROWESS OF THE OMNIPRESENT PHYLUM *BACTEROIDETES*

6.1 – Abstract

Although *Bacteroidetes* inhabit effectively every type of environment ranging from natural and engineered ecosystems to even the human body, the ecological contributions of many lineages remain unclear. We perform unprecedented comprehensive phylum-wide comparative genomics of 1600 genomes spanning 215 genera, including novel uncultured lineages illuminated through metagenomics and metatranscriptomics. This discerns relationships between *Bacteroidetes* phylogeny, habitat, and physiology and uncovers widespread and unique metabolic traits that overhaul our understanding of *Bacteroidetes*. Strikingly, we discover overlooked habitat-associated metabolic capacities including nitrogen respiration in the open ocean, H_2 oxidation in soil and freshwater ecosystems, syntrophy and formate generation in engineered environments, and H_2 cycling in the human gut. Thus, this holistic genomic investigation provides critical steps forward in recognizing the ecological impact of *Bacteroidetes* across biological disciplines.

6.2 – Introduction

As one of the most abundant and prevalent bacterial phyla on Earth, *Bacteroidetes* is ubiquitously found in natural (*e.g.*, soil (*253*), marine (*254*), and freshwater (*255*)), human/animal-associated (*e.g.*, gastrointestinal tract (*256*)), and artificial (*e.g.*, wastewater treatment process (*257*)) ecosystems. Reflecting this plethora of habitats, *Bacteroidetes* conquer extremes of many physiological spectrums: obligate aerobes to anaerobes, psychrophiles to thermophiles, phototrophs (*258*), and even pathogens and endosymbionts (*259*). In conjunction with such versatility, their unique and well-known capacity to degrade diverse polymeric organic substances allows *Bacteroidetes* to indispensably drive carbon turnover across the globe. It is no wonder that this phylum has been a major hub leading to discoveries of links between gastrointestinal microbiome and human health (*260, 261*), the relationship of insect diversification and endosymbiosis (*259, 262*), survival strategies in the oligotrophic open ocean (*254*), and novel polymer-decomposing enzymes for biotechnological application (*263*), just to name several. Maximizing our knowledge of these fascinating microorganisms is indisputably essential for furthering many biological disciplines.

The phylum consists of five known classes, typically affiliated with anaerobic human/animal-associated environments (*Bacteroidia*) and aerobic natural ecosystems (*Chitinophagia*, *Cytophagia*,

Flavobacteria, and *Sphingobacteriia*) (*264*). Although these organisms are generally characterized as hydrolytic (*i.e.*, polymer-degrading) organotrophs, beyond that, their habitat- and phylogeny-specific metabolic capacities and ecological impacts in biogeochemical cycles, engineered ecosystems, and human health remain largely unclear. Despite the thousands of *Bacteroidetes* isolates and genomes obtained over the past decades, we still lack definitive insight into (a) explicit relationships between *Bacteroidetes* phylogeny, habitat, and physiology, (b) characteristic traits of specific genera and families, and (c) rare functional capabilities across the phylum. Moreover, based on phylogeny defined by Greengenes 16S rRNA database (*212*), more than one third of *Bacteroidetes* families (11 out of 31) remain uncharted, suggesting we have yet to recognize the full spectrum of *Bacteroidetes*' capabilities. Thus, in this study, we metagenomically recover genomes for uncharacterized *Bacteroidetes* and perform rigorous comprehensive phylum-wide comparative genomics (1600 genomes spanning 215 genera) to contextualize each individual's physiology, provide novel insight into the ecophysiological range and versatility of *Bacteroidetes*, and elucidate overlooked ecological contributions across natural, engineered, and human-associated biological processes on Earth.

6.3 – Results and Discussion

6.3.1 – Towards comprehensive phylogenetic coverage of the phylum

The lack of representatives for 35% of *Bacteroidetes* family-level lineages is a severe shortcoming in understanding the ecophysiology of the phylum. Inspecting the Greengenes database (*212*) reveals that 82% of the uncultivated *Bacteroidetes* families (9 of 11) affiliate with *Bacteroidia* in engineered environments, and more specifically methanogenic bioreactors. To acquire genomes for such lineages, we perform metagenomics on a full-scale anaerobic digester hosting diverse uncultivated *Bacteriodia* families and successfully recover 12 high quality *Bacteroidia* genomes (>92% genome completeness and <6% contamination) (Appendix E Table E.1; available under GenBank BioProject PRJNA321808). Eight of these genomes indeed represent seven novel Bacteroidia families given that the phylogenomic distance to their closest relatives is at least 15 and 31 % larger than the average maximum distance within cultured families and average minimum distance between cultured families (Fig. 6.1A, 6.1B, and Appendix E Table E.1). By the same criteria, rumen- (*265*) and estuary- (*266*) associated lineages also comprise novel families. In total, we successfully double the phylogenetic range of our insight into *Bacteroidia* (9 cultured and 9 novel uncultured families) and also expand the number of family-level lineages of *Bacteroidetes* by 45% (Fig. 1A). Combining the significantly increased phylogenetic coverage of the phylum with the publically available genomes, we have a holistic genome dataset sufficiently representative of all *Bacteroidetes* classes appropriate for phylum-wide comparative genomics.

6.3.2 – A major division in the phylum: *Bacteroidia* vs. non-*Bacteroidia* classes

We predict the physiology of each *Bacteroidetes* member through rigorously annotating polymer hydrolysis (glycosylhydrolases, peptidases, and lipases) and catabolism of monomers (sugars and amino acids [AAs]) along with downstream respiration (aerobic and anaerobic), fermentation, intracellular electron transduction necessary to facilitate these pathways, and other types of energy metabolism (*e.g.*, hydrogen and formate metabolism). This level of specificity is necessary because widely employed annotation techniques (*e.g.*, COG) do not effectively show relationships between phylogeny and function (Fig. S1). Further multiple correspondence analysis (MCA) successfully reveals separation of the phylum into largely two ecotypes: (i) anaerobic *Bacteroidia* in human/animal-related and engineered environments and (ii) aerobic non-*Bacteroidia* classes (*i.e.*, *Chitinophagia*, *Cytophagia*, *Flavobacteria*, and *Sphingobacteriia*; or CCFS) in natural ecosystems (*i.e.*, freshwater, marine, and soil) (Appendix E Fig. E.2A). Logically, phylogenetically and environmentally diverse facultative (an)aerobes are found in an intermediate position and functionally distinct ecto/endosymbionts differentiate from all other ecotypes. Moreover, reflecting the respiratory between the two major groups, MCA confirms that CCFS have higher aerobic AA degradation capacity and *Bacteroidia* have higher capacity for anaerobic AA degradation, fermentation, and intracellular electron transduction (Appendix E Fig. E.2B). All observations agree with the well-accepted traits of *Bacteroidetes* classes and supports the efficacy of this study's approach.

Division of *Bacteroidetes* into *Bacteroidia* and CCFS reveals contrasting metabolic traits. For polymer hydrolysis, the two groups have similar distributions of carbohydrate hydrolysis capacity (*i.e.*, number of glycosylhydrolase families), but CCFS generally have a higher range extracellular peptidase and lipase families (Fig. 2A). Correspondingly, while few glycosylhydrolases are exclusive to either group, many peptidase (13) and lipase (6) families are only found in CCFS (Appendix E Table E.2), suggesting CCFS may encounter and degrade a wider variety of polypeptides and lipids. As for monomer catabolism, though CCFS genera tend to only degrade a moderate range of AA and sugar types (median of 6 types and maximums of 13 and 11 types respectively), *Bacteroidia* genera have more diverse capacities and can even reach up to 17 and 18 types of AAs and sugars respectively (Fig. 6.2A). Reflecting this, CCFS have normal distributions of monomer substrate ranges, whereas *Bacteroidia* does not (R^2>0.98 and <0.95 correspondingly for linear regressions of normal probability plots). The inverse trends of polymer hydrolysis and monomer catabolism suggest that CCFS members tend to hydrolyze a wide range of polymers and only degrade specific monomers and *Bacteroidia* contrastingly includes a diverse array of genera that specialize in scavenging a narrow or wide range monomers. Overall, this indicates that *Bacteroidia* and CCFS members fill contrasting ecological roles.

To decipher environment-dependent differences between CCFS and *Bacteroidia*, this study further compares metabolic capacities between genera grouped by their habitat: gastrointestinal tract, host-associated (*e.g.*, oral cavity and wounds), marine, freshwater, soil, and bioreactors (Fig. 6.2B). First focusing on CCFS, soil-affiliated genera consistently encode the most diverse array of hydrolytic enzymes for all three polymer categories, likely due to the high biochemical and ecological complexity of terrestrial soils. In contrast, they only degrade moderate ranges of monomers. Freshwater and marine genera show similar properties, but encode lower hydrolytic enzyme diversity. As opposed to environmental CCFS, gastrointestinal and other host-associated genera generally encode fewer glycosylhydrolase families and sugar degradation pathways, suggesting they may only have limited saccharide-degrading capacities. Next, for *Bacteroidia*, all genera have relatively consistent hydrolytic enzyme diversity, yet genera from different habitats have contrasting monomer degradation capacities (Fig. 6.2B). Bioreactor, marine, and freshwater *Bacteroidia* genera tend to have the capability to degrade a much wider range of AAs than those from other environments. For sugars, bioreactor and host-associated *Bacteroidia* can catabolize a narrow range of sugars compared to those from other habitats. Notably, several genera affiliated with rumen and gasotrintestinal tracts have the unique capacity to degrade very diverse range of sugars. This reemphasizes that CCFS are more environmentally diverse in polymer hydrolysis and *Bacteroidia* in monomer degradation. Lastly, across both CCFS and *Bacteroidia*, ecto- and endo-symbiotic members (*e.g.*, "*Ca.* Azobacteroides") have the lowest metabolic capacities in all categories, reflecting their host-dependent nature.

6.3.3 – Versatile energy metabolism CCFS as novel components of environmental ecosystems

In-depth MCA reveals association between CCFS phylogeny, habitat, general metabolic capacities, and specific catabolic pathways with ecological relevance to global biogeochemical cycles (Fig. 3). While marine CCFS are heterogeneous based on MCA, those from freshwater and soil habitats are relatively homogeneous at both genus and family level (Fig. 6.3A); however, overall, marine genera clearly have distinct metabolic qualities from freshwater and soil genera. First, regarding catabolism of specific monomers, CCFS members commonly encode degradation of growth-conducive sugars and AAs that lead directly to glycolysis, pyruvate, or tricarboxylic acid cycle (TCA) (*i.e.*, glucose, fructose, Ala/Cys/Asp/Asn/Glu/Gln/Pro; Fig. 3B and S3A). By the same line of logic, though less common, degradation of most pentoses (arabinose, fucose, L-ribulose, xylose, D-xylulose) are found evenly across environmental genera (*i.e.*, marine, freshwater, and soil). However, catabolism of several sugars and other AAs either specifically coincide with CCFS found in marine (D-mannose; Met/Gly/Ser/His), marine and soil (D-ribulose; Arg), or freshwater and soil (Ile, Leu, Trp, and Tyr; D-ribose and N-acetylglucosamine).

Further, marine-affiliated AA degradation and environmental sugar catabolism have contrasting sugar-, polysaccharide-, lipid-degradation capacity (low and high respectively; Fig. 6.3C). Thus, CCFS sugar, AA, and polymer degradation surprisingly vary across natural environments, suggesting divergent adaptation across the aerobic biosphere.

Remarkably, the aforementioned marine CCFS AA degradation and environmental sugar degradation coincide with distinct types of energy metabolism. MCA reveals correlation between marine AA degradation and the abilities to harness light energy via proteorhodopsin and also reduce nitrogen oxides (both denitrification and ammonification; Fig. 6.3B). Indeed, proteorhodopsin and most nitrogen oxide reducing enzymes (*i.e.*, Nar, Nir, Nor, and Nrf) are found in a larger fraction of marine genera than freshwater or soil (Appendix E Fig. E.4). Although most nitrogen oxides are typically not tested (*i.e.*, nitrite, nitric oxide, and nitrous oxide), many marine CCFS can reduce nitrate under aerobic conditions (*267*). From this, we can deduce that reduction of nitrogen oxides is likely a supplemental energy metabolism and not respiration. Reflecting the surprising prevalence of nitrogen oxide reduction in marine CCFS, only marine genera (*i.e.*, *Aquiflexum*, *Belliella*, *Cecembia*, and *Imtechella*) encode full denitrification. However, most marine CCFS do not encode the full spectrum of denitrifying enzymes, suggesting that they may scavenge denitrification intermediates produced by other organisms in the open ocean and also generate partially reduced nitrogen oxides for others to metabolize further. Thus, we propose that marine *Bacteroidetes* may degrade AAs and release ammonia, generate additional ammonia through nitrite reduction, perform step-wise denitrification through interaction with other marine organisms, and supplement this with phototrophy, whereby making unique contributions to ecology revolving around organotrophy and nitrogen cycling in the open ocean.

On the other hand, we discover that H_2 metabolism corresponds to freshwater and soil CCFS (Fig. 6.3B and Appendix E Fig. E.4). Unexpectedly large fractions of these genera (27.8 and 44.4% respectively) encode novel *Bacteroidetes* [NiFe] Hox-, Hup-, and Hya-type hydrogenases (Figure 6.3B). Compared to Hox, Hup- and Hya-based H_2 metabolism is more common among CCFS members. Hup and Hya are both membrane-bound uptake hydrogenases, suggesting these CCFS members may oxidize H_2 either as a primary or supplementary energy source. In addition, these uptake hydrogenases often coincide with versatile sugar catabolism: >93% of Hup/Hya-harboring genera can degrade at least five sugar types (phylum median) and >50% eight types (phylum third quartile). This implicates that H_2 oxidation is an overlooked component of sugar degradation by CCFS genera, especially those from freshwater and soil environments. Thus, CCFS driving mineralization of carbohydrates may also simultaneously make significant contributions to the global H_2 cycle by scavenging either atmospheric or anaerobically generated H_2. However, an H_2-oxidizing *Bacteroidetes* has yet to be reported, so further studies are necessary to

elucidate the physiological function of H_2 oxidation and fully understand *Bacteroidetes*-mediated carbohydrate decomposition in natural ecosystems.

Although CCFS genera are generally aerobic, we discover diversity even in their aerobic respiration. Nearly all CCFS (97%) encode cytochrome c oxidase (Cox), suggesting it is the core terminal oxidase supporting aerobic respiration for CCFS. In contrast, only 72% encode a cytochrome bd-type quinol oxidase (Cbd), evenly distributed across different habitats (Appendix E Fig. E.4). Cbd are associated with high oxygen affinity and can also confer resistance to environmental inhibitors (*e.g.*, hydrogen peroxide, antibiotics, and cyanide) (*268, 269*), so genera harboring such terminal oxidases may use Cbd to tolerate microaerobic conditions and other environmental stresses, perhaps contributing to the success of CCFS in diverse ecosystems. To better understand CCFS aerobic respiration, comparison to the respiratory pathways of facultatively aerobic *Bacteroidia* is critical. Of the 9 known facultatively aerobic genera, 5 encode Cox (*i.e.*, *Draconibacterium*, *Sunxiuqinia*, *Geofilum*, *Saccharicrinis*, and *Marinifilum*), *Prolixibacter bellariivorans* uniquely harbors a cytochrome bo3 quinol oxidase (Cbo), and most other genera only possess Cbd (*e.g.*, Marinilabilia), indicating that Cox, Cbd, and also Cbo can support *Bacteroidia* aerobic respiration. Notably, Cbd is found in many obligately anaerobic *Bacteroidia* (73% of genera), likely for the purpose of combating oxidative stress (*270*) or even taking advantage of O_2 as a supplemental non-respiratory electron acceptor (*271*). In total, while *Bacteroidetes* can utilize diverse terminal oxidases for aerobic respiration as a phylum, CCFS specialize in respiration through Cox, and perhaps Cbd under certain circumstances. Additionally, the diverse respiratory pathways of *Bacteroidia* clearly suggests that CCFS and *Bacteroidia* underwent divergent adaptation to aerobic life.

6.3.4 – Unique metabolism of *Bacteroidia*: fermentation, hydrogen, and formate

Comparison of *Bacteroidia* physiology across habitats also reveals striking environmental variation; moreover, close inspection of their genomes reveal many novel metabolic niches, perhaps because anaerobic metabolism is generally less studied than aerobic. Although degradation of simple hexoses (glucose, fructose, and mannose) and labile AAs (TCA-linked – Asp/Asn/Glu; 2-oxoacid-generating – Ala/Cys/Met/Thr) are found commonly across the class, the class can largely be divided into three types based on the association of AA and sugar degradation with differing habitats (Fig. 6.3E). Degradation of pentose-phosphate-pathway-associated pentoses (xylose, D-xylulose, L/D-ribulose, and L/D-arabinose) affiliate with environmental *Bacteroidia*; however, pentoses (rhamnulose, fuculose, rhamnose, fucose) and amino-sugars (N-acetylglucosamine and N-acetylmuramic acid) that require dedicated catabolic pathways are almost exclusively found in gastrointestinal genera (Fig. 6.3E and Appendix E Fig. E.3B). While rumen genera unbiasedly degrade most of the aforementioned sugars, methanogenic bioreactor and host-

associated members tend to lack the capacity to degrade many of these sugars. As for AAs, bioreactor genera clearly affiliate with degradation of many AA types, especially branched-chain (Ile/Leu/Val), glutamate-dependent (Arg/Pro/His) and formate-generating (Gly/Ser) AAs (Fig. 6.3E and Appendix E Fig. E.3B and E.5). In addition, the degradation of several other AAs (Lys/Trp/Tyr/Gln) associate with both bioreactor and aquatic (*i.e.*, marine and freshwater) *Bacteroidia*. Generally, catabolism of such AAs do not coincide with gastrointestinal, rumen, and host-associated environments. In addition to these trends in sugar/AA degradation, we identify clear correlation between sugar metabolism and polymer hydrolysis (Fig. 6.3F). Thus, in total, we can largely divide *Bacteroidia* monomer catabolism into three types: human/animal-associated genera with versatile sugar degradation and polymer hydrolysis, bioreactor genera with diverse AA catabolism, and aquatic genera with intermediate capacities for both.

For anaerobic life, fermentation, H^+-reducing H_2 production, and CO_2-reducing formate generation (*272*) are vital processes supporting organotrophy that substituting respiration for electron disposal. Although fermentative production of acetate, TCA-affiliated acids (*i.e.*, malate, succinate, and propionate), and alcohol are common throughout the class *Bacteroidia* (≥66% of genera in each habitat), acetate-reducing butyrate fermentation is quite rare (only five genera in *Bacteroidia*). Besides these, different fermentation pathways correspond to distinct environments and catabolism (Fig. 6.3E). For human/animal-associated environments, L-lactate fermentation is associated with genera from gastrointestinal, rumen, and host-associated environments (80/83/85% respectively and ≤38% for others) and, though less common, D-lactate fermentation with genera from those environments and soils (47/33/54/33% and ≤25% for others). In addition, we observe 1,2-propanediol fermentation in genera from digestion-related ecosystems (gastrointestinal and rumen) (47/50% and ≤25% for others). These fermentation pathways tend to coincide with sugar degradation, which is logical as sugar catabolism often leads to pyruvate and lactaldehyde that can be respectively converted to L/D-lactate and propanediol. On the other hand, fermentative generation of formate from 2-oxoacids (*e.g.*, pyruvate) is associated with bioreactor and marine environments (50/56% and ≤34% for others). Thus, *Bacteroidia* clearly have hitherto overlooked diversity in fermentative capacities. Notably, the unexpectedly high frequency of D-lactate fermentation (47%) and low frequency of butyrate fermentation (6.7%) in gastrointestinal genera sheds new light on the contribution of *Bacteroidetes* to gut acid production and human health (*273-275*).

Hydrogenogenesis is a particularly critical step in the carbon and sulfur cycles, linking organotrophy to methanogenesis and sulfate reduction (*2*). Depending on the substrate, H_2-producing catabolism can be thermodynamically inhibited by H_2 accumulation (*133*), so H_2-producers often depend on "syntrophic" interactions with H_2-oxidizing partners (*e.g.*, methanogens and sulfate reducers) (*13*). Many *Bacteroidia* from bioreactor, marine, and freshwater ecosystems (67, 67, and 100% of genera) encode

[FeFe]-type hydrogenases indicative H_2 generation, especially those with high AA-degrading capacity (Fig. 6.3E). However, *Bacteroidia* may have multiple H_2 generation strategies given that they encode two types of energy-conserving hydrogenases: trimeric HydABC (10% of genera) and tetrameric HydABCD (40%) (*6*). Perhaps, *Bacteroidia* in these environments rely on different kinds of H_2 production and/or H_2-mediated syntrophic partnerships to drive anaerobic polymer decomposition and organotrophy, especially of AAs. Notably, many facultatively aerobic marine *Bacteroidia* intriguingly harbor energy-conserving [FeFe] hydrogenases (*Draconibacterium*, *Sunxiuqinia*, *Saccharicrinis*, and *Marinilabilia*), suggesting they have achieved the exceptional feat of adapting to both extremes of the redox spectrum (H^+ and O_2 respiration) at the aerobic-anaerobic interface in sediment oxyclines.

As an alternative to H^+-reducing H_2 production, we discover the first *Bacteroidetes* members (ADurb.Bin008 and *Lentimicrobium*) that can extraordinarily also employ CO_2 as an anaerobic electron acceptor and produce formate. The two organisms' formate dehydrogenases are enzymatically and phylogenetically quite distinct: ADurb.Bin008 encodes a syntrophic *Deltaproteobacteria*-related NADH-dependent formate dehydrogenase and *Lentimicrobium* encodes a *Firmicutes*-related putatively electron-bifurcating formate dehydrogenase (Appendix E Fig. E.6A and E.6B). Interestingly, the organisms also have contrasting catabolic capacities where ADurb.Bin008 is a versatile AA degrader while *Lentimicrobium* is only known to degrade sugars (*276*). Thus, the newly identified *Bacteroidetes* formate metabolism may support different formate-generating lifestyles. In total, many *Bacteroidia* members show clear and diverse adaptation to H^+-respiring anaerobic life and some even resort to implementing CO_2 as an electron acceptor and generating formate, a metabolism thought to be primarily restricted to *Deltaproteobacteria* and *Firmicutes*.

To characterize and contextualize this specialized anaerobic metabolism, we metatranscriptomically investigate the metabolic behavior of ADurb.Bin008 in its native environment – an anaerobic digester. Among all *Bacteroidetes*, ADurb.Bin008 encodes degradation of the highest number of AA types (17 types), even including those that require syntrophic interactions (*245*) (Appendix E Fig. E.5). Accordingly, this organism expresses degradation of many AAs (7~10 types), multiple extracellular peptidases, and overall high activity both *in situ* (7.4% of total metatranscriptome) and in the presence of model protein substrates (*i.e.*, casein or gelatin; 6.8 and 7.3% respectively) (Appendix E Fig. E.7A, E.7C, and E.7D). In addition, ADurb.Bin008 predominates when fed with AAs whose degradation depends on syntrophy (*i.e.*, Ile, Leu, or Val; up to 51.9%) and highly expresses the appropriate catabolic pathways (Appendix E Fig. E.7B). Through inspection of its gene expression *in situ* using metatranscriptomics, we find that this organism has high activity (7.4% of metatranscriptomic reads) and expresses many AA (10 types) degradation pathways, extracellular peptidases, and a hydrogenase, indicating capacity to perform

H_2-generating protein hydrolysis and AA fermentation (Appendix E Fig. E.7A, E.7C, and E.7D). Indeed, when the anaerobic digester community is fed with model protein substrates (*i.e.*, casein and gelatin), ADurb.Bin008 retains similar activity and also expresses H_2-evolving degradation of many AAs (7~8 types). In addition, when we feed the anaerobic digestion community with branched-chain amino acids (BCAAs) whose degradation requires syntrophic interactions (*i.e.*, Ile, Leu, or Val), ADurb.Bin008 clearly predominates (up to 51.9% of metatranscriptomic reads) and highly expresses the corresponding catabolic pathways and H_2 generation (Appendix E Fig. E.7A and E.7B). Under all *in situ*, protein-fed, and BCAA-fed conditions, a H_2-oxidizing methanogen (*Methanoculleus* ADurb.Bin009) has consistently high activity, suggesting H_2 transfer between ADurb.Bin008 and ADurb.Bin009. We also observe that formate generation and transfer is especially critical during BCAA degradation given that ADurb.Bin008 displays significantly higher expression of the formate dehydrogenase and complementary energy conservation pathways compared to *in situ* (>25% higher and $p<0.05$). Indeed, the partner ADurb.Bin009 correspondingly expresses formate-driven methanogenesis (Appendix E Fig. E.7B). In total, the capacity to degrade BCAAs through H_2 and formate transfer to a partner methanogen demonstrates that ADurb.Bin008 is capable syntrophic AA degradation. Thus, we put forth ADurb.Bin008 as the first discovered syntrophic *Bacteroidetes*, featuring unique CO_2-reducing formate generation to accomplish degradation of protein and AAs through metabolic partnership with methanogens. To highlight this functionally novel *Bacteroidetes*, we propose the taxonomic name of "*Candidatus* Aminosyntropha formicigenes" gen. nov., sp. nov. (A.min.o.syn.tro'pha. M.L. adj. *Amino* amino; Gr. adj. *syn* together with; Gr. fem. n. *trophos* one who feeds; M.L. masc. n. *Aminosyntropha* referring to the capacity to degrade amino acids in syntrophic association with hydrogenotrophic organisms) (for.mi.ci.ge′nes. N.L. adj. *formicicum* from L. n. *formica* ant; Gr. v. *gennaio* produce; N.L. adj. *formicigenes* producing formic acid)

In contrast with this specialized H_2/formate producer, *Bacteroidia* H_2 generation is not a prevalent process in host-associated, soil, rumen, and gastrointestinal genera as relatively few of them harbor [FeFe] hydrogenases (15, 33, 38, and 38%). Moreover, the small proportion of hydrogenase-encoding species in genera associated with those environments (*e.g.*, 25 and 36% of gastrointestinal *Alistipes* and *Bacteroides* species) implicates that H_2 generation is not an essential component of their lifestyle and ecological contribution. However, we discover that select obligately anaerobic gastrointestinal and potentially pathogenic *Porphyromonadaceae* genera (*Barnesiella*, *Corpobacter*, *Parabacteroides*, and *Tannerella*) rather harbor [NiFe]-type Hya uptake hydrogenases closely related to those found in CCFS (~60% amino acid sequence similarity; Appendix E Fig. E.6C). This implies that these genera can curiously oxidize H_2 without exogenous electron acceptors, a completely overlooked function of *Bacteroidetes*. We suspect that H_2 oxidation can contribute to resistance against oxidative stress (*277*), proton motive force generation, and

reduction of endogenous metabolites for energy acquisition (*e.g.*, fumarate respiration) or biosynthesis (*278*).

To investigate the role of H_2 oxidation and its potential relationship with catabolism and fermentation in the human gut, we closely inspect publically available gastrointestinal tract metatranscriptomes (*279*). The microbial community activity is quite variable across subjects, in terms of both activity of individual species and their carbohydrate degradation (Appendix E Fig. E.8A and E.8C); however, we uncover predominantly active gut species (7 *Bacteroides*, 2 *Parabacteroides*, 1 *Alistipes*, and 1 *Prevotella*) that collectively account for up to 76% of the gut *Bacteroidetes* gene expression (Appendix E Fig. E.8A). Among them, only *Parabacteroides distasonis* harbors an Hya uptake hydrogenase and also consistently expresses this hydrogenase across subjects (83.3% of subjects; Fig. S8B). Incidentally, it is also the only gut organism capable of and actively degrading most AAs. Thus, H_2-driven *P. distasonis* stress resistance and supplemental metabolism is not only unexpectedly prevalent in the gastrointestinal tract, but also linked to gastrointestinal protein digestion. Such Hya-mediated consumption of gastrointestinal H_2 may also potentially contribute to host health, given that gastrointestinal H_2 production can associate with obesity (*280*) and pathogens can use H_2 as an energy source during infection (*278*). *P. distasonis* can also generate H_2 itself through an [FeFe] hydrogenase, suggesting that this bacterium may even intracellularly recycle self-produced H_2 (*281*). In total, we discover that a highly active *Parabacteroides* species may serve a unique role in gastrointestinal microbial ecology revolving around hitherto overlooked gastrointestinal H_2 cycling and AA decomposition. Clearly, many links between gut *Bacteroidia* and host health remain to be uncovered.

6.4 – Conclusion

Although *Bacteroidetes* are prevalent across the globe and often detected as major populations, they are often dismissed as simple polymer, carbohydrate, or peptide degraders. However, performing rigorous phylum-wide comparative genomics sheds light on links between *Bacteroidetes* phylogeny, habitat, and metabolism and unearths novel and highly relevant ecological niches that *Bacteroidetes* members can fill. Investigation of uncultivated *Bacteroidetes* in engineered environments unveils "*Ca.* Aminosyntropha" with the capacity to hydrolyze protein and degrade almost the full spectrum of AAs through syntrophy, using a wide array of electron transduction pathways and the unique capacity to respire both H^+ and CO_2 (Fig. 6.4A). Genomically revisiting the physiology of isolates also reveals distinctive roles that *Bacteroidetes* can take, including several that we highlight here. In the gastrointestinal tract, we discover that *Parabacteroides distasonis* can impact human health by single-handedly driving AA degradation and performing H_2 cycling through typically mutually exclusive hydrogenases (Fig. 6.4A). In the open ocean, organisms like

Aquiflexum balticum can stand at the interface of polymer decomposition, nitrogen cycling, and phototrophy with a full spectrum of denitrification and ammonification enzymes and proteorhodopsin (Fig. 6.4B). In marine sediments, *Sunxiuqinia elliptica* possess a striking respiratory arsenal including a hydrogenase, ammonifying nitrite reductase, and terminal oxidase to support flexible carbohydrate and protein degradation across the oxic-anoxic interface in marine sediments (Fig. 6.4B). Although just a subset of the findings, these discoveries change our understanding of how *Bacteroidetes* can contribute to biotechnology, human health, and marine biogeochemical cycling. In parallel to the on-going efforts uncovering uncultivated phyla known as microbial dark matter (*143, 217*), we present this study's holistic approach as an effective way to revisit familiar and ubiquitous phyla and fully exploit the massive and ever-expanding genomic database. Sometimes, the darkest place is just beneath the lighthouse.

6.5 – Materials and Methods

All publically available *Bacteroidetes* genomes were downloaded from Genbank and Joint Genome Institute integrated genome and metagenome server (*282*). To acquire genomes for uncultivated *Bacteroidetes*, we collected samples from a full-scale anaerobic digester located in Urbana, Illinois at three consecutive months (October, November, and December 2013), extracted genomic DNA using the MP Biomedicals (Santa Ana, California, USA) FastDNA SPIN Kit for Soil DNA Extraction, sequenced, pretreated, and assembled as previously described (*150*), binned the metagenomic assembly using MetaBAT v0.26.3 and MaxBin 2.1 (*55, 56*), assessed genome completeness and contamination using CheckM v1.0.1 (*58*), and finally retained high quality *Bacteroidetes* genomes from each binning software for downstream analysis (>90% completeness and <7.5% contamination). All *Bacteroidetes* genomes were annotated through Prokka v1.11 (*208*). The annotation was further manually curated with assistance from functional domain analysis through CD-Search with its corresponding conserved domain database (*283, 284*); signal peptide and transmembrane domain prediction through SignalP v4.1 (*285*); carbohydrate-active enzyme, peptidase, and lipase prediction through dbCAN 5.0 (*286*), MEROPS (*287*), and lipase engineering database (*288*); and hydrogenase annotation using HydDB (*7*). In addition, to further verify the function, we compared the sequence similarity of each gene to a database containing enzymes with experimentally verified catalytic activity and genes with extensive genetic, phylogenetic, and/or genomic characterizations with a 40% amino acid similarity cutoff. For enzymes that have divergent functions even with a 40% similarity cutoff (*e.g.*, FeFe hydrogenases, 2-oxoacid oxidoreductases, glutamate dehydrogenases, and sugar kinases), phylogenetic trees were constructed with reference sequences to identify association of the query sequences to phylogenetic clusters containing enzymes with characterized catalytic activity. Specifically for hydrogenases, formate dehydrogenases, and electron transduction complexes (*e.g.*, Rnf and Nqr) that are

often composed of multiple subunits and tend to organize together on the genome, we only annotate the function of the complex if all subunits were identified in a gene cassette or divided onto two contigs (*i.e.*, two ends of the operon on the ends of two contigs).

In addition to gene annotation, metabolic capacities/traits were predicted based on the strict criteria that all enzymes necessary for the pathway could be identified. Specifically, sugar degradation (18 types), amino acid degradation (20 types), electron transduction mechanisms (*e.g.*, NADH:quinone oxidoreductase), respiration (O_2 and nitrogen species), H_2 metabolism, formate metabolism, polymer hydrolysis (glycosylhydrolase, extracellular peptidase, and extracellular lipase families), phototrophy (*i.e.*, proteorhodopsin), and diazotrophy (*i.e.*, nitrogenase). In particular, anaerobic metabolism is difficult to predict due to the generally poorer insight compared to aerobic metabolism and high complexity of energy conservation, electron transduction, fermentation, and anaerobic respiration. Moreover, while sugar-metabolizing pathways are quite consistent between aerobes and anaerobes, pathways for degrading amino acids can differ significantly. Thus, for anaerobic amino acid degradation, the complete pathway was only assigned when the genome also harbored complementary electron disposal pathways (*i.e.*, fermentation, H_2 generation, and/or formate generation) and appropriate electron transduction. As an example, a pathway that produces one nicotinamide adenine dinucleotide (NADH) per substrate degraded would require electron transduction from NADH to ferredoxin to utilize a ferredoxin-oxidizing hydrogenase to complete the metabolic pathway. Finally, to define the average metabolic capacity for each *Bacteroidetes* genus, for each metabolic trait, we calculated the percentage of genomes (*i.e.*, strains) with the predicted capacity for each genus. In addition, the overall capacity to degrade monomers (sugars and amino acids) and macromolecules (polysaccharides, peptides, and lipids) were respectively defined by the number of metabolizable sugars/amino acids and hydrolase families. Further, a metabolic trait was considered part of each genus' "overall" capacity only if at least 10% of the genomes encoded the function. Finally, using R and the R packages FactoMineR and ggplot2 (*289-291*), multiple correspondence analysis was performed to compare the overall metabolic traits of *Bacteroidetes* genera.

To study the gene expression of anaerobic digester *Bacteroidetes*, samples were taken from the aforementioned anaerobic digester, inoculated into 120mL serum bottles (8mL sludge with 72 mL anaerobic media and 80:20 N_2-CO_2 headspace prepared as previously described (*153*)), and fed with gelatin, casein, isoleucine, leucine, valine, or olive oil. Gelatin and casein were added at a concentration of 1.25g L^{-1} and olive oil at 20mL L^{-1}. To mimic *in situ* conditions where monomeric substrates are available at low concentrations, the branched-chain amino acid treatments were operated as fed-batch cultures in which 500uM isoleucine, leucine, and valine was fed into the individual treatments when methane production from the previous feeding plateaued. All treatments were executed in triplicate. Headspace gas composition

was monitored using a Shimadzu GC-2014 Gas Chromatograph. For the macromolecule-fed samples, samples were collected when methane production entered exponential phase (5, 5, and 14 days respectively for gelatin-, casein-, and olive oil-fed treatments) and immediately stored in a freezer at -80˚C. As for the branched-chain amino acids, samples were collected the cultures completed 10 fed-batch cycles (*i.e.*, 5mM isoleucine/leucine/valine). RNA was extracted for the above treatments and raw anaerobic digester sample collected for metagenomics as previously described (*292*).

RNA was prepared for plate-based sequencing using the PerkinElmer (Waltham, MA, USA) Sciclone NGS robotic liquid handling system using Illumina's (San Diego, CA, USA) Ribo-Zero rRNA Removal Kit (Bacteria) and the TruSeq Stranded Total RNA HT sample prep kit following the protocol outlined by Illumina with the following conditions: 100ng total RNA starting material per sample and library amplification through 10 cycles of polymerase chain reaction (PCR). KAPA Biosystem's (Wilmington, MA, USA) next-generation sequencing library quantitative PCR kit was used to quantify the prepared libraries (run on a Roche (Basel, Switzerland) LightCycler 480 real-time PCR instrument). After multiplexing the quantified libraries, an Illumina HiSeq sequencing platform was used with a TruSeq paired-end cluster kit v4 and cBot instrument to generate a clustered flow cell for sequencing the combined pool. Downstream sequencing was performed through the Illumina HiSeq2500 sequences with HiSeq TruSeq SBS sequencing kits v4 using a 2x150 indexed run recipe. The generated metatranscriptomic sequences were trimmed using Trimmomatic v0.30 with a quality cutoff of 30, sliding window of 6bp, and minimum length cutoff of 50 base pairs (*155*) and mapped to metagenomic bins using BBmap (Bushnell B. - sourceforge.net/projects/bbmap/). The gene expression levels were calculated separately for individual bins as reads per kilobase transcript per million reads (RPKM) normalized to the bin's median RPKM (excluding genes with no mapped reads).

6.6 – Figures

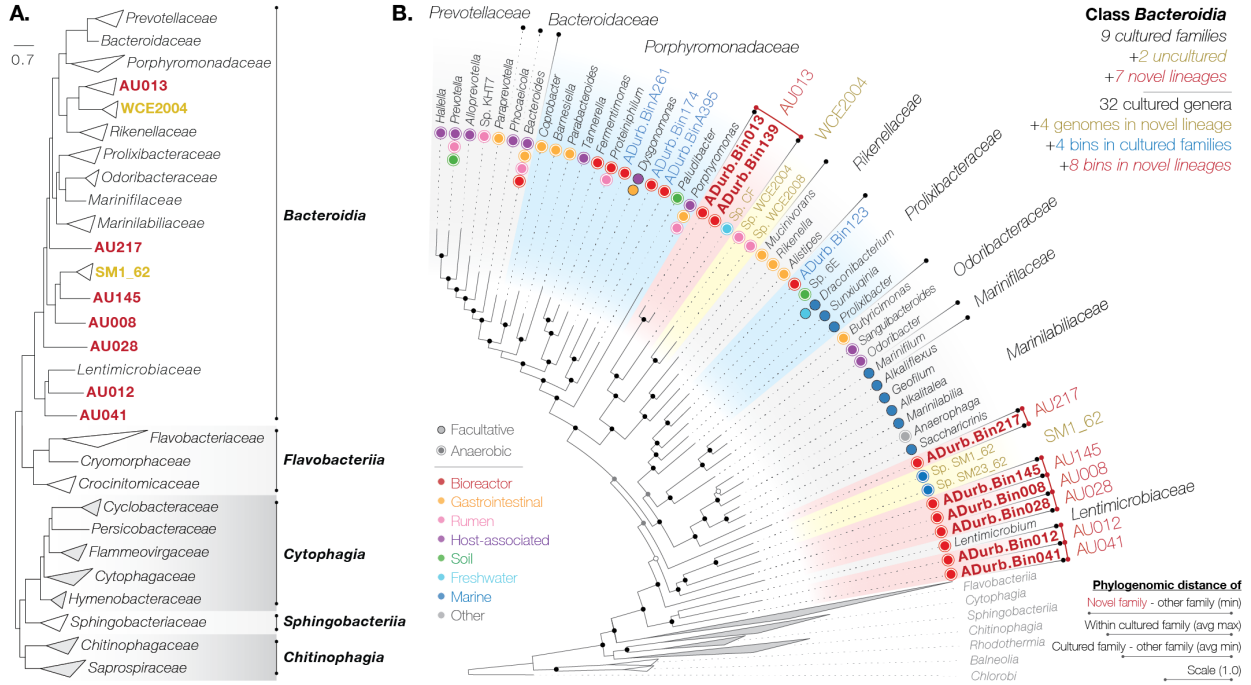

Figure. 6.1. Novel *Bacteroidia* lineages revealed by phylogenomic analysis of anaerobic digester bins. (A) Phylogenomic tree of all *Bacteroidetes* families and novel uncultivated families. (B) *Bacteroidia*-focused tree showing 9 cultured families along with newly identified families consisting of anaerobic digester bins (red label) and several previously published genomes with unknown phylogeny (yellow label) with tentative family-level lineage names assigned. These families are defined based on the criterion of having phylogenomic distances to the next phylogenetically nearest *Bacteroidia* genome relative higher than both (i) average maximum distance within cultured families and (ii) average minimum distance between cultured families. Several anaerobic digester bins also fall within cultivated families (blue label and background). Each genus/genome is also labeled by their associated habitat (colored circle) and whether they are strictly anaerobic (double circle) or facultatively aerobic (outlined circle).

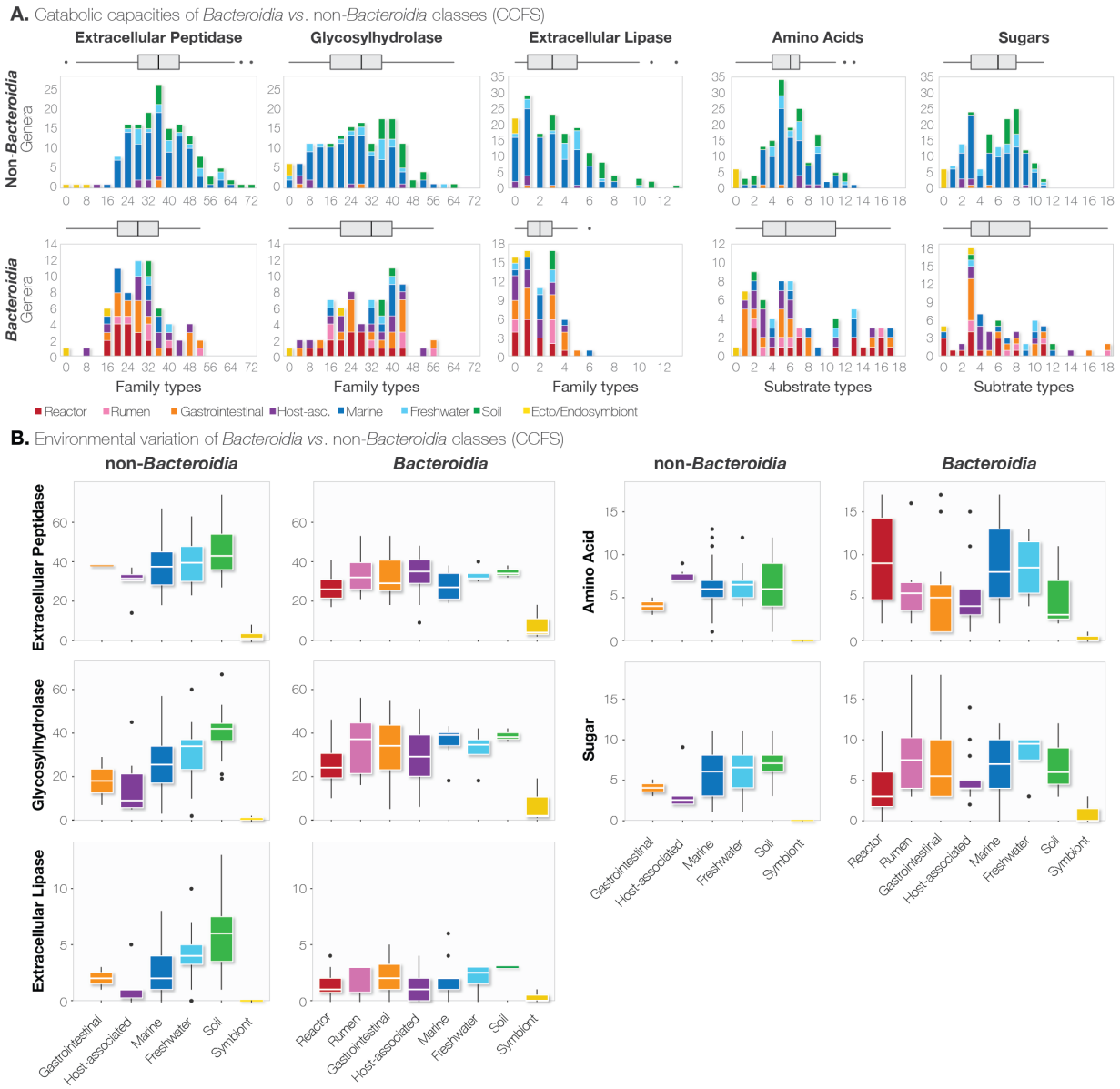

Figure 6.2. Comparison of *Bacteroidetes* metabolic capacities based on phylogeny and habitat. The overall capacity to degrade monomers (sugars and amino acids) and macromolecules (polysaccharides, peptides, and lipids) were respectively defined by the number of metabolizable sugars/amino acids and hydrolase families (glycosylhydrolase, extracellular peptidase, and extracellular lipase). A metabolic trait was considered part of each genus' "overall" capacity only if at least 10% of the genomes encoded the function. (A) Divided into *Bacteroidia* and non-*Bacteroidia* classes (*Chitinophagia*, *Cytophagia*, *Flavobacteriia*, and *Sphingobacteria*), histograms of the metabolic capacities of all genera (colored by habitat) are shown with summarizing boxplots. (B) Also divided by phylogeny, boxplots of the metabolic capacities of genera affiliated with specific habitats are shown.

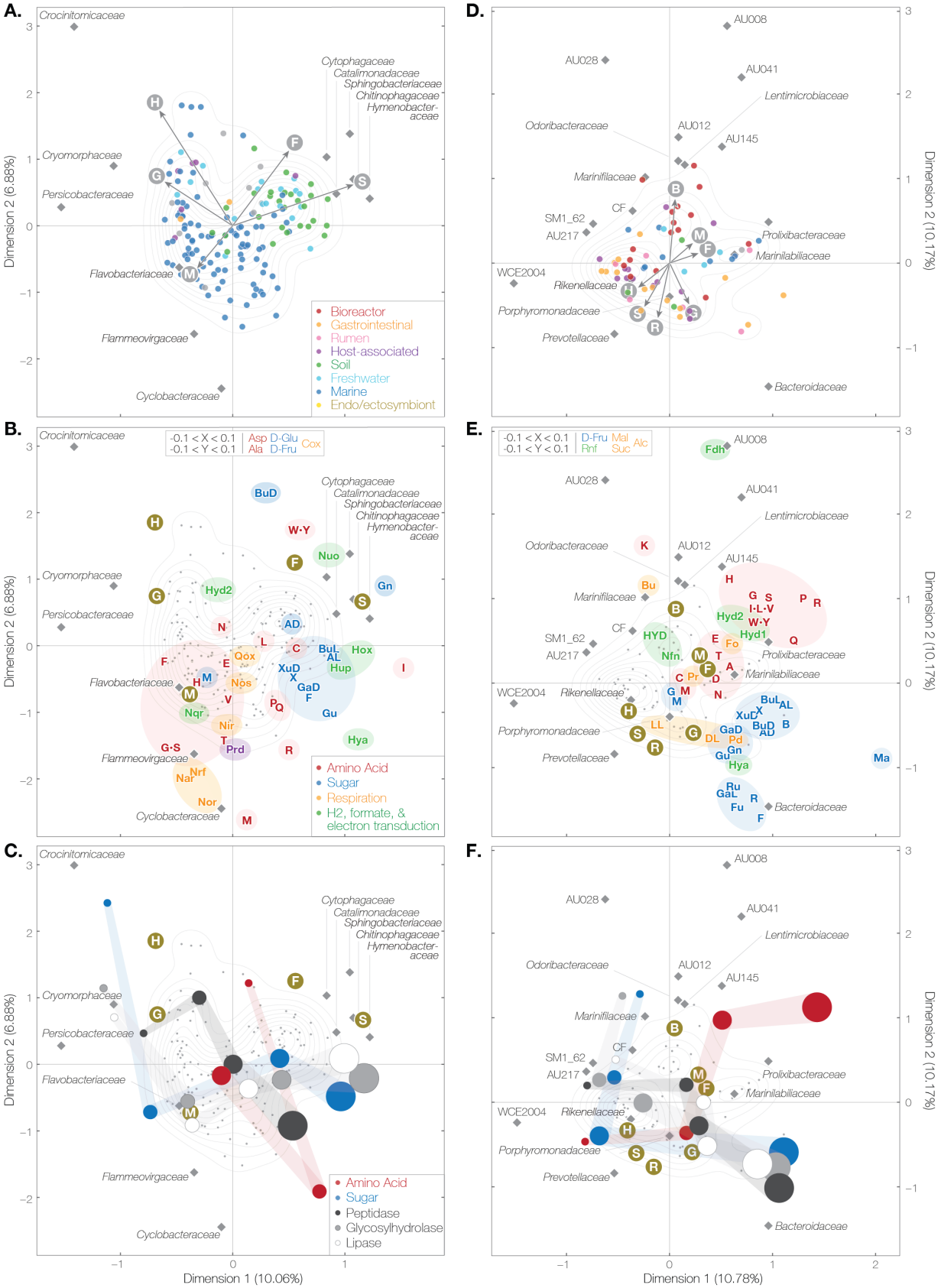

Figure 6.3. Multiple correspondence analysis (MCA) of *Bacteroidetes* genera metabolic capacities

Figure 6.3. (cont.) divided into *Bacteroidia* and non-*Bacteroidia* classes. Each genus is colored by their most associated habitat: bioreactor (red), gastrointestinal (orange), rumen (pink), host-associated (purple), marine (blue), freshwater (cyan), soil (green), and ecto/endosymbionts (yellow). The association of specific characteristics/traits (*i.e.*, metabolic, phylogenetic, and habitat) to these genera are shown as vectors (labeled at their respective coordinates, *i.e.*, weight and direction). For non-*Bacteroidia* classes (*Chitinophagia*, *Cytophagia*, *Flavobacteriia*, and *Sphingobacteria*), the results from a single MCA are divided into three panels. For the top and middle panels, the genera are respectively shown with (A) phylogeny (family) and habitat and (B) degradation of specific amino acids (red) and sugars (blue), types of electron transfer and H_2/formate metabolism (green), respiratory pathways (orange), and phototrophy (purple). (C) The bottom panel shows four levels of summarized metabolic capacities (increasing circle size with colored background connecting consecutive levels): number of degradable amino acids (red)/sugars (blue) and hydrolytic enzymes families (glycosylhydrolase – gray, peptidase – dark gray, and lipase – white). For amino acid degradation, the levels are defined by percentage of all standard amino acids (<5, <10, <15, and ≤20). For sugar degradation, the levels are defined by quartiles of all genera spanning the entire phylum ($<1^{st}$, $<2^{nd}$, $<3^{rd}$, and $\leq 4^{th}$ quartile). Similarly, hydrolytic capacities are divided by quartiles. Quartile values are shown in Appendix E figure E.2. The corresponding analyses for *Bacteroidia* are shown in (D), (E), and (F).

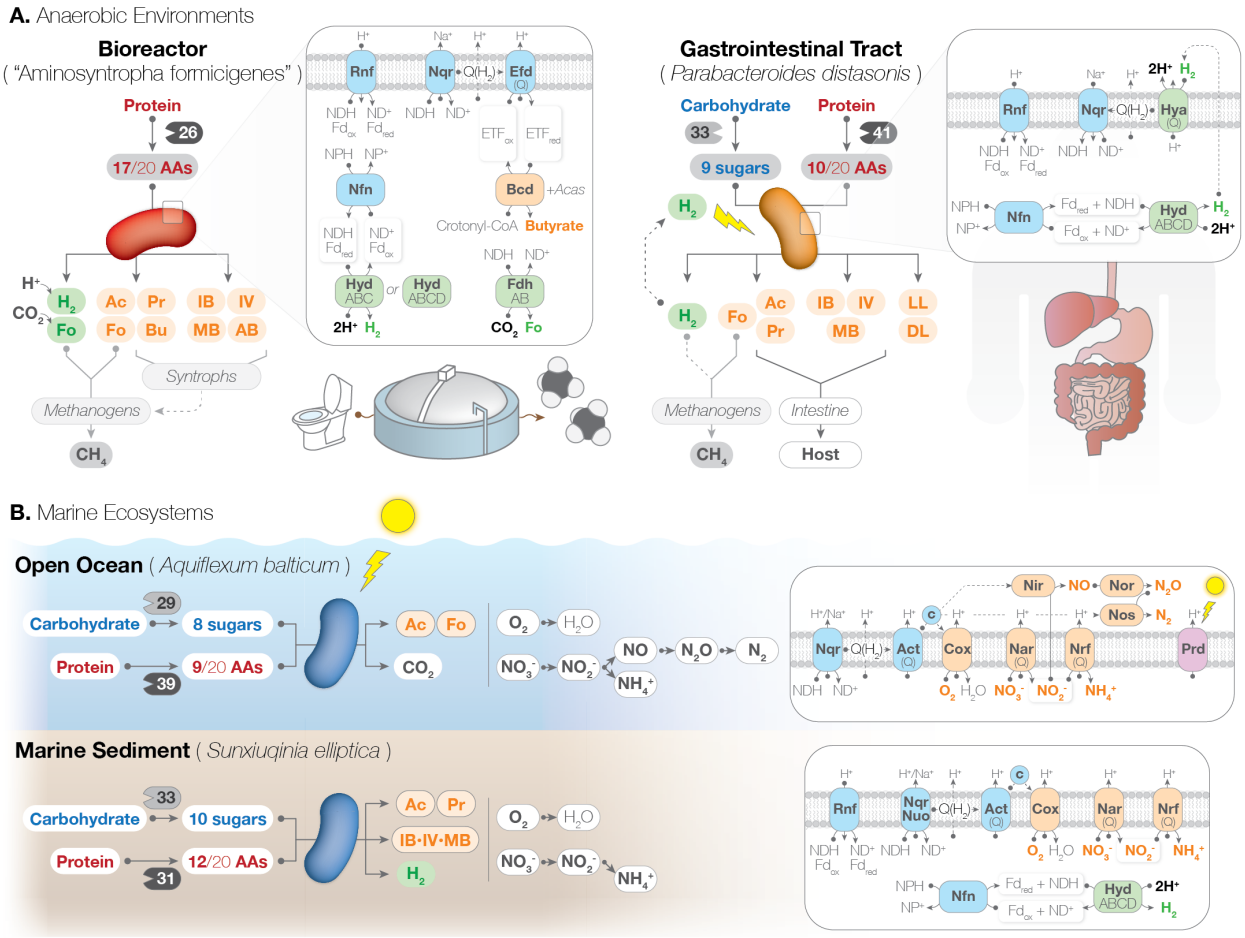

Figure 6.4. Highlighting *Bacteroidetes* with unique metabolic traits and novel ecological roles. (A) In anaerobic environments, bioreactor-affiliated "*Candidatus* Aminosyntropha formicigenes" can degrade the highest number of amino acids among all *Bacteroidetes*, fermentatively generate butyrate, and syntrophically interact with methanogens through H^+- and CO_2-reducing H_2 and formate (Fo) generation while gastrointestinal *Parabacteroides distasonis* can generate, oxidize, and even recycle H_2 through implementation of two distinct hydrogenases (electron-confurcating hydrogenase Hyd and quinone-reducing membrane bound hydrogenase Hya). (B) In marine ecosystems, *Aquiflexum balticum* can contribute to heterotrophy, phototrophy, and nitrogen cycling in the open ocean can harvest light energy (proteorhodopsin – Prd) and reduce various nitrogen species through both denitrification and ammonification while *Sunxiuqinia elliptica* can have ecological impact across marine sediment oxyclines by flexibly wielding O_2, NO_3^-/NO_2^-, and H^+ respiration. For each bacterium, the number of glycosylhydrolase and extracellular peptidase families and degradable amino acids and sugars are shown. Abbreviations: Rnf – nicotinamide adenine dinucleotide (NDH):ferredoxin (Fd) oxidoreductase; Nqr – Na^+-translocating NDH:quinone (Q) oxidoreductase; Nuo – NDH:Q oxidoreductase; Efd – electron transfer flavoprotein (ETF) dehydrogenase; Nfn – NADH-dependent Fd:NDH phosphate (NADPH) oxidoreductase; Bcd – ETF-dependent butyryl-CoA dehydrogenase; Fdh – formate dehydrogenase; Act – alternative complex III; Cox – cytochrome oxidase; Nar – nitrate reductase; Nir – nitrite reductase; Nor – nitric oxide reductase; Nos – nitrous oxide reductase; Nrf – nitrite reductase (ammonifying); Ac – acetate; Pr – propionate; Bu – butyrate; IB – isobutyrate; MB – 2-methylbutryrate; IV – isovalerate; AB – 4-aminobutyrate; LL – L-lactate; DL – D-lactate.

CHAPTER 7

CONCLUSION

Methanogenic wastewater treatment is an attractive biotechnology due to low operation cost and generation of methane gas as a recoverable fuel, but development of the technology has been hindered by our lack of insight into the biological processes taking place in the reactors. With the advent of molecular biology, engineers and ecologists could finally peer into the microbial communities in wastewater treatment bioreactors in hopes to gain insight for engineering solutions for improving treatment efficiency and also methane recovery (*18, 19*). This is an especially critical mission for this generation as increasing awareness of environmental issues necessitates environmental stewardship and measures to acquire sustainable energy. However, such efforts revealed overwhelming microbial diversity, including many populations whose contributions could not be addressed due to difficulties in cultivation (*18, 19*). The research presented in the preceding chapters shows one of the first efforts in implementing "metagenomics" and "metatranscriptomics" as novel approaches for the prediction of metabolic functions and ecological roles of uncultivated *Bacteria* and *Archaea* in methanogenic wastewater treatment processes. Furthermore, we step beyond conventional "omics" approaches and integrate cutting-edge sequencing technology (single-cell genomics in Chapter 5 and long-read sequencing in Chapter 6) to achieve high quality genomes for uncultivated organisms. As a result, we successfully unravel the physiology of hitherto uncharacterized syntrophic metabolizers, methanogens, and fermenters and delineate overlooked microbial catabolism, interactions, and ecology driving the conversion of organic waste to methane in both industrial and municipal processes.

"Omics"-based approaches can significantly accelerate exploration of yet-to-be-cultured organisms (*53, 59*) and thereby characterize the uncharted majority of wastewater treatment microbial populations to reveal previously unnoticed biological phenomena. We can further deduce the impact of such phenomena on wastewater treatment and use such inferences as seeds for inspiring novel engineering solutions. Investigation of uncultured syntrophic metabolizers driving treatment of petrochemical industry wastewater reveals novel widespread aromatic compound degradation pathways that produce novel byproducts (*i.e.*, butyrate and benzoate) and have unique thermodynamic properties (Chapter 2 and 3). Based on thermodynamic theory, the presence and/or accumulation of these compounds likely inhibits syntrophic aromatic compound catabolism. Thus, such overlooked inhibitors clearly require specific attention. Indeed, butyrate is a common fermentation product and benzoate is often present in many petrochemical wastewater types, suggesting that (i) exploration of the influence of these byproducts on aromatic pollutant treatment is necessary as a next step and (ii) current processes ought to be engineered

to prevent or minimize the inhibition imposed by these compounds. Although not part of this dissertation, on-going experiments confirm that benzoate accumulation does stunt degradation of other aromatic compounds.

Likewise, further investigation of uncultured organisms uncovers ecological processes that engineers may be able to exploit for bioreactor optimization. Characterization of a novel archaeon reveals a unique mode of methanogenesis contributing to an overlooked component of methane production in wastewater treatment (Chapter 4). This organism can produce methane from methyl compounds, which are often associated with animal, plant, and algae waste and specific types of industrial wastewater. This suggests that such waste/wastewater can be an excellent source for harvesting methane gas. For example, food waste is rich in methyl compounds (*e.g.*, phenolic esters and pectin) and is a major issue in the U.S. (~31% of food wasted according to USDA's 2013 Economic Research Service); methanogenic treatment would decrease the environmental impact and allow energy recovery from a major domestic waste. In addition, with the growing attention on algae bioreactors for biofuel production, treatment of algae waste containing high concentrations of methyl compounds (*e.g.*, dimethylsulfoniopropionate) through methanogenic bioreactors may support concomitant biofuel and biogas production from a single process. Thus, metabolic characterization of uncultivated organisms can inspire new processes relevant to current societal issues.

The rigorous application of metagenomics also allowed identification of an organism indispensable to municipal wastewater treatment. Despite the long history of anaerobic digestion since 1895, we present the first characterization of a highly abundant, prevalent, and active microorganism in digesters responsible for proteinaceous waste degradation and capable of single-handedly catabolizing most protein-derived amino acids (Chapter 6). As protein is the largest constituent of waste fed into digestion (30~40%) (*293*), optimization of digester operation for the activity of this unique organism may lead to improvement of the process (*e.g.*, stable protein removal and acid production). Moreover, addition of this organism to digesters with poor performance may be a viable bioaugmentation strategy. Currently, neither biostimulation nor bioaugmentation strategies are implemented for the improvement of anaerobic digestion; thus, further biochemical and reactor-based experiments to fully characterize the properties of this novel organism may lead to new engineering solutions of methanogenic biotechnology.

Given the insight obtained through the research presented in this dissertation, the application of omics technology has the clear potential to open new windows for wastewater treatment engineering. On the other hand, it is critical to recognize and delineate the advantages, disadvantages, and limitations of this approach to responsibly transmit insight at the microbial level to engineers. Omics' advantages are (i) acceleration of uncovering the microbial ecology of bioreactors by circumventing time-consuming

cultivation- and reactor-based experiments, (ii) in-depth physiological characterization of key microbial populations that experimental work may not necessarily reveal, and (iii) targeted investigation of metabolism and interactions taking place inside of reactors rather than in the laboratory (*i.e.*, *in situ* rather than *in vivo*). However, all such advantages hinge on (a) the quality of genomes recovered through metagenomics and (b) the validity of "metabolic reconstruction" (*i.e.*, prediction of metabolic behavior from genomes).

Although algorithms for extracting microbial population genomes from metagenomes have drastically improved even over the past few years, it is nearly impossible to obtain complete genomes and eliminate contamination (*i.e.*, genome fragments were falsely categorized that actually belong to a different organism). One predicts that an organism can perform certain metabolism by detecting corresponding complete metabolic pathways in the genome. Thus, we may falsely predict that a target organism is incapable of certain metabolism if a necessary gene is encoded on the missing fragments of an incomplete genome. Likewise, even if a target organism lacks a metabolic pathway in reality, contaminating genes may encode this pathway and lead to false prediction of metabolic capacity. Moreover, software may incorrectly predict the function of particular genes, leading to further convolution. Thus, researchers employing metagenomics must take measures to minimize such false negatives and positives.

In this dissertation, we attempt to address the aforementioned issues through four approaches: multi-algorithm genome recovery, manual genome curation, manual gene annotation, and holistic metabolic prediction. For genome recovery, we employ at least two methods (*e.g.*, MetaBAT and MaxBin), compare their results, and synthesize the information to identify mis-categorized DNA sequences. Although too detailed to elaborate here, testing current software on many methanogenic bioreactor metagenomes has revealed that particular organisms are prone to poor genome recovery and also tend to contaminate other genomes. Thus, with accumulated experience in bioreactor metagenomics, we can preemptively and deliberately search for mis-categorized DNA sequences and manually curate sequence categorization to both improve genome recovery for problematic organisms and also reduce contamination of other organisms. For each refined genome, we manually predict the functions of genes involved in target metabolic pathways through careful investigation of individual gene's phylogeny (*e.g.*, homology and protein trees), structure (*e.g.*, protein domain, active sites, and protein folding), and organization (*e.g.*, operons). For example, phylogeny can tell us how likely the target gene may have the predicted function based on phylogenetic proximity to proteins with known biochemical activity; structure can reveal fine protein domains that determine the biochemical activity; and gene organization can elucidate potential association between multiple genes that may form coherent metabolic pathways. Finally, for an organism to be able to degrade a specific compound, it must not only encode the corresponding metabolic pathway,

but also ancillary pathways for energy recovery and biosynthesis from the particular compound. Thus, we only report catabolism if all three pathways are present. By integrating the aforementioned rigorous approaches, we can confidently predict microbial function from genomes. We strongly believe this level of fidelity to precision is necessary to generate coherent results; otherwise, metagenomics can be enormously deceptive and lead to confusion in the field of engineering.

We would like to add that implementation of omics alone has major limitations regardless of how thorough we are. At best, metagenomics can only reveal the capacity to perform certain metabolism; an organism may not necessarily actively use the annotated metabolism in the studied ecosystem. Metatranscriptomics (RNA-based whole community gene expression profile) is a major step forward that allows us to determine which metabolic pathways are actually active *in situ* and, whereby, properly define microbial function, interactions, and ecology. Further, we can implement thermodynamic theory to judiciously evaluate the viability of particular metabolic functions and even predict novel ecological interactions (Chapter 3, 4, and 5). However, it is essential to be aware that omics-based predictions cannot address the activity level, kinetics, temporal variation, and spatial distribution of microbial metabolism within bioreactors. For example, although identification of a novel metabolic pathway with favorable thermodynamics is interesting, it is virtually irrelevant if the metabolism only reaches 5% of the reaction rate of a conventional pathway and only occurs under specific anomalous reactor conditions. On the other hand, if a newly discovered metabolism is found to be significantly faster and more widely implemented compared to a known pathway, we may overhaul our previous understanding and focus efforts on characterizing the novel metabolism. In any case, investigation of such behavior and properties require meticulous laboratory- and reactor-based experiments. Thus, we believe that omics-based studies ought to inspire complementary experiments, and *vice versa* – continuous feedback between the two research approaches is paramount for properly furthering applied microbiology in the context of wastewater engineering.

To address the limitations outlined above and achieve the full potential of metagenomics, we intend to utilize omics in combination with cultivation-based experiments for future investigations of microbial ecology in methanogenic bioreactors. Each recovered genome contains an enormous amount of information that would theoretically allow us to peer into individual organism's physiology that other methods may not reveal. With each metagenome containing hundreds of genomes, the insight we can obtain for each ecosystem is unfathomable. However, with accumulated experience in how and where to look in the genomes, we can efficiently pinpoint uncultivated organisms with novel, relevant, and exploitable metabolic features. Furthermore, by predicting physiological properties through the genomes, we can design organism-specific cultivation strategies to isolate target organisms for downstream

laboratory-based characterization of the aforementioned metabolic features that omics cannot predict. With insight attained through "genome-informed" cultivation of major microbial populations, we can refine ecological models of the studied methanogenic wastewater treatment processes. In parallel, we also plan to characterize the novel metabolic pathways by artificially inserting these pathways into model organisms (*e.g.*, *Escherichia coli* and *Geobacter sulfurreducens*) and studying the biochemical and kinetic properties, a field known as "synthetic biology." We believe the combination of metagenomics with genome-informed cultivation and synthetic biology will allow microbial ecologists and engineers to reach new frontiers for development of engineering solutions for the protection of our environment.

APPENDIX A

SUPPLEMENTARY MATERIAL TO CHAPTER 2

A.1 – Supplementary Figures

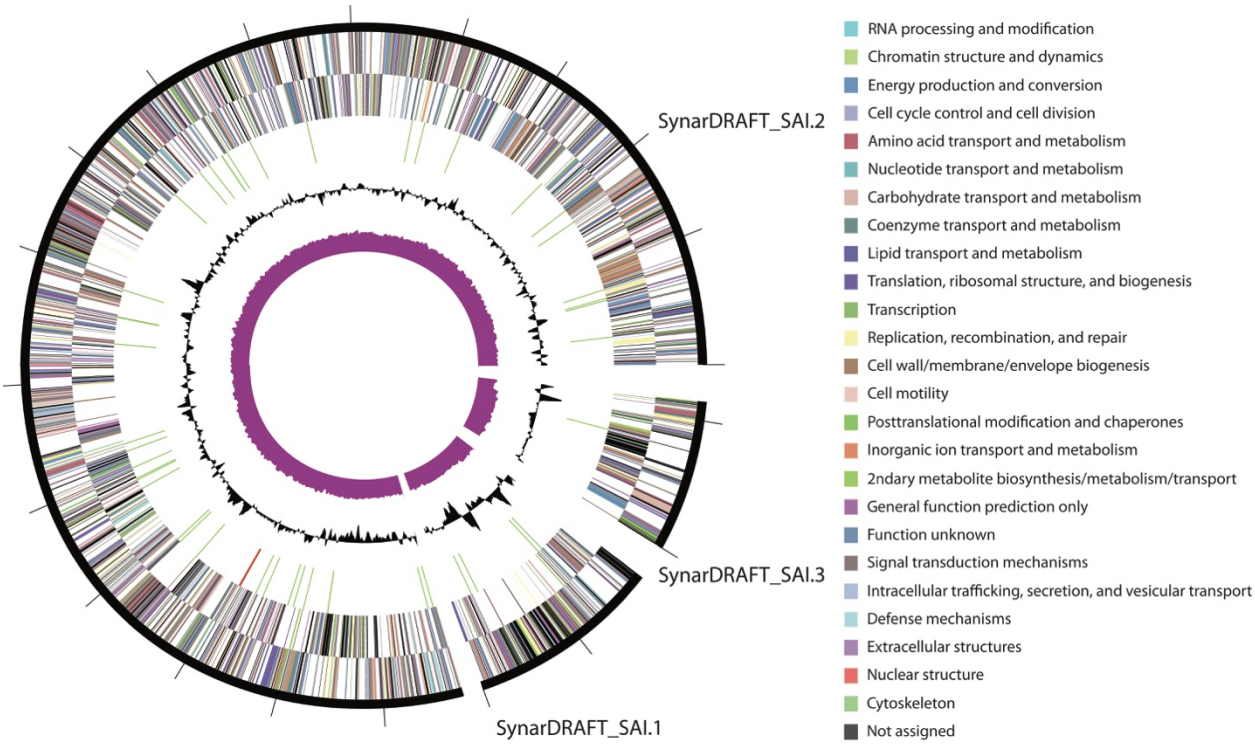

Figure A.1. A circular map of the strain UI genome (from outer to inner circle) with forward strand genes and reverse strand genes colored by COG categories, RNA genes (green tRNA and red rRNA), GC content, and GC skew.

A.2 – Supplementary Tables

Refer to the supplementary file "Nobu2017_AppendixA_Tables.xlsx" for the following tables:

Table A.1. *Syntrophorhabdus aromaticivorans* strain UI genes relevant to general physiological function.

Table A.2. *Syntrophorhabdus aromaticivorans* strain UI genes encoding aromatic compound degradation and syntrophic electron flow and their top blastp hits.

Table A.3. *Syntrophorhabdus aromaticivorans* strain UI genes encoding aromatic compound degradation and syntrophic electron flow and their similarity to relevant organisms' genes.

APPENDIX B

SUPPLEMENTARY MATERIAL TO CHAPTER 3

B.1 – Supplementary information

B.1.1 – Metagenome results

For community U1, samples for metagenomic sequencing were taken from the same sampling port on three different dates (February 23rd, April 1st, and May 28th 2014) from a lab-scale bioreactor treating synthetic purified terephthalic acid process wastewater. For community B three samples were taken from different sampling ports on July 14th, 2014 from a full-scale bioreactor treating purified terephthalic acid process wastewater. Detailed operation data for both reactors are shown later in the supplementary material. For metagenomes of both communities, Illumina HiSeq2500 paired-end reads from the three samples described above were co-assembled. The metagenomic assembly for community U1 consisted of 244088 contigs with a total length of 936.8 Mb, N50 of 7607 bp, and N90 of 1343 bp. Likewise, the community B metagenome consisted of 287604 contigs with a total length of 960.6 Mb, N50 of 5288 bp, and N90 of 1279 bp. Binning the assembled metagenomes for community U1 and B respectively yielded 127 and 97 bins that had at least 65% genome completeness and <10% contamination as estimated by CheckM (Table B.1 and B.2) (*58*).

B.1.2 – Metatranscriptomic sequencing and mapping overview

Metatranscriptomes were sequenced for community U1 and B under benzoate, terephthalate, and trimellitate amendment at the time points shown above in duplicate (Fig. B.1) and directly from each bioreactor on April 1st, 2014 and July 14th, 2014. For bioreactor U1, the sampling port is located in the middle of the sludge bed. For bioreactor B, the sampling port is located 3 feet from the bottom of the sludge bed. Metatranscriptomic sequencing generated 36.2~40.9 million single-end 100bp reads for each sample (Table B.4). For each metatranscriptome, 37.6~95.9% of the reads were mapped using a 100% nucleotide similarity cutoff.

B.1.3 – Reactor specifications

Communities U1 and B are respectively derived from lab-scale and full-scale methanogenic granular upflow anaerobic sludge blanket bioreactors treating purified terephthalic acid process wastewater, which typically contains acetate, benzoate, terephthalate, trimellitate, isophthalate, orthophthalate, 4-methylbenzoate, and 4-carboxybenzaldehyde (*147, 294*). The reactors were sampled when operation and

physical parameters were stable (*e.g.*, neutral pH ~7, temperature of 37°C, and >90% removal for most above compounds). For bioreactor B, samples were taken from different sampling ports on the same day.

B.1.4 – Aromatic compound metabolism

Previous studied have thoroughly characterized syntrophic benzoate (BA) catabolism (*27, 84, 114, 295, 296*), making annotation quite simple. On the other hand, we have yet to properly characterize or identify enzymes involved in terephthalate (*33, 87, 143*) and trimellitate metabolism. Given that degradation of terephthalate (TA) and trimellitate (TM) very likely proceeds through a benzoyl-coenzyme A intermediate, we suspect that these aromatic compounds undergo coenzyme A (CoA) activation and subsequent decarboxylation (Fig. 2.2). Through comparison with previously identified putative CoA ligases and decarboxylases, we identify a large variety of homologs in *Pelotomaculum* and *Syntrophorhabdaceae* bins obtained in this study (Table B.3).

For TA degradation, *Pelotomaculum* bins encode putative TA--CoA ligases and terephthalyl-CoA decarboxylases (*e.g.*, A4E54_02221 and A4E54_02223) closely related to those found in our previous metatranscriptomics- and metaproteomics-based studies on a hyper-mesophilic (46-50°C) TA-treating methanogenic bioreactor (HMTAb) (*110*). However, gene expression data acquired from TA-amended treatments shows very little expression of such genes (Table B.4). Moreover, a *Pelotomaculum* (PtaB.Bin117) that only encodes aromatic decarboxylases related to those found in HMTAb does not have any notable activity during TA amendment. In contrast, the *Pelotomaculum* and *Syntrophorhabdaceae* bins that respond to TA amendment all encode a well-conserved and highly expressed gene cassette containing homologs of the HMTAb CoA ligase and decarboxylase (*e.g.*, A4E56_03078 and A4E56_03079; Table B.4). Thus, we suspect that two types of TA degradation enzymes must exist – mesophilic and thermophilic.

The newly identified TA degradation gene cassette contains four to five genes – cation:aromatic acid symporter, CoA ligase, aromatic decarboxylase, decarboxylase partner protein, and regulator. This gene cassette has consistent structure and homology across the bins and is also highly expressed under TA amendment. This suggests that this is a terephthalate degradation (Tpd) gene cassette encoding a cation:TA symporter (TpdT), TA--CoA ligase (TpdY), terephthalyl-CoA decarboxylase (TpdA), terephthalyl-CoA decarboxylase partner protein (TpdB) and an associated regulator (TpdR).

$$TA_{out} + Na^{+}_{out} \rightarrow TA_{in} + Na^{+}_{in} \quad \text{(eq. 1)}$$

$$TA_{in} + ATP + CoA \rightarrow \text{Terephthalyl-CoA} + AMP + PP_i \quad \text{(eq. 2)}$$

$$\text{Terephthalyl-CoA} + H_2O \rightarrow \text{Benzoyl-CoA} + HCO_3^- \quad \text{(eq. 3)}$$

For the hypermesophilic Tpd gene cassette, we correspondingly assign abbreviations as TpdT2, Y2, A2, and B2. This cassette also contains a hypothetical protein (TpdC2) and a carbonic anhydrase (TpdD2). Presence of a carbonic anhydrase suggests that the hypermesophilic TA decarboxylation reaction adds one more step:

$$\text{Terephthalyl-CoA} + H_2O \rightarrow \text{Benzoyl-CoA} + HCO_3^- \quad \text{(eq. 3)}$$
$$HCO_3^- + H^+ \rightarrow CO_2 + H_2O \quad \text{(eq. 4)}$$
$$\text{Terephthalyl-CoA} + H^+ \rightarrow \text{Benzoyl-CoA} + CO_2 \quad \text{(eq. 5)}$$

Thus, the net hypermesophilic reaction generates CO_2 rather than HCO_3^-. However, the reason for having such a reaction at higher temperatures remains unclear. With a temperature increase from mesophilic (37°C) to 50°C, the carbonic acid dehydration (eq. 4) rate increases by 700%, given the carbonic acid dehydration activation energy (E_a) of 66.6 kJ mol^{-1}. Thus, a carbonic anhydrase would be more useful at lower temperatures where the dehydration rate is lower. On the other hand, a temperature increase to 50°C also incurs a 26.7% decrease in CO_2 solubility, given enthalpy of CO_2 dissolution ($\Delta_{sol}H$) of 19.95 kJ mol^{-1} at 298.15 K. Assuming *Pelotomaculum* regulates its cytosolic bicarbonate at a fixed concentration (*e.g.*, 50mM) regardless of temperature, then the temperature-induced CO_2 solubility decrease would also decrease the ΔG of eq. 4 by 1.2 kJ mol^{-1}. Therefore, *Pelotomaculum* may exploit this marginal energy through a carbonic anhydrase to increase the thermodynamic favorability of TA degradation at higher temperatures. Another possibility is that the carbonic anhydrase serves as a H^+ sink to prevent cytosol acidification; however, the reason that such a reaction would become necessary at higher temperatures is unclear.

In addition, we identify a Tpd-like gene cassette in the *Syntrophorhabdus* bins that have high homology with genes found in *Syntrophorhabdus aromaticivorans* strain UI (*e.g.*, A4E62_01832-35; Table B.3). Given that strain UI degrades isophthalate (IA) and that none of the genes in these cassettes respond to TA or TM addition, we suspect that these cassettes encode IA degradation (Ipd) through a cation:IA symporter (IpdT), IA--CoA ligase (IpdY), isophthalyl-CoA decarboxylase (IpdA), and two decarboxylase partner proteins (IpdB and IpdC). Moreover, for Ipd-encoding bins, the expression of these Ipd genes are greater than the top octile *in situ*, but extremely low under any of the studied conditions (Table B.5 and B.9).

As for TM degradation, we find that the four *Syntrophorhabdaceae* members that respond to TM amendment all encode and highly express a putative TM degradation (Tmd) gene cassette containing Tpd homologs – cation:TM transporter (TmdT), TM--CoA ligase (TmdY), aromatic acid decarboxylase (TmdA), decarboxylase partner protein (TmdB), and associated regulator (TmdR) (Table B.3 and B.7). While we expected that two decarboxylases are necessary to complete the conversion of TM to benzoyl-CoA, we could only identify one decarboxylase (TmdA) with high gene expression in all TM degraders.

Thus, it remains unclear whether TmdA can accomplish sequential decarboxylation of trimellityl-CoA to benzoyl-CoA or the second decarboxylase was simply not detected due to the metagenomic nature of this study.

The above analysis indicates that TA, IA, and TM are all likely transported into the cell through a cation:aromatic acid symporter. However, we could not identify any homologs of such transporters or known H^+:BA symporters (*e.g.*, BenE), implicating that BA is transported through an alternative mechanism by syntrophs. Based on comparative genomics, we identify an uncharacterized ABC-type transporter conserved amongst anaerobic BA degraders with genomes available – *Syntrophus aciditrophicus* strain SB, *Syntrophorhabdus aromaticivorans* strain UI, and *Geobacter metallireducens* strain GS-15 (*27, 295*). The putative BA transporter (Btr) contains a periplasmic ligand-binding receptor (BtrX; *e.g.*, SYN_00789 or Gmet_1823), two permease subunits (BtrAB; *e.g.*, SYN_00788-787 or Gmet_1822-1821), and two ATP-binding subunits (BtrCD; *e.g.*, SYN_00786-785 or Gmet_1820-19). Indeed, these genes are also conserved among the syntrophic aromatic compound degraders in this study. Moreover, we confirm expression of BtrXABCD by these syntrophs during BA amendment (Table B.3 and B.8). Interestingly, during TA/TM amendment, the TA and TM degraders express BtrXABCD and BA--CoA ligase, which are both unnecessary during TA and TM catabolism. This suggests that TA and TM degraders can convert benzoyl-CoA (a TA/TM degradation intermediate) to BA and further secrete this BA. In agreement, we also detect other non-TA/TM-degrading syntrophs expressing BA degradation, including BtrXABCD.

B.2 – Supplementary Figures

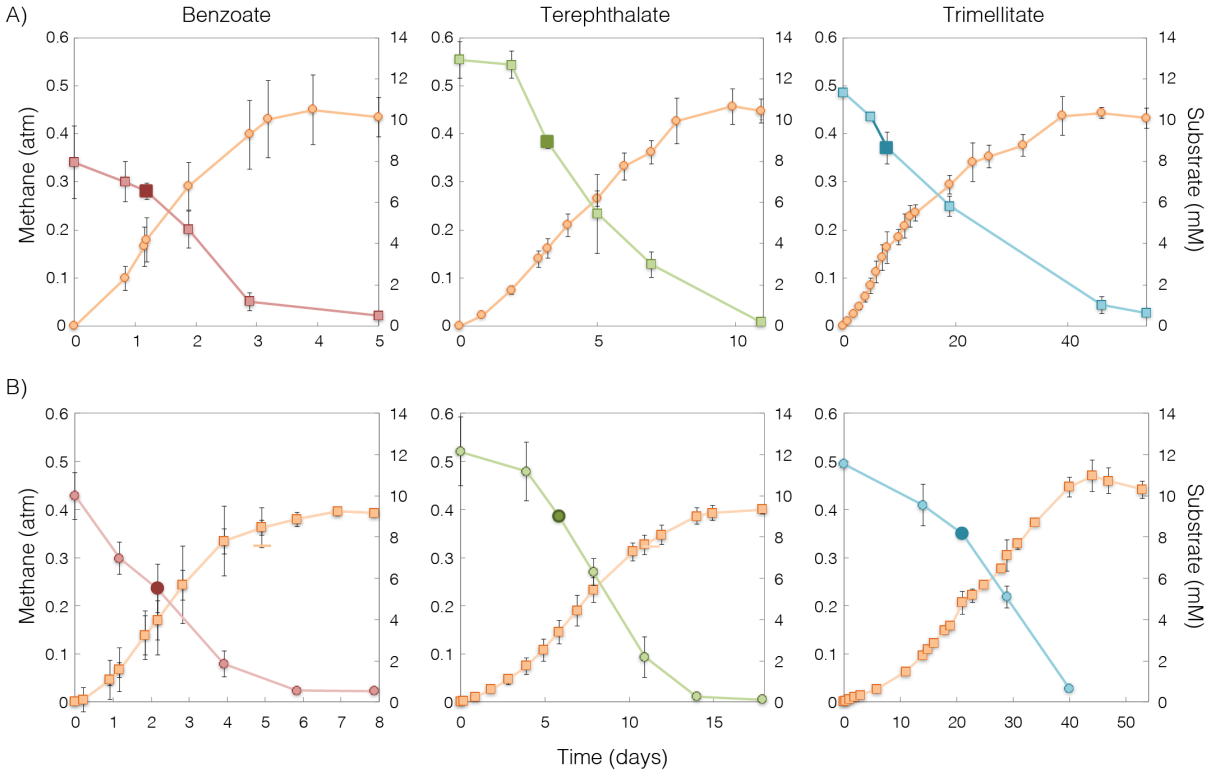

Figure B.1. Aromatic compound degradation and methane production during benzoate, terephthalate, and trimellitate amendment for (A) community U1 and (B) community B. The left- and right-hand axes correspondingly indicate atmospheric methane concentration (atm) and substrate concentration (mM). Time points at which samples for metatranscriptomics were taken are shown with a darker marker.

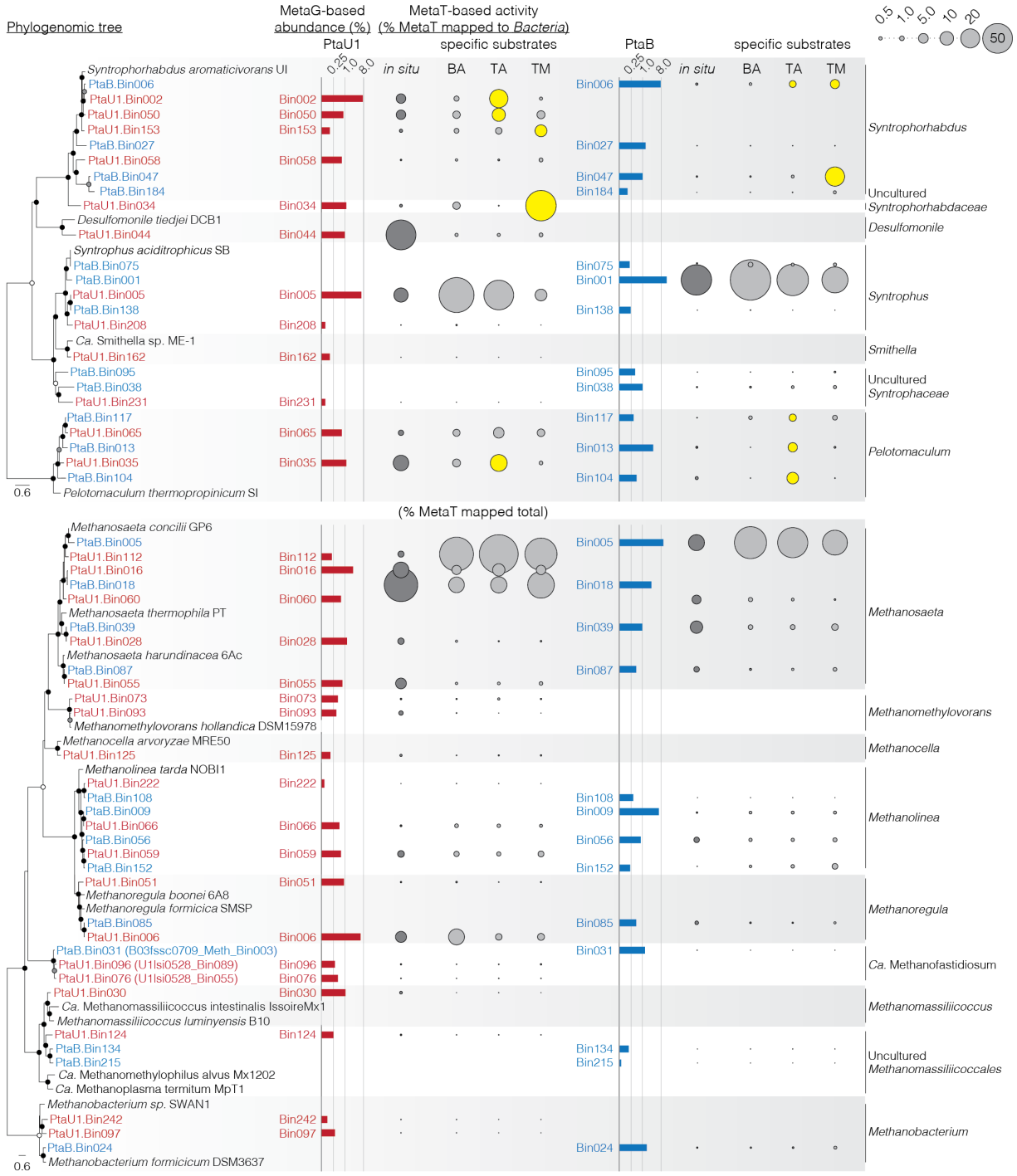

Figure B.2. Metagenomics-based abundance and metatranscriptomics-based activity of syntrophs and methanogens. The phylogeny of the syntroph-related populations identified in communities U1 and B are shown with their corresponding abundance estimated from their genomes' metagenomic coverage calculated as the percentage of metagenomic reads mapped to each genome relative to the total reads mapped to all constructed bacterial draft genomes (both *Bacteria* and *Archaea*). The estimated activity of these syntrophs in response to specific substrates (benzoate – BA, terephthalate – TA, and trimellitate – TM) are shown as the percentage of metatranscriptomic reads mapped to each genome relative to total reads mapped to all constructed bacterial draft genomes. Likewise, the activity of the methanogens

Figure B.2. (cont.) are shown as the percentage of reads mapped to each genome relative to the total reads mapped to all constructed draft genomes (both *Bacteria* and *Archaea*). For TA and TM amendment, the circle representing activity level is marked yellow if the population's estimated activity increased at least three-fold relative to the BA treatment. Syntroph populations with notable abundance and activity (>1% for both estimated abundance and activity) are assigned population names and symbols (colored circles with abbreviation of population name) for the purposes of this study.

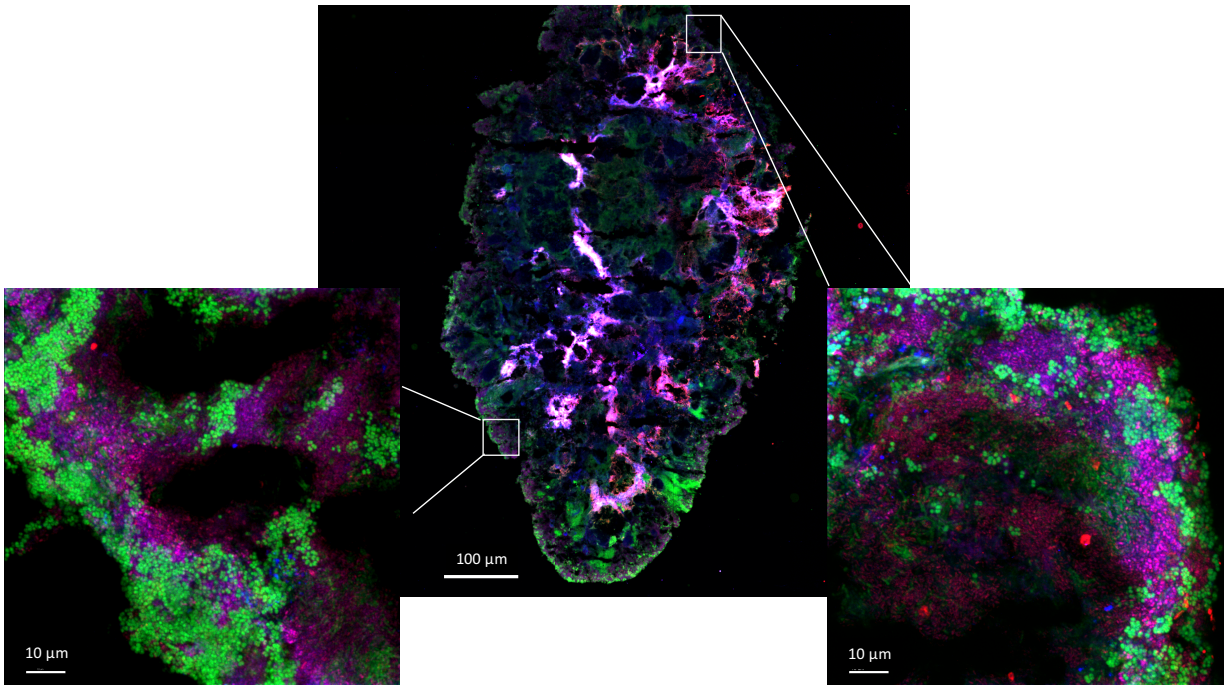

Figure B.3. Visualization of Deltaproteobacteria and methanogens inside the community U1 granules using fluorescence in situ hybridization. Granules collected from bioreactor U1 were fixed, sectioned, and attached to glass slides. All *Bacteria* (blue), *Deltaproteobacteria* (red, appearing purple in combination with blue from *Bacteria*), and *Archaea* (green) were visualized using the following probes: EUB338mix (EUB, EUBII, EUBIII), DELTA495 (DELTA495a, b, and c), and ARC915 respectively. The distribution of *Deltaproteobacteria* syntrophs and methanogens in the granule is seemingly random, unlike layered granules observed in some bioreactors treating other types of wastewater (*e.g.*, brewery wastewater). The diverse syntroph-syntroph and syntroph-methanogen distances likely create a heterogeneous environment for niche partitioning and segregation. Moreover, we can observe several morphotypes for both syntrophs (short and large rods) and methanogens (coccoid *Methanoregula* (*297*) and long blunt rod *Methanosaeta* (*298*)).

B.3 – Supplementary Tables

Refer to the supplementary file "Nobu2017_AppendixB_tables.xlsx" for the following tables:

Table B.1. Metagenomic bins reconstructed from metagenomes of the full-scale reactors in this study and their estimated phylogeny, completeness, and contamination based on CheckM; phylogeny based on phylogenomic tree; and metagenomic coverage.

Table B.2. Metagenomic bins reconstructed from metagenomes of the lab-scale reactor in this study and their estimated phylogeny, completeness, and contamination based on CheckM; phylogeny based on phylogenomic tree; and metagenomic coverage.

Table B.3. Locus tags for genes encoding aromatic compound degradation, H2/formate generation, and electron transduction found in Pelotomaculum (blue), Syntrophorhabdaceae (red), and Syntrophus (green) bins recovered from community U1 and B.

Table B.4. Characteristics of RNAseq metatranscriptomes from lab-scale and full-scale reactors and corresponding experiments.

Table B.5. Rank of gene expression level (RPKMB) for each bin. Each gene's rank is shown as a number (1 is highest expression level) and also the top X% of ranked genes it belongs to. Genes with RPKMB ranks greater than the top octile (>87.5% of genes) are marked red; genes less than the top quartile but greater than the next octile (or third quartile; >75%) a gradient from red to white with decreasing rank; and remaining genes a gradient from white (third quartile) to blue (minimum).

Table B.6. Gene expression level for aromatic compound degradation, H2/formate generation, and electron transduction genes in bins from community U1 and B during terephthalate amendment. The expression level is calculated separately for individual bins as reads per kilobase transcript per million reads mapped to the bin (RPKMB), averaged between two metatranscriptomes. Pelotomaculum (blue), Syntrophorhabdaceae (red), and Syntrophus (green).

Table B.7. Gene expression level for aromatic compound degradation, H2/formate generation, and electron transduction genes in bins from community U1 and B during trimellitate amendment. The expression level is calculated separately for individual bins as reads per kilobase transcript per million reads mapped to the bin (RPKMB), averaged between two metatranscriptomes. Pelotomaculum (blue), Syntrophorhabdaceae (red), and Syntrophus (green).

Table B.8. Gene expression level for aromatic compound degradation, H2/formate generation, and electron transduction genes in bins from community U1 and B during benzoate amendment. The expression level is calculated separately for individual bins as reads per kilobase transcript per million reads mapped to the bin (RPKMB), averaged between two metatranscriptomes. Pelotomaculum (blue), Syntrophorhabdaceae (red), and Syntrophus (green).

Table B.9. Gene expression level for aromatic compound degradation, H2/formate generation, and electron transduction genes in bins from community U1 and B in situ. The expression level is calculated separately for individual bins as reads per kilobase transcript per million reads mapped to the bin (RPKMB). Pelotomaculum (blue), Syntrophorhabdaceae (red), and Syntrophus (green).

APPENDIX C

SUPPLEMENTARY MATERIAL TO CHAPTER 4

C.1 – Supplementary Figures

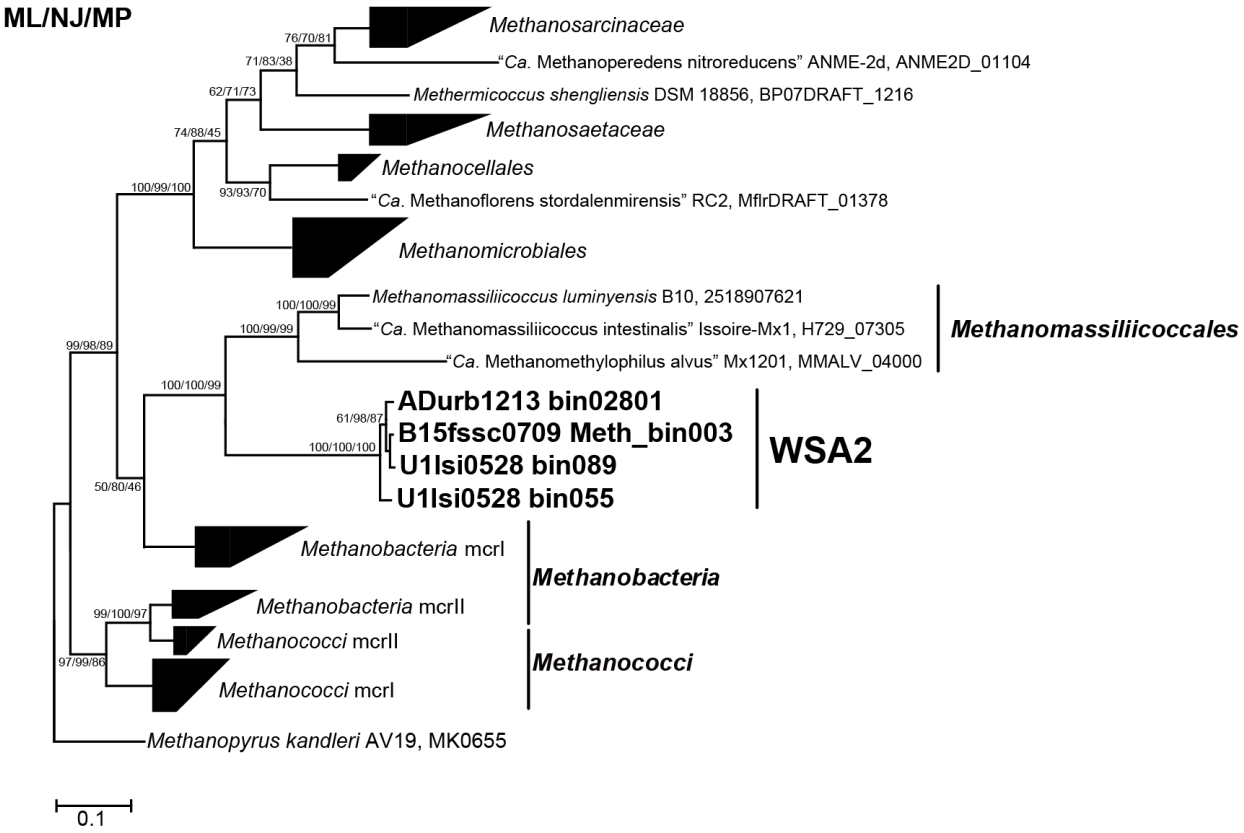

Figure C.1. WSA2 phylogeny based on methyl coenzyme M reductase subunit A. The tree was calculated using maximum likelihood (ML), neighbor-joining (NJ), and maximum parsimony (MP). The bootstrap values for all methods are shown. The bar indicates 10% base substitution.

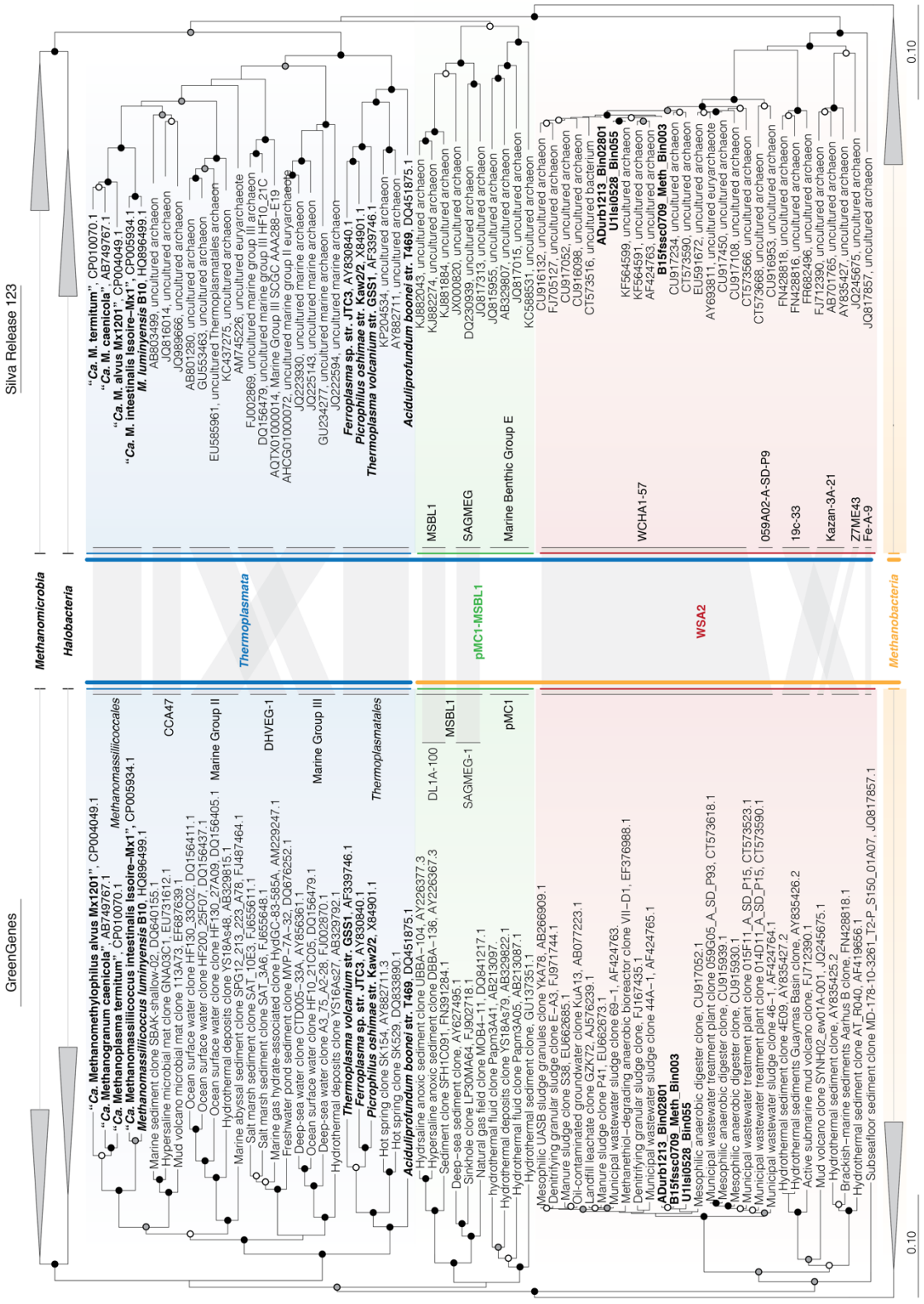

Figure C.2. (see caption on next page)

Figure C.2. (cont.) 16S rRNA-based phylogeny of WSA2 (red) compared to closely related *Euryarchaeota* classes and clusters (pMC1 and MSBL1) using both Greengenes (left) and Silva v123 (right) databases. The trees were constructed using ARB neighbor-joining algorithm with 5000 bootstrap replications, sequences at least 1200bp in length, and *Methanobacteria* as the outgroup. Bootstrap values greater than 90% (black), 75% (gray), and 50% (white) are shown. Clades corresponding to each other in the two databases are indicated. The current classification of each clade to *Thermoplasmata* (blue) and *Methanobacteria* (orange) classes are shown at the center with thick vertical bars. The renewed classification is depicted as thin adjacent lines, categorizing the sequences as *Thermoplasmata*, pMC1-MSBL1, WSA2, or *Methanobacteria*.

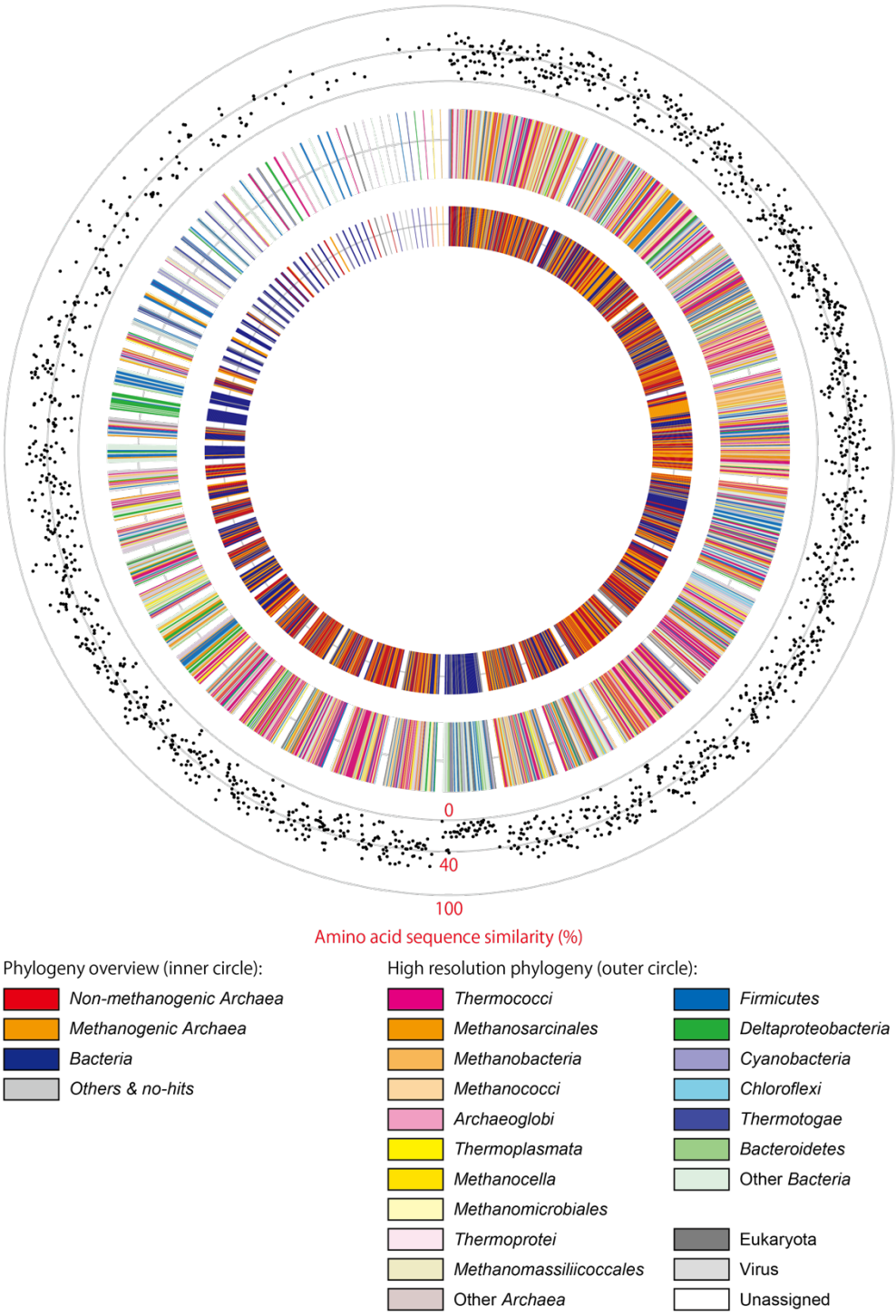

Figure C.3. The phylogeny of genes in the anaerobic digester WSA2 genome based on blastp against the current GenBank non-redundant protein sequence database. Contigs are shown in order of decreasing length from the top in the clock-wise direction. Genes are colored individually based on their closest phylogenetic relative using warmer and cooler colors for *Archaea* and *Bacteria* respectively at low- (domain; inner circle) and high- (ranging between phylum and order) resolution phylogenetic classification. The outer circle shows blastp similarity of the corresponding WSA2 gene to their top hit.

Figure C.4. Phylogeny of methylated thiol Coenzyme M methyltransferase (Mts) against associated corrinoid proteins and methanol- (MtaC) and methylamine- (MtmC, MtbC, and MttC) specific methyltransferase corrinoid proteins. The tree was constructed using Clustalw with default parameters for alignment and neighbor-joining clustering method for tree construction with 5000 bootstrapped replications. For Mts families associated with *Methanosarcinales* (blue), *Methanomassiliicoccales* (orange), and WSA2 (red), lines are drawn connecting neighboring methyltransferases and corrinoid proteins. For Mts families that have discernable association with specific corrinoid proteins, backgrounds are provided behind these lines. Nodes are marked for bootstrap values greater than 90% (black), 75% (gray with outline), and 50% (gray).

C.2 – Supplementary Tables

Refer to supplemental file "Nobu2017_AppendixC_Tables.xlsx" for the following tables:

Table C.1. Genome characteristics of WSA2 extracted from metagenomes of an anaerobic digester (AD) in Urbana, Illinois and full-scale and lab-scale purified terephthalic acid (PTA) process wastewater-treating methanogenic bioreactors. Genomes were identified for four WSA2 populations and one representative genome was selected for each WSA2 population (bolded). Bins (or strains) from the same population in different samples (i.e., time points for Urbana AD and reactor depths for Full-scale PTA) (marked with *) are shown with estimated genome completeness calculated using 40 marker genes. (abbreviations: CDS - coding sequence).

Table C.2. Methanogenesis-related catabolic genes of the WSA2 genomes, including core genes (black), genes found in only some genomes (blue), and genes missing from all genomes.

Table C.3. Biosynthesis pathways of the WSA2 genomes, including core genes (black), genes found in only some genomes (blue), and genes missing from all genomes.

Table C.4. Methanogenesis cofactor biosynthesis of the WSA2 genomes, including core genes (black), genes found in only some genomes (blue), and genes missing from all genomes.

APPENDIX D

SUPPLEMENTARY MATERIAL TO CHAPTER 5

D.1 – Supplementary Information

D.1.1 – Metagenomics and phylogenetic binning

The metagenomic reads from HiSeq2000 were trimmed with a minimum quality cutoff of 3 and further screened to be at least 50bp in length, have an average quality score >20, and contain less than 3 ambiguous nucleotides (N's). The trimmed and screened paired-end reads were assembled using SOAPdenovo v1.05 (*44*) with default settings ("-K 81 –p 32 –R –d 1") and a range of kmers between 85 and 105. Short contigs (<1800bp) were assembled using Newbler (flags: -tr, -rip, -me 98, -ml 80) (Roche) and the longer contigs and Newbler-assembled sequences were further assembled using minimus 2 (flags: -D MINID=98 –D OVERLAP=80) (*299*). Raw reads were mapped to the final contigs using BWA to estimate individual read depths (*156*). In parallel, to elucidate the microbial community composition, 16S rRNA genes were assembled from trimmed reads using EMIRGE with default settings (*300*), curated using ChimeraSlayer included in QIIME (*209*), aligned with the PyNAST algorithm against the 2010 Greengenes core dataset (*212, 301*), classified using the Greengenes server, and further manually classified in the Greengenes dataset phylogenetic tree using ARB (*213*). Genes were annotated using the Joint Genome Institute Integrated Microbial Genomes (IMG) pipeline (*282*) and manually curated using EBI-InterproScan (*100*), BLAST+ blastp (*215*), and Phobius (*302*). The metagenome and SAG data are correspondingly accessible on IMG/M with taxon object ID 3300001095 and Genbank with accession numbers noted in Table D.2).

For phylogenetic binning, PhylopythiaS, ClaMS, and Metawatt were tested. Default settings were used for PhylopythiaS. For ClaMS, the de-Bruijn chain algorithm was used with a k-mer length of 4 and a 0.005 distance cut-off. As for Metawatt, a minimum bin size of 100kb was used. For all k-mer-based binning, contigs shorter than 500bp were omitted due to limitations of the method. The Metawatt BLAST annotation was performed by cleaving metagenomic contigs into 500bp fragments and performing blastn with an e-value cut-off of 10^{-3} against reference genomes from the Genbank genome database and TA-degrading community single amplified genomes (SAGs) (Table D.2).

For rigorous scaffold annotation and draft genome construction, four additional contig-discrimination methods were used to manually curate PhylopythiaS-generated bins: average contig read coverage (*99*), BLAST, k-mer frequency (*51*), and essential single-copy genes (*303*). For each taxonomic bin, clusters of contigs with homogeneous read coverage, blastp-based phylogenetic distribution of genes, and k-mer frequency were identified and the remaining contigs were rejected from the bin. Taxonomic identity

of each bin was manually re-evaluated and corrected based on gene phylogenetic distribution and 16S rRNA gene phylogenetic classification determined using ARB with GreenGenes database (*212, 213*). Within each bin, we also rejected contigs with consistently high similarity to other clades found in the TA-degrading community (>60% to Genbank-derived reference genome), overlap with contigs from SAGs of other clades (>98%), inconsistent k-mer frequency based on IMG k-mer frequency principle component analysis using a fragment window of 5000bp, fragment step of 500bp, k-mer size of 4, and minimum variation of 10. Contigs from Metawatt bins of the same clade and contigs rejected from other PhylopythiaS bins were manually binned to the target clade only if they met the above criteria. Finally, redundant essential single-copy genes were identified to set a stringent read coverage cut-off and eliminate binning overlap between populations of the same clade. Bins and SAGs that correspond to the same taxon (16S rRNA gene sequence similarity >99%) and/or share contig overlap (>98% nucleotide similarity) were grouped together and treated as a species-level pangenome. This procedure allows generation of high quality draft genomes based on homogeneous k-mer frequency, no repeats in essential single-copy genes, and rejection of high-homology essential single-copy genes on contigs with relatively low or high read coverage. Genome completeness for bacterial clades was estimated using bacterial essential single-copy genes (*303*). Similarly, for archaeal clades, we estimated genome completeness by identifying single-copy genes (COG-based) shared amongst all publically available genomes within the closest related order (*Methanosarcinales* or *Methanomicrobiales*) and calculating the coverage of these genes in the metagenome- or SAG- derived draft genome. The relative abundance of these microbial populations in the community was estimated by grouping (97% similarity cutoff) the 16S rRNA gene sequence derived from the draft (pan)genomes with EMIRGE-constructed sequences and their abundance information.

D.1.2 – EMIRGE-based abundance of each metagenomic bin

To estimate the abundance of each population corresponding to the constructed metagenomic draft genomes, we matched bin-associated 16S rRNA gene sequences with EMIRGE-assembled sequences, which have calculated population abundances. Genomes of known clades (EMIRGE-based percent abundance noted if available) included three *Methanolinea* (clone HMTAb4, 31.3% of archaeal community), four *Methanosaeta*, two *Methanospirillum* (HMTAb46, 6.3%), three *Pelotomaculum* (HMTAb16, 5.5%; HMTAb50, 3.2%; and HMTAb322, 2.6% of bacterial community), three *Syntrophorhabdaceae*, *Pseudothermotoga* (HMTAb26, 0.6%), *Mesotoga* (HMTAb101, 3.7%), and *Aminiphilus* (HMTAb171, 6.0%) (Fig. 2 and Supplementary Table D.1 and Table D.2). For target uncultivated phyla, SAG-sequencing and binning generated genomes for Atribacteria (HMTAb134, 0.8%), Cloacimonetes (HMTAb147, 1.7%), Hydrogenedentes (HMTAb329, 0.8%), Marinimicrobia (HMTAb93, 17.2%), and WS1 (HMTAb109, 0.2%);

poorly understood phyla *Armatimonadetes* and *Chloroflexi* (HMTAb180, 0.7%); and other uncultivated phylogenetic branches including the *Caldiserica* 31q06 previously described (HMTAb66, 0.8%), novel *Caldisericia* sister clade (HMTAb62, 1.1%), *Syntrophus*-related clade (HMTAb213, 1.2%), *Spirochaetes* SA-8 (HMTAb43, 1.7%), *Synergistetes* TTA-B6 (HMTAb273, 0.59%), *Planctomycetales*-related clade (HMTAb164, 3.4%), and *Phycisphaerae*-related cluster WPS-1 (HMTAb11, 12.2%) (*18, 231, 232*). Of these clades, pangenomes were constructed for Atribacteria, Cloacimonetes, Hydrogenedentes, Marinimicrobia, *Chloroflexi*, and *Syntrophus*-related clade. Several methanogen, *Syntrophorhabdaceae*, and *Spirochaetes* SA-8 clade draft genomes lacked 16S rRNA gene sequences or only had truncated sequences (<600 bp).

D.1.3 – Syntrophic metabolizer energy conservation pathway survey

A survey across publically available genomes of organisms capable of syntrophic metabolism confirms that syntrophs generally divide in to two groups, encoding either Rnf (NADH:Fd oxidoreductase)(*12*) or Hdr-Ifo (heterodisulfide-reductase-associated putative ion-translocating Fd:NADH oxidoreductase)(*228*). These complexes are thought to perform proton-motive-force-driven reverse electron transport from NADH to Fd_{ox}. This generates Fd_{red}, a valuable high-energy electron carrier, that can facilitate exergonic H^+ reduction to H_2. Syntrophs harness this energy margin to drive concomitant endergonic NADH-oxidizing H^+ reduction using a Fd_{red}-dependent electron-confurcating hydrogenase (ECHyd; HydABC)(*35, 120*). While Rnf and ECHyd are known to support anaerobic H_2 metabolism other than syntrophic degradation, the Hdr-Ifo gene cassette is well-conserved amongst organisms capable of syntrophic metabolism. Furthermore, as Hdr-Ifo co-occurs with ECHyd in all available syntroph genomes, we believe the genomic co-occurrence of these two systems with catabolic pathways for oxidation of a syntrophic substrate(*133*) serves as a rigorous biomarker for the capacity to perform Hdr-Ifo-type syntrophic catabolism. To further support this, it is necessary to identify that no fermentative or other respiratory pathways are expressed.

Previous studies have postulated that Fix (electron-transfer-flavoprotein-oxidizing hydrogenase) may play an important role in RET-driven H_2 generation during syntrophic butyrate oxidation(*35, 234*). In agreement, many syntrophs capable of metabolizing simple carboxylates (*e.g.*, propionate, butyrate, and glycolate) possess the FixABCX gene cassette(*12*). While many non-syntrophic organisms encode the Fix system, the genomic association between Fix and carboxylate metabolism may guide identification of carboxylate-oxidizing syntrophic organisms. For example, even though Fix may support electron transduction from alkyl dehydrogenation to H^+ reduction, another electron flow path is necessary to facilitate NADH reoxidation. Therefore, if a carboxylate-metabolizing syntroph utilizes Fix for RET-driven H_2 production, it is likely to also possess ECHyd complemented with RET to facilitate Fd-mediated NADH-

oxidizing H^+ respiration. Following this logic, all currently available genomes of FixABCX-encoding syntrophs also harbor ECHyd with Rnf or Hdr-Ifo. Thus, similar to Hdr-Ifo, syntrophic carboxylate degraders may be genomically characterized through coinciding identification of carboxylate metabolism, Fix, ECHyd, and Rnf or Hdr-Ifo.

On the other hand, we also identify organisms capable of syntrophic metabolism lacking ECHyd (*e.g.*, *Clostridium ultunense*) or Rnf and Ifo (*e.g.*, *Thermacetogenium phaeum*). Coincidentally, these organisms are both capable of syntrophic acetate oxidation. They may generate NADPH from acetate-degrading Wood-Ljungdahl pathway methyl-tetrahydrofolate (THF) oxidation and may require energy conservation mechanisms connecting NADPH to H_2/formate generation, different from other syntrophs who do not rely on $NADP^+$ as a catabolic electron carrier. *C. ultunense* encodes Rnf, an electron-bifurcating NAD^+-dependent NADPH:Fd_{ox} oxidoreductase (NfnAB), and energy-conserving formate dehydrogenase (ECFdh; HylABC-FdhF2) structurally and functionally similar to ECHyd(*249*) (Table D.12). Although only electron bifurcation activity has been tested for ECFdh, we suspect that its activity is reversible much like ECHyd(*120, 123*). Thus, we postulate that NfnAB drives electron transfer from NADPH to NAD^+ and Fd_{ox}, Rnf balances NADH and Fd_{red} ratio to 1:1, and ECFdh performs electron confurcation from NADH and Fd_{red} (1:1) to CO_2 reduction, whereby completing electron flow from substrate oxidation to formate generation. Interestingly, *C. ultunense* lacks any known enzymes for further H^+-reducing formate oxidation.

While *T. phaeum* lacks Rnf, Hdr-Ifo, and NfnAB, we identify ECHyd, NADPH-dependent formate dehydrogenase (FdhAB), and several RET mechanisms, including NAD(P) transhydrogenase (Pnt), group 4 Fd-dependent hydrogenase (Hyf), and Hyf-linked formate dehydrogenase (FdhH). Besides electron confurcation through ECHyd, these complexes theoretically allow oxidation of NADH, NADPH, and Fd_{red} coupled with Pnt, FdhAB, and Hyf respectively. Although FdhAB would generate formate as an intermediate, Hyf-linked FdhH can perform proton-extruding formate oxidation and H^+ reduction. In this way, these enzymes form a complete electron flow path re-oxidizing NADH, NADPH, and Fd_{red} through H_2 production. Albeit uncommon amongst syntrophs with sequenced genomes, organisms capable of syntrophy can clearly deviate from the predominant Rnf- and Hdr-Ifo- type catabolism and form unique energy-conserving electron flow independent of Rnf, Hdr-Ifo, or ECHyd. Regardless, we observe that electron bi(con)furcation and RET is essential for syntrophic metabolism and that certain energy conservation strategies can correlate with particular catabolic capabilities.

D.1.4 – Homoacetogen energy conservation pathway survey

In order to improve our understanding of energy conservation revolving around homoacetogenic metabolism, we compared the genomes of known homoacetogenic bacteria, *Acetobacterium woodii(304,*

305), *Acetonema longum(306)*, *Blautia product* (*307, 308*), *Blautia schinkii (309)*, *Clostridium autoethanogenum(310)*, *C. carboxidivorans (311)*, *C. drakei (311)*, *C. ultunense (312)*, *C. ljungahlii (313, 314)*, *Eubacterium limosum*, *Moorella thermoacetica (28, 308)*, *Sporomusa ovata* (*315, 316*), *Tepidanaerobacter aectatoxydans (317, 318)*, *Thermacetogenium phaeum* (*319*), and *Thermoanaerobacter kivui* (*320*), *Treponema azotonutricium (321)*, and *Treponema primitia (321)*. In chemolithotrophic homoacetogenesis mediated by the Wood-Ljungdahl pathway, acetogens are thought to employ unique energy conservation strategies to harness energy from exergonic methylene-tetrahydrofolate (THF) reduction ($\Delta G^{o'}$=-22kJ/mol) (*304, 322*), drive endergonic H_2 and formate oxidation, and facilitate energy acquisition from high energy electron carriers (*i.e.*, Fd_{red} and NADPH). Comparative genomics reveals that acetogens divide into five distinct methylene-THF reductase (MetF) genotypes (Table D.12). We identify that *A. woodii*, *B. hydrogenotrophica*, and *B. schinkii* encode the previously postulated bifunctional electron-bifurcating MetVF-RnfC2 (*304*) (type I) and associated ion-translocating NADH:Fd_{ox} oxidoreductase (Rnf). While *C. ljungdahlii(313)*, *T. acetatoxydans*, *C. carboxidivorans*, *C. autoethanogenum, E. limosum*, and *T. kivui* possess MetVF as well, they lack RnfC2 and *T. kivui* is missing Rnf; thus, this MetVF (type II) is likely restricted to Rnf-independent electron bifurcation. *M. thermoacetica*, *T. phaeum*, and *Sporomusa ovata* encode a heterodisulfide reductase (HdrABC)-associated MetF thought to perform unknown energy-conserving reduction (*323*). In contrast some acetogenic MetF are non-energy-conserving: *A. longum* and *S. ovata* encode a Fd-oxidizing MetF while *B. producta(324)* and *T. primitia* have an NADH-dependent MetF. Interestingly, some of these MetF only share <20% amino acid similarity; furthermore, *T. azotonutricium*, *C. drakei*, and *C. ultunense* lack any identifiable MetF. Thus, homoacetogens harbor functionally and phylogenetically diverse MetF despite most belonging to a single phylum, *Firmicutes*, and must possess unidentified MetF with novel functionality.

While we find that hydrogenotrophic acetogens consistently encode energy-conserving H_2 oxidation through an NAD^+- and Fd_{ox}- dependent electron-bifurcating hydrogenase (HydABC)(*123*), we also discover that acetogens encode a variety of other cytoplasmic, membrane-bound, and periplasmic hydrogenases and formate dehydrogenases. The hydrogenases include energy-conserving Fd_{ox}-reducing group 4 hydrogenases (*325*) (Hyf; encoded by *M. thermoacetica* and *T. phaeum*), novel Hyf-like hydrogenase cassette (Cluster A; *S. ovata*, *A. longum*, and *C. carboxidivorans*), NADP-reducing electron-bifurcating hydrogenase (*326*) (Hyt; *C. ljungdahlii*, *C. carboxidivorans*, and *C. autoethanogenum*), cytoplasmic $NADP^+$-reducing hydrogenase (*M. thermoacetica*), signaling-kinase-associated putative sensory hydrogenase (*239, 327, 328*) (HfsABC or HydS; *A. woodii*, *B. hydrogenotrophica*, *B. schinkii*, *A. longum*, *T. acetatoxydans*, *C. carboxidivorans*, *E. limosum*, *B. producta*, *T. primitia* str. ZAS-2, and *T. azotonutricium*) and a novel conserved hydrogenase gene cassette (*B. schinkii*, *B. producta*, and *C.*

ultunense) (Table D.12). As for formate metabolism, we identify quinone-reducing (FdoGHI; *M. thermoacetica*), cytoplasmic $NADP^+$-reducing (FdhAB; *M. thermoacetica*, *T. phaeum*, and *S. ovata*), and two electron-bifurcating(*249*) (HylABC-FdhN; *A. longum*, *B. producta*, and *C. ultunense* & Hyt-linked FdhA; *C. ljungdahlii*) formate dehydrogenases. In addition, several acetogens encode H_2-oxidizing CO_2 reduction complexes, including an *A. woodii* gene cassette (cluster I; *A. woodii*, *B. schinkii*, *T. kivui*, and *T. primitia*) and Hyf-linked formate dehydrogenase H (FdhH; *Moorella* and *Thermacetogenium*). Thus, as with MetF, homoacetogens hydrogenases and formate dehydrogenases are remarkably diverse. Moreover, the presence of these substrate oxidation pathways does not correlate with the observed MetF genotypes, suggesting that further investigation is necessary to identify the reason for such high diversity within this semi-monophyletic niche. In addition, we find that Rnf, NfnAB, HydABC, and HfsABC are well conserved amongst many aceotgens, suggesting that they may serve a core function to homoacetogenesis. On the other hand, we currently have limited genomic insight into non-hydrogenotrophic formate-oxidizing acetogens (*e.g.*, *C. ultunense*), so we could not determine genomic trends for such organisms.

H_2 and formate oxidation by enzymes discussed above generate a variety of reduced electron carriers. However, specific steps in the Wood-Ljungdahl pathway require particular electron donors. In most cases, methenyl-THF(*329*) and CO(*322, 330*) reduction respectively require NAD(P)H and Fd_{red} as electron donors. In addition, most methylene-THF reduction mechanisms and other physiological processes rely on NADH for reducing power. Logically, all homoacetogen genomes must harbor protein complexes for electron transfer between low- (*e.g.*, Fd and NADP) and high- (*e.g.*, NADH) potential electron carriers to facilitate electron flow from substrate oxidation to reductive acetyl-CoA synthesis. In agreement, we discover that H_2- and formate- oxidizing homoacetogens all encode at least one of the following energy-conserving electron transfer complexes: Rnf, NAD(P) transhydrogenase, and NADPH:Fd-NAD^+ oxidoreductase (Nfn)(*236*). In conclusion, while there is seemingly little genotypic convergence between homoacetogens, we identify that MetF is not necessarily well-conserved; uncharacterized acetogenic MetF must exist; acetogens depend on energy-conserving electron transfer between physiological electron carriers (*e.g.*, Nfn); and that specific energy conservation pathways (*i.e.*, RnfA-G, NfnAB, HydABC, and HfsABC) are shared widely amongst homoacetogens. Thus, this holistic homoacetogen genome comparison builds upon previous acetogen genome analyses(*28, 304, 310, 313, 316, 331, 332*) to reveal insight that could facilitate genomic identification of novel acetogens and also support future acetogen metabolism analysis.

D.1.5 – *Mesotoga* and *Pseudothermotoga* syntrophic acetate degradation

Although *Pseudothermotoga lettingae* strain TMO is a known acetate-degrading syntroph, its genome does not encode the Wood Ljungdahl pathway (CO dehydrogenase/acetyl-CoA synthase + tetrahydrofolate pathway) thought to oxidize acetate by performing reverse acetogenesis(*333*). Thus, *Thermotogae* syntrophic acetate degradation must utilize a different pathway. As *P. lettingae* strain TMO is known to degrade acetate to CO_2, it may use the tetrahydrofolate (THF) pathway as a means of connecting acetate degradation to one-carbon metabolism. Supporting that *P. lettingae* strain TMO may use the THF pathway for syntrophic catabolism, strain TMO is also known to syntrophically degrade methanol, a one-carbon compound whose degradation presumably requires the THF pathway(*240*). Moreover, the strain TMO MetF is distinct (<20% amino acid similarity) from those encoded by acetate-degrading syntrophs possessing the Wood-Ljungdahl pathway, which supports the existence of an alternative acetate oxidation strategy. Based on analysis of the strain TMO genome and TA-degrading community *Mesotoga*/*Pseudothermotoga* draft genomes, we identify that these organisms encode genes for connecting acetate degradation to the THF pathway through the glycine cleavage system (Fig. D.2 and Table D.5). In this pathway, acetyl-CoA is condensed with formate (generated at the end of the pathway) to produce pyruvate, and then aminated to form serine. Using THF, serine is split into glycine and 5,10-methylene-THF. The glycine cleavage system oxidatively degrades glycine to generate another 5,10-methylene-THF. The THF pathway can then further oxidize these 5,10-methylene-THF to yield formate. Finally, this formate can be oxidized to generate H_2 and CO_2 using a formate hydrogenlyase (Fhl) complex(*334*). While the TA-degrading community *Thermotogae* draft genomes encode this Fhl complex, strain TMO does not possess any discernable formate dehydrogenase or Fhl to perform the final formate degradation step. However, given that strain TMO is known to degrade formate, it must have an unidentified formate oxidation mechanism(*240*).

As discussed with regard to homoacetogens, methenyl-THF reduction (or methylene-THF oxidation for acetate catabolism) is thought to require NADP(H) as an electron carrier. NADPH re-oxidation is paramount for these *Thermotogae* organisms to utilize the THF pathway for acetate oxidation. While neither *Mesotoga* nor *Pseudothermotoga* encode NADPH-dependent hydrogenases, we identify that they encode NfnAB for electon bifurcation of NADPH to Fd_{ox} and NAD^+. In total, the postulated acetate oxidation pathway generates 1 Fd_{red}, 2 NADH, 1 ATP, 4 H_2, and 2 CO_2 per acetate. To further oxidize Fd_{red} and NADH, all three *Thermotogae* genomes encode Rnf and ECHyd. Using these complexes to perform energy-conserving electron transfer from Fd_{red} and NADH to H_2 generation would require approximately 2/3 ATP investment per acetate, assuming 3.33 H^+ extruded per ATP via ATPase, 1 H^+ translocated per electron transferred from NADH to Fd_{red}, and 2 H^+ extruded through formate-oxidizing H_2 generation (*e.g.*,

formate hydrogen lyase or Hyf-FdhN). Thus, this pathway has an energy yield of 1 ATP per acetate after thermodynamically sound, complete electron balance.

In comparison with the oxidative Wood-Ljungdahl pathway, this novel pathway circumvents carbon monoxide metabolism, avoids endergonic 5-methyl-THF oxidation, and generates organic metabolic intermediates that can be easily used for biosynthesis, and conserves energy by using formate (an end-product) to drive pyruvate synthesis (pyruvate-formate lyase) (Fig. D.2). This pyruvate synthesis from acetyl-CoA would otherwise require Fd_{red}, a valuable high-energy electron carrier. Further, in comparison to the *Thermotogae*-type pathway with a yield of 1 ATP per acetate, the current understanding of the Wood-Ljungdahl pathway suggests a puzzling theoretical yield of 0 ATP per acetate assuming (a) 3.33 H^+ are translocated by ATPase ATP hydrolysis, (b) methylene-THF oxidation generates NADPH, (c) endergonic 5-methyl-THF oxidation proceeds through Fd_{red}-dependent NADH-yielding electron confurcation, (d) Nfn and Rnf mediate electron transfer between carriers, and (e) energy-conserving formate-oxidizing H_2 generation extrudes 2 H^+ across the membrane (*e.g.*, formate hydrogen lyase or Hyf-FdhN). While we clearly lack critical insight into the energetics of syntrophic acetate oxidation through the Wood-Ljungdahl pathway, this analysis suggests that the *Thermotogae*-type and Wood-Ljungdahl acetate catabolism may have different ATP yields. Furthermore, this metabolic reconstruction warrants future metabolic and physiological comparison of *Pseudothermotoga* with other acetate- and C1- metabolizers to elucidate the nature of syntrophic C1 metabolism under methanogenic conditions.

D.1.6 – *Syntrophus*-related clade member branched fatty acid degradation

Branched-chain fatty acids (BCFA) are often overlooked when testing anaerobic metabolic capabilities; thus, we have very little insight into such pathways, let alone specific genes. Nevertheless, previous studies presented valuable observations for syntrophic isovalerate (IV), isobutyrate (IB), and 2-methylbutyrate (2MB) degradation(*242, 335-338*). These BCFA have been observed to be correspondingly degraded to 3 mol acetate + 1 mol H_2, 2 mol acetate + 2 mol H_2, and 1 mol acetate + 1 mol propionate + 2 mol H_2. Previous studies also present evidence for specific steps in BCFA degradation. For IB degradation, a previous study hypothesized that a unique "isobutyryl-CoA mutase" isomerizes isobutyryl-CoA to butyryl-CoA as one of the first steps (Fig. D.5) (*336*). For IV, another study hypothesized that an ion-translocating 3-methyl-but-2-enoyl-CoA carboxylase, enoyl-CoA hydratase, and 3-hydroxy-3-methylglutaryl-CoA lyase are necessary (Fig. D.5) (*242*). For 2MB, a study on *Pseudomonas putida* isolates a 3-hydroxy-2-methylbutyryl-CoA dehydrogenase specific to 2-methylbutyrate degradation, albeit not syntrophic(*338*). Based on such observations, we propose metabolic pathways for degrading these BCFA (Fig. D.5).

The *Syntrophus*-related clade member expresses candidate genes that may perform metabolic steps specifically necessary for BCFA catabolism. For IB, this *Syntrophus*-related clade member expresses an acyl-CoA mutase complex (100112619-100112620) that has high homology (>80%) to that of *Syntrophothermus lipocalidus* (Slip_0758-0759), the only known IB-degrading syntroph with a sequenced genome (Table D.3)(*339*). For IV, the *Syntrophus*-related clade member expresses a gene cassette (100003690-100003694) with putative ion-translocating 3-methyl-but-2-enoyl-CoA carboxylase, enoyl-CoA hydratase, and 3-hydroxy-3-methylglutaryl-CoA lyase highly homologous (52-75%) to a similar gene cassette found in *Desulfoccocus multivorans,* a known anaerobic IV degrader (WP_020878000-WP_020878002) (Table D.3). For 2MB, this *Syntrophus*-related clade member expresses a 3-hydroxyacyl-CoA dehydrogenase (100044538) homologous (53%) to the *Pseudomonas putida* 3-hydroxy-2-methylbutyryl-CoA-specific dehydrogenase (WP_020192115) necessary oxidation of 3-hydroxy-2-methylbutyryl-CoA to 3-keto-2-methylbutyryl-CoA (Fig. D.5). As a result, we identify putative BCFA-degrading genes in known BCFA degrader genomes and find that the *Syntrophus*-related clade member expresses genes with high homology.

D.1.7 – *Caldiserica* cluster 31q06 unlikely to be an acetogen

Our previous study proposed that *Caldiserica* cluster 31q06 member may perform H_2-oxidizing acetate and butyrate production mediated by ribulose-1,5-bisphosphate carboxylase (RuBisCO) and butyryl-CoA dehydrogenase(*87*). This RuBisCO was hypothesized to perform carbon fixation through the Calvin-Benson-Bassham cycle. We also identified this RuBisCO gene (100196811) in the *Caldiserica* draft genome. Despite possessing this gene, this genome lacks most of the genes encoding the CBB cycle. ClustalW-based reevaluation of the RuBisCO gene revealed that it falls into the type IV RuBisCO-like proteins that lacks the carboxylation active site residues and instead catalyzes enolization of a ribulose-1,5-bisphosphate analogue, 2,3-diketo-5-methylthiopentyl-1-P(*340*). In type IV RLP-encoding organisms, including *Bacillus subtilis* and *Rhodospirillum rubrum*, the RLP enolase was found to participate in methionine salvage(*341, 342*). In this pathway, a byproduct of polyamine biosynthesis, S-Methyl-5'-thioadenosine (MTA), is converted to methionine(*341*). In addition, the butyryl-CoA dehydrogenase does not relate to the clostridial electron-bifurcating butyryl-CoA dehydrogenase and cannot perform energy-conserving butyrate production. Therefore, we determine that 31q06 can neither oxidize H_2 acetogenically nor generate butyrate, but low genome completeness prohibits further accurate identification of metabolic capabilities it may have.

D.1.8 – Novel electron transfer complexes found in Hydrogenedentes

We identify a unique Hydrogenedentes gene (100000978) that appears to be a fusion protein between a methanogenesis-associated c-type cytochrome and hydrogenase-associated cytochrome b. The N-terminal cytochrome c has a transmembrane anchor (7-26) followed by seven periplasm-side CXXCH heme motifs (112-116, 186-190, 238-242, 310-314, 342-346, 375-379, and 404-408) similar to a *Methanosarcina acetivorans* strain C2A methanogenesis-related multiheme c-type cytochrome. The C-terminal cytochrome b contains five consecutive transmembrane domains (917-936, 957-979, 994-1016, 1049-1070, and 1090-1112) thought to bind two hemes between the helices and also one CXXCH heme motif (861-865) on a periplasmic strand. Based on this structure, it is reasonable to infer that the cytochrome b allows quinol oxidation and donates electrons to the c-type cytochrome, or *vice versa*.

In addition, the Hydrogenedentes genome also encodes a periplasmic Fe-only hydrogenase (10001196) that may receive electrons from the c-type cytochromes for H_2 generation, analogous to the well-studied *Desulfovibrio vulgaris* H_2-producting electron transduction system(*343*). Although this hydrogenase does not contain a signal peptide, the gene directly upsteam is an uncharacterized periplasmic electron transfer protein encoding a signal peptide with thioredoxin, alpha-helical ferredoxin, and FAD-binding domain. Moreover, the upstream ferredoxin- and thioredoxin- like genes are also both periplasmic. Thus, these non-cytoplasmic redox genes may be co-translocated into the periplasm using the thioredoxin signal peptide (*344*). In combination with the expressed Na^+-translocating NADH:quinone oxidoreductase (Nqr), the TA-degrading community Hydrogenedentes may perform unique membrane-associated NADH-oxidizing H_2 production through Nqr NADH-oxidizing quinone reduction, cytochrome b quinol-oxidizing cytochrome c reduction, and periplasmic hydrogenase for final cytochrome-c-oxidizing H^+ reduction.

We also identify another unique hydrogenase (100000927) containing an N-terminal Fe-hydrogenase domain and C-terminal signal transduction histidine kinase. We also identify 4Fe-4S and ferredoxin-type 4Fe-4S domains immediately upstream and downstream of the Fe-hydrogenase domain, respectively. In addition, this hydrogenase is coordinated with another signal transduction histidine kinase. Thus, we speculate that these genes may encode a sensory hydrogenase and signaling system, similar to the clostridial hydrogenase with PAS-domain associated with many acetogens.

D.1.9 – Candidatus nomenclature

Mesotoga member: "*Ca.* Mesotoga acetoxidans"
acet.o'xi.dans. L. n. *acetum*, vinegar; N.L. v. *oxido* (from Gr. adj. *oxys*, acid or sour and in combined words indicating oxygen), to oxidize; N.L. part. adj. *acetoxidans*, acetate-oxidizing.

Pseudothermotoga member: "*Ca.* Pseudothermotoga acetoxidans"
acet.o'xi.dans. L. n. *acetum*, vinegar; N.L. v. *oxido* (from Gr. adj. *oxys*, acid or sour and in combined words indicating oxygen), to oxidize; N.L. part. adj. *acetoxidans*, acetate-oxidizing.

Syntrophus-related clade member: "*Ca.* Caldisyntrophus multiacidovorans"
Cal.di.syn.tro'phus. L. adj. *caldus*, hot; Gr. adj. *syntrophos*, having grown up with one, living with; N.L. masc. n. *Caldisyntrophus*, one living syntrophically under thermophilic conditions with another so that each produces a nutrient required by the other.
mul.ti.a.ci.do.vo'rans. L. adj. multus, many; N.L. neut. n. acidum (from L. adj. acidus -a -um, sour), an acid; L. v. voro, to eat, to devour; N.L. part. adj. multiacidovorans, devouring many acids.

Chloroflexi subphylum I member: "*Ca.* Anaeroacetigenium thermophilum"
An.a.e.ro.a.ce.ti.ge'ni.um. Gr. pref. *an*, not; Gr. n. *aer aeros*, air; N.L. n. *acidum aceticum*, acetic acid; L. suff. - *genium* (from L. v. *gigno*) producing; N.L. neutr. n. *Anaeroacetigenium*, producing acetic acid anaerobically.
ther.mo'phi.lum. Gr. n. *thermê*, heat; N.L. adj. *philus* (from Gr. masc. adj. *philos*), loving; N.L. neutr. adj. *thermophilum*, heat-loving.

"*Ca.* Hydrogenedentes" (NKB19) member:"*Ca.* Hydrogenigenium lipolyticum"
Hy.dro.ge.ni.ge'ni.um. N.L. n. *hydrogenium*, hydrogen; L. suff. - genium (from L. v. gigno) to produce; N.L. n. *Hydrogenigenium* producing hydrogen.
li.po.ly'ti.cum. Gr. n. *lipos*, fat; Gr. v. *lyo*, to dissolve; Gr. suff. –*ikos*, belonging to; N.L. neutr. adj. *lipolyticum*, dissolving fat, referring to the ability to hydrolyze lipids to glycerol and fatty acids.

Planctomycetales member: "*Ca.* Lipofermentans glycerini"
Li.po.fer.men'tans. Gr. n. lipos, fat; L. part. adj. *fermentans*, fermenting; N.L. masc. adj. *lipofermentans*, fat fermenting.
gly.ce.ri'ni. N.L. n. *glycerinum*, glycerine, glycerol; N.L. gen. n. *glycerini*, of glycerol, referring to utilization of glycerol.

"*Ca.* Marinimicrobia" (SAR406) member: "*Ca.* Anteamarinum acidaminivorans"
An.te.a.ma.ri'num. L. adv. *antea*, before; L. adj. *marinus*, marine; N.L. neutr. n. *Anteamarinum*, a microbe that was formerly in the marine environment.
a.cid.a.mi.ni.vo'rans. L. n. *acidum*, an acid; N.L. n. *aminum*, an amine; -i- connecting vowel; L. part. adj. vorans, eating. N.L. adj. *acidaminivorans*, amino acid eating.

Phycisphaerae-related cluster WPS-1 member: "*Ca.* Proteinisyntropha apolusis"
Pro.tei.ni.syn.tro'pha. N.L. n. *prǫteĭnum*, protein; -i- connecting vowel; Gr. adj. *syntrophos*, having grown up with one, living with; N.L. fem. n. *Proteinisyntropha*, one who degrades protein in syntrophic association with partner organisms.
a.po'lu.sis. Gr. n. *apólousis*, sewage; N.L. gen. n. *apolusis*, of or related to waste water and sewage.

Caldisericia sister clade member: "*Ca.* Caldimicrobium aminisyntrophus"
Cal.di.bac.te'ri.um. L. adj. *caldus*, hot; N.L. n. *microbium*, a small living entity; N.L. neutr. n. *Caldimicrobium*, a thermophilic microbe of *Caldiserica* ancestry.
a.mi.ni.syn.tro'phus. N.L. n. *aminum*, an amine; Gr. adj. *syntrophos*, having grown up with one, living with; N.L. masc. adj. *aminisyntrophus*, degrading amino acids in syntrophic association with partner organisms.

Aminiphilus member: "*Ca.* Aminiphilus thermophilus"
ther.mo'phi.lus. Gr. fem. n. *thermê*, heat; N.L. adj. *philus* -a -um (from Gr. adj. philos -ê -on), loving; N.L. masc. adj. *thermophilus*, heat loving.
Synergistetes cluster TTA-B6 member: "*Ca.* Aminitrophus cloacae"
A.mi.ni.tro'phus. N.L. n. aminum, an amine; Gr. adj. *trophos*, one who feeds; N.L. masc. adj. *Aminitrophus*, amino acid eating.
clo.a'cae. L. fem. n. *cloaca*, sewer; N.L. gen. n. *cloacae*, of or related to sewage.

Spirochaetes clade SA-8 member: "*Ca.* Argininisyntropha thermophila"
Ar.gi.ni.ni.syn.tro'pha. N.L. n. *arginínum*, arginine; Gr. n. *syntrophos*, having grown up with one, living with; N.L. fem. n. *Argininisyntropha*, one who degrades arginine in syntrophic association with partner organisms.
ther.mo'phi.la. Gr. fem. n. *thermê*, heat; N.L. adj. *philus* -a -us (from Gr. adj. philos -ê -on), loving; N.L. fem. adj. *thermophila*, heat loving.

"*Ca.* Atribacteria" clade JS1 (OP9 clade JS1) member: "*Ca.* Atricorium thermopropionicum"
Atri.co'ri.um. L. adj. *ater -tra –trum*, black; L. neut. n. *corium*, layer or coating. N.L. neut. n. *Atricorium*, black film.
ther.mo.pro.pi.o'ni.cum. Gr. fem. n. thermê, heat; N.L. n. *acidum propionicum*, propionic acid; N.L. neut. adj. *thermopropionicum*, thermophilic and pertaining to propionate.

"*Ca.* Cloacimonetes" (WWE1) member: "*Ca.* Cloacimonas propionivorans"
Clo.a.ci.mo'nas. L. n. *cloaca*, sewer; -i- connecting vowel; N.L. n. *monas* a unit, N.L. fem. n. *Cloacimonas* a unit from a sewer
pro.pio.ni.vo'rans. N.L. n. *acidum propionicum*, propionic acid; L. v. *voro*, to eat, to devour; N.L. part. adj. *propionivorans*, devouring propionic acid.

D.2 – Supplementary Figures

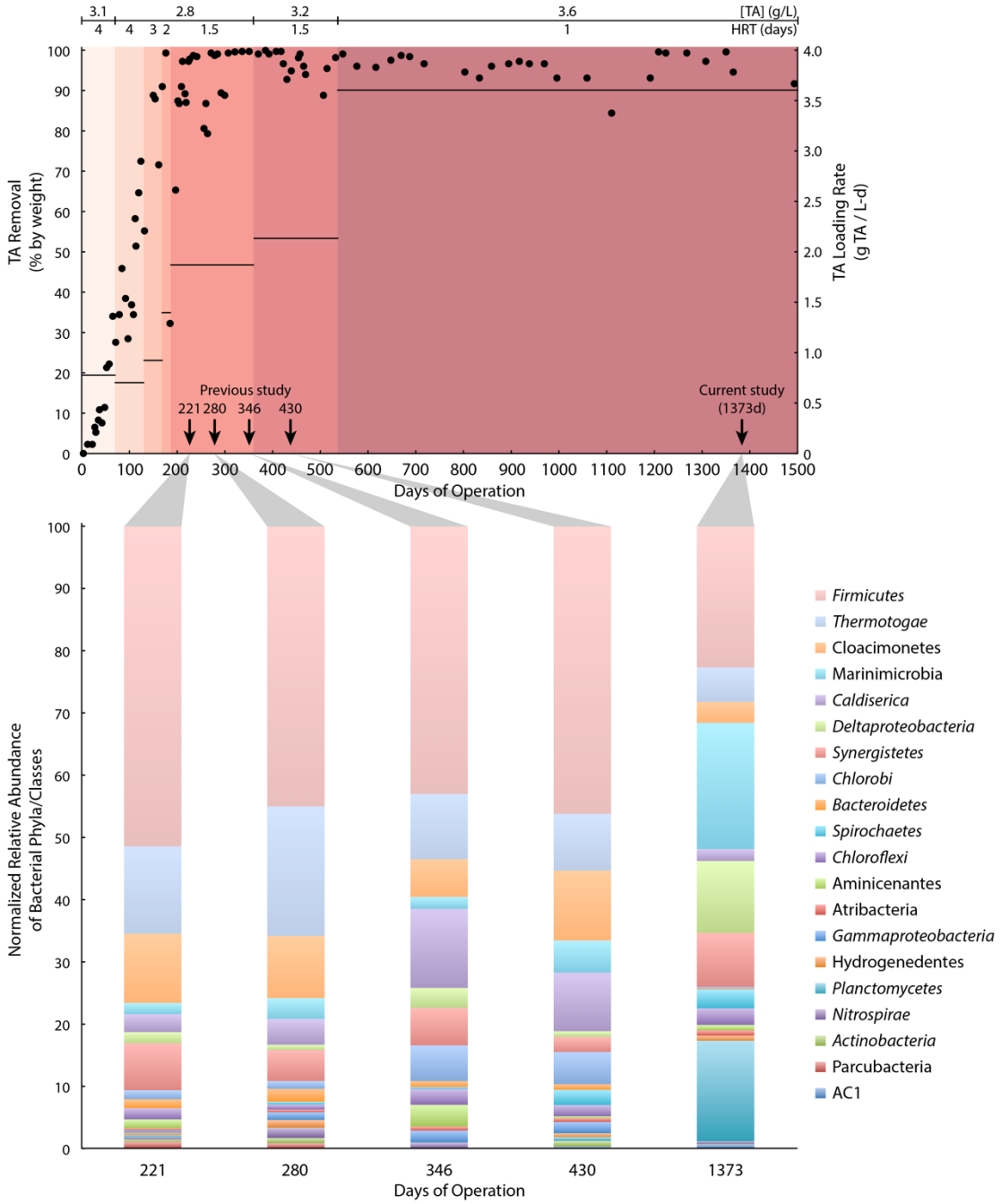

Figure D.1. Reactor performance over 1500 days of operation with increasing terephthalate (TA) concentration and decreasing hydraulic retention time (HRT) leading to gradual increase in TA loading rate (horizontal lines corresponding to right-hand vertival axis in top panel). TA removal (effluent concentration / influent concentration; right-hand vertical axis) was measured over time (black points in top panel). DNA sampling points from the previous study (221, 280, 346, and 430 days) and current study (1373 days) are shown with corresponding phylum/class-level bacterial community composition normalized to total bacterial sequence number.

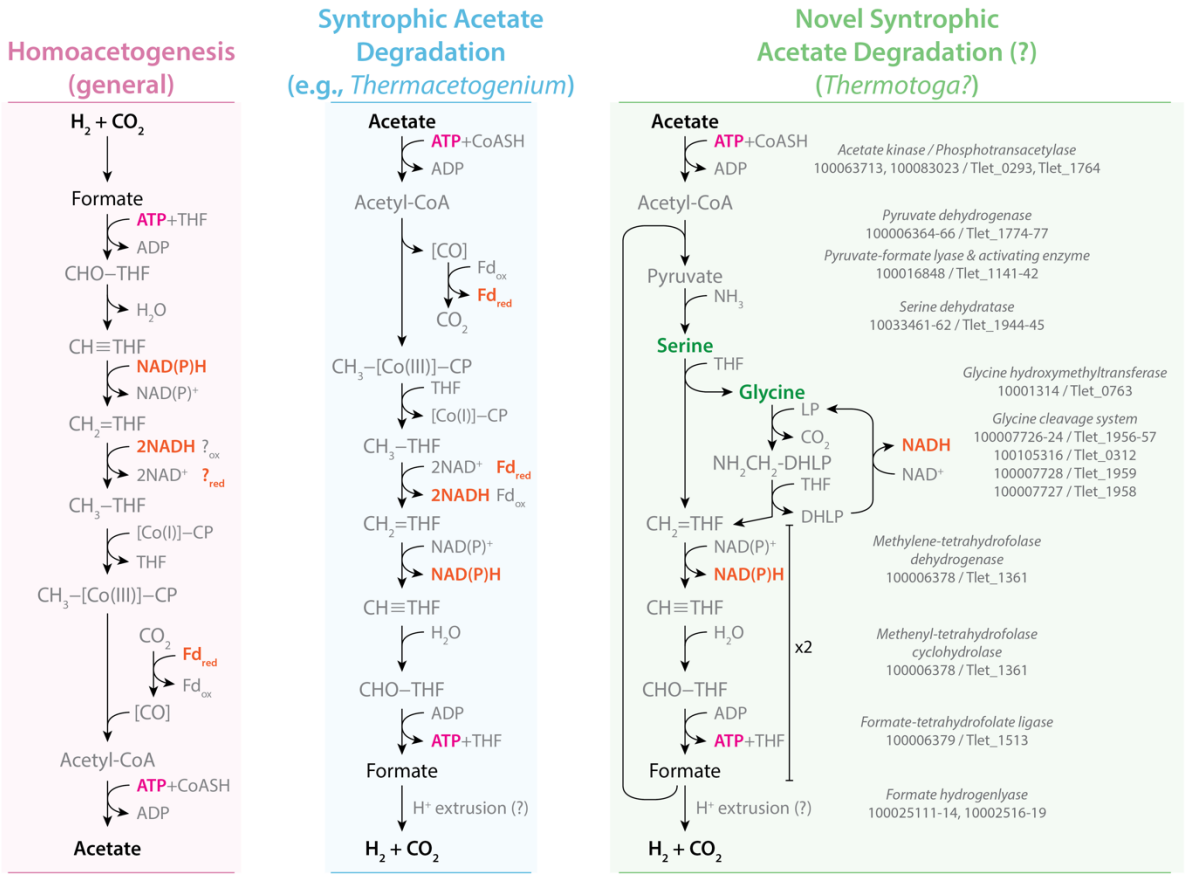

Figure D.2. Comparison of the general acetogenesis pathway (red) with previously proposed (blue) and hypothesized syntrophic acetate degradation pathways (green). The newly proposed syntrophic acetate degradation pathway (green) is mediated by serine synthesis, glycine cleavage system, tetrahydrofolate pathway, and final formate conversion to hydrogen. For each step, the expressed genes are shown for the TA-degrading community *Mesotoga*. The corresponding *Pseudothermotoga lettingae* strain TMO loci are also noted. Co-reactants and co-products include adenosine triphosphate (ATP, magenta), adenosine diphosphate (ADP), oxidized/reduced ferredoxin (Fd_{ox} and Fd_{red}, orange), coenzyme A (CoA), tetrahydrofolate (THF), lipoylprotein (LP), oxidized/reduced nicotinamide adenine dinucleotide (NAD^+ and NADH, orange), and oxidized/reduced nicotinamide adenine dinucleotide phosphate ($NADP^+$ and NADPH, orange).

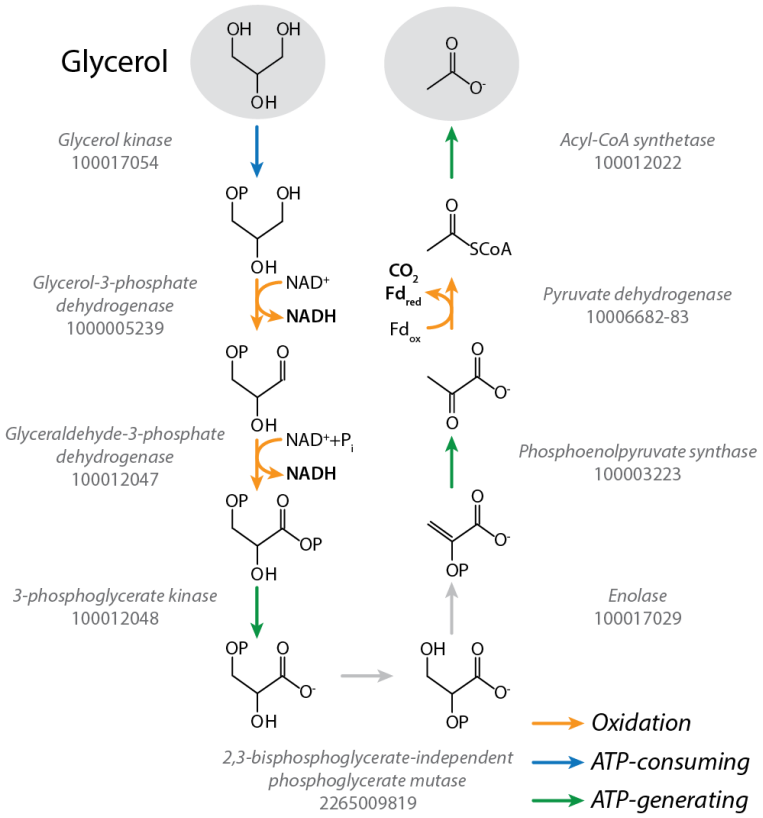

Figure D.3. The putative syntrophic glycerol degradation pathway employed by the TA-degrading community Hydrogenedentes and *Planctomycetales* member. This pathway degrades glycerol through a phosphoenolpyruvate intermediate to acetate. For each step, the expressed genes are shown for Hydrogenedentes. Oxidative (orange) reactions show the oxidized and reduced (bolded) electron carriers. ATP -consuming and -generating steps are shown in blue and green respectively. Co-reactants and co-products include coenzyme A (CoA) oxidized/reduced ferredoxin (Fd_{ox} and Fd_{red}) and oxidized/reduced nicotinamide adenine dinucleotide (NAD^+ and NADH).

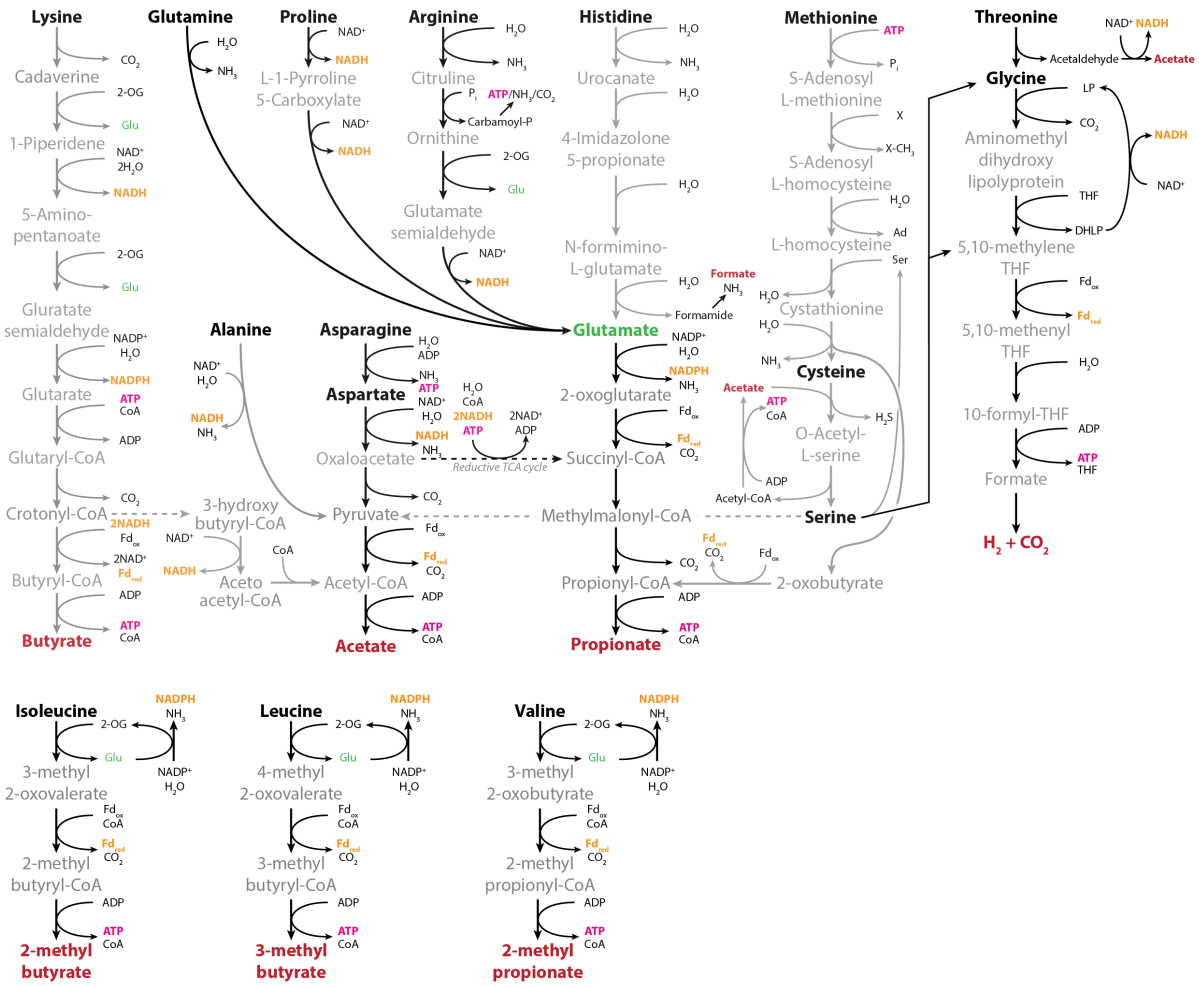

Figure D.4. Theoretical syntrophic amino acid degradation pathways. The specific pathways that are expressed by the TA-degrading community Marinimicrobia, *Phycisphaerae*-related WPS-1 cluster, and *Spirochaetes* clade SA-8 members are denoted with black arrows. Phenylalanine, tyrosine, and tryptophan metabolism is not shown because anaerobic degradation of these compounds has not been demonstrated under methanogenic conditions. Co-reactants and co-products include adenosine triphosphate (ATP, magenta), adenosine diphosphate (ADP), coenzyme A (CoA), tetrahydrofolate (THF), lipoylprotein (LP), 2-oxoglutarate (2-OG), glutamate (Glu), serine (Ser), adenosine (Ad), oxidized/reduced ferredoxin (Fd_{ox} and Fd_{red}, orange), oxidized/reduced nicotinamide adenine dinucleotide phosphate ($NADP^+$ and NADPH, orange) and oxidized/reduced nicotinamide adenine dinucleotide (NAD^+ and NADH, orange).

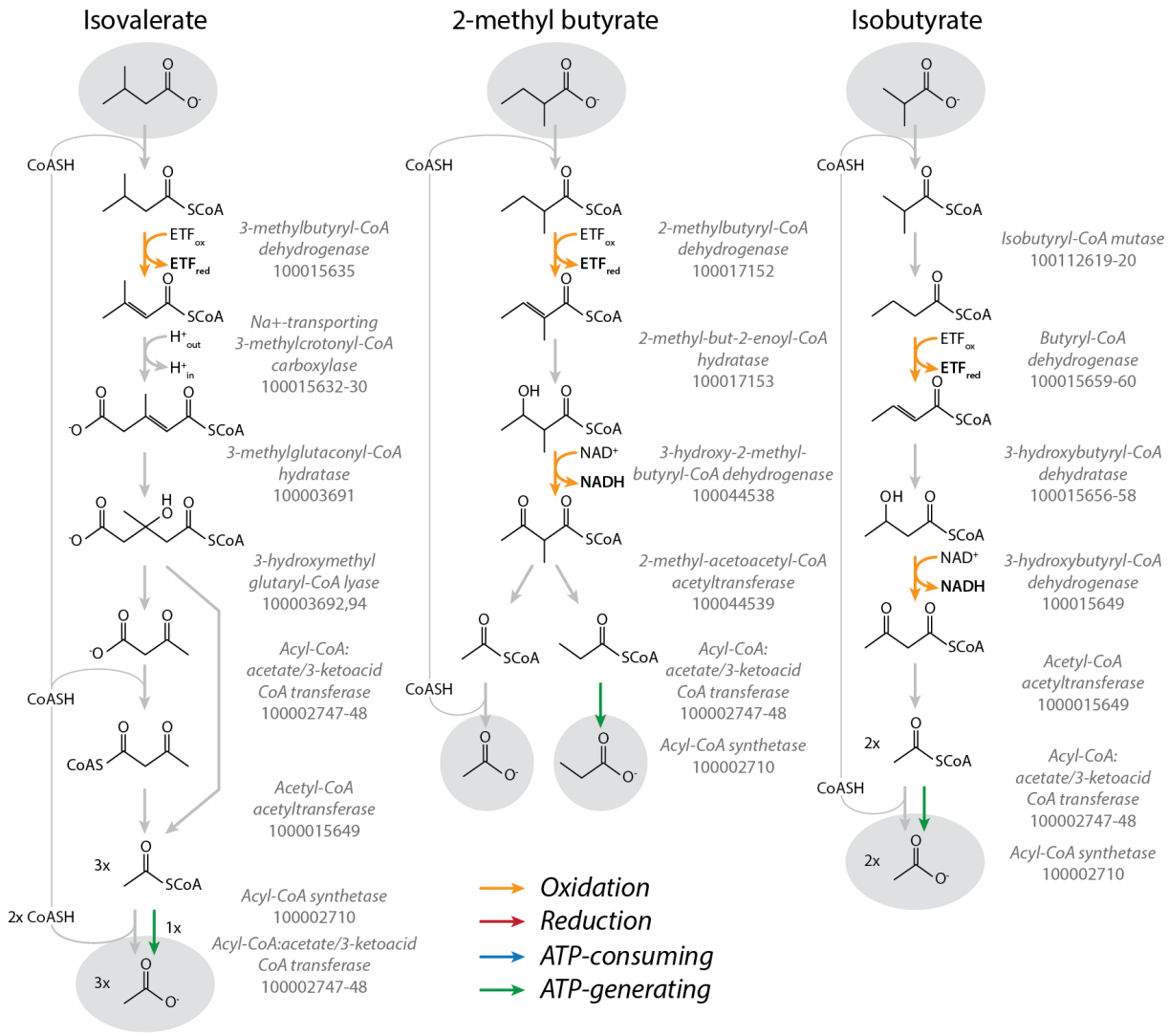

Figure D.5. Newly proposed syntrophic isovalerate, 2-methylbutyrate, and isobutyrate degradation pathway. Isovalerate is oxidatively degraded through a membrane-gradient-driven carboxylation to ultimately generate acetate. 2-methyl butyrate is metabolized by oxidatively producing a 2-methyl-acetoacetyl-CoA intermediate that can be thiolytically cleaved to yield acetate and propionate. Isobutyrate is isomerized to form a straight-chain fatty acid for downstream butyryl-CoA degradation to acetate. For each step, the expressed genes are shown for the TA-degrading community *Syntrophus*-related clade member. Oxidative (orange) reactions show the oxidized and reduced (bolded) electron carriers. ATP -consuming and -generating steps are shown in blue and green respectively. Co-reactants and co-products include coenzyme A (CoASH or SCoA), cytosolic and non-cystosolic membrane-associated proton (H^+_{in} and H^+_{out}) oxidized/reduced electron transfer flavoprotein (ETF_{ox} and ETF_{red}), oxidized/reduced ferredoxin (Fd_{ox} and Fd_{red}) and oxidized/reduced nicotinamide adenine dinucleotide (NAD^+ and NADH).

D.3 – Supplementary Tables

Refer to supplementary file "Nobu2017_AppendixD_Tables.xlsx" for the following tables:

Table D.1. Microbial community structure based on EMIRGE-assembled 16S rRNA gene sequences.

Table D.2. Metagenomic bins, single-cell amplified genomes (SAGs), and pangenome specifications, 16S rRNA genes, and energy conservation genes.

Table D.3. Binning results from PhylopythiaS, Metawatt 1.7 (N4 medium used for Metawatt), and BLAST annotation by Metawatt.

Table D.4. Terephthalate metabolism and acetate/butyrate production pathways found in *Pelotomaculum* and *Syntrophorhabdus* metagenomic bins and single-cell genomes.

Table D.5. Methanogenic hydrogen, formate, and acetate metabolism of *Methanolinea, Methanospirillum,* and *Methanosaeta* metagenomic bins and single-cell genomes.

Table D.6. Newly proposed syntrophic acetate degradation pathway found in *Mesotoga* and *Pseudothermotoga* metagenomic bins and single-cell genomes and also *Pseudothermotoga lettingae* strain TMO.

Table D.7. Syntrophic butyrate oxidation and newly proposed syntrophic branched fatty acid degradation pathways found in the *Syntrophus*-related clade pangenome.

Table D.8. Acetogenesis pathways found in the *Chloroflexi* subphylum I member pangenome.

Table D.9. Lipid hydrolysis and glyceorl fermentation pathways found in the Planctomycetales-related clade bin and Hydrogenedentes pangenome.

Table D.10. Protein hydrolysis and amino acid degradation pathways found in the Marinimicrobia pangenome and *Planctomycetes* cluster WPS-1, *Caldisericia* sister clade, *Aminiphilus, Synergistetes* cluster TTA-B6, and *Spirochaetes* cluster SA-8 draft genomes.

Table D.11. Syntrophic propionate degradation pathways found in the Atribacteria and Cloacimonetes pangenomes.

Table D.12. Acetogen genome survey for energy conservation genes and pathways.

APPENDIX E

SUPPLEMENTARY MATERIAL TO CHAPTER 6

E.1 – Supplementary Figures

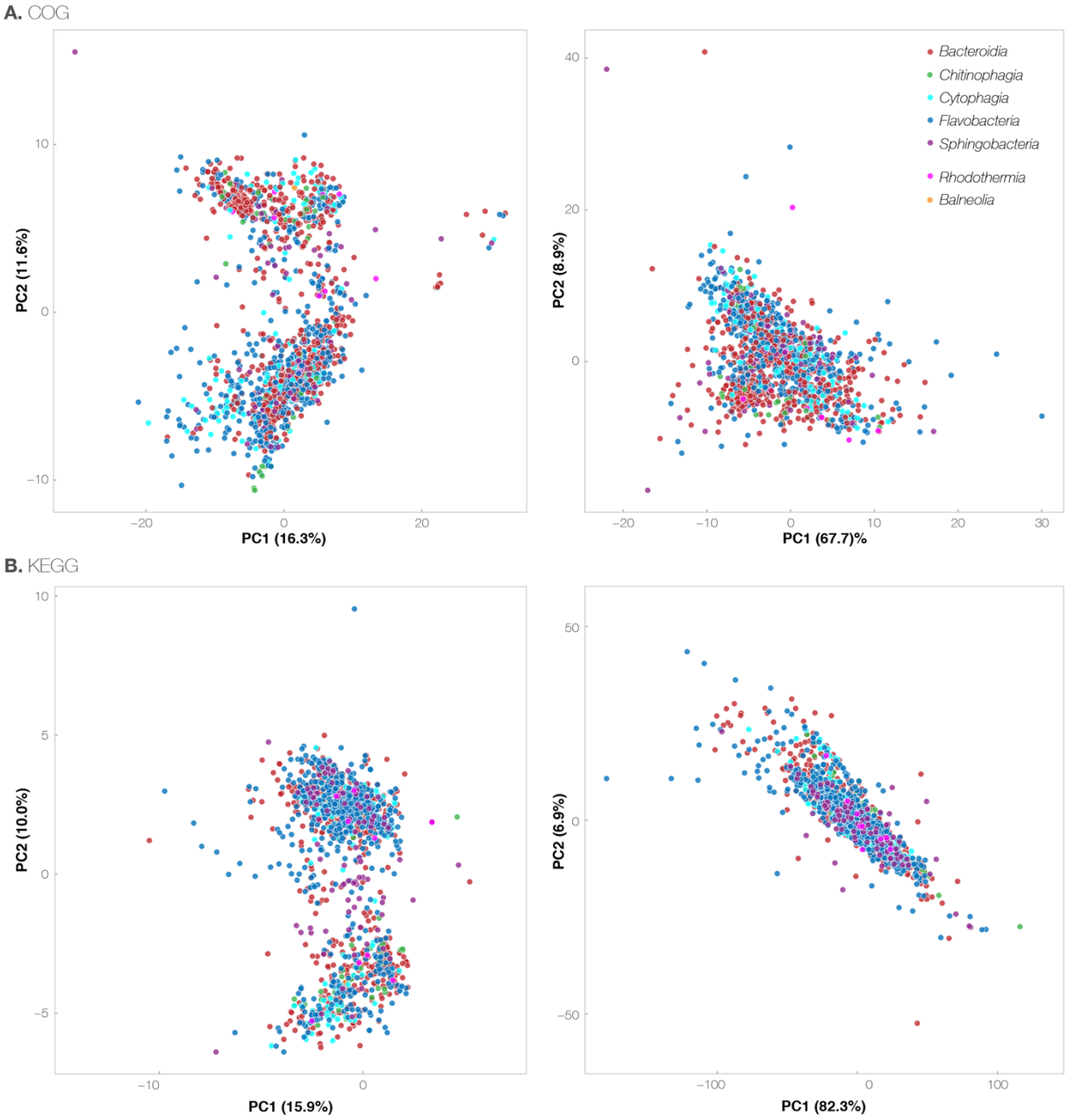

Figure E.1. Comparative genomics of *Bacteroidetes* genomes based on COG and KEGG gene/pathway annotation. Principal component analysis of presence/absence of function based on (A, left) cluster of orthologous gene (COG), (A, right) COG pathway, (B, left) Kyoto Encyclopaedia of Genes and Genomes (KEGG), and (B, right) KEGG pathway annotation available on Joint Genome Institute's IMG. Each genome is colored based on their phylogeny. *Balneolia* and *Rhodothermia* of the phylum *Rhodothermaeota*, closely related to *Bacteroidetes*, are also included for reference.

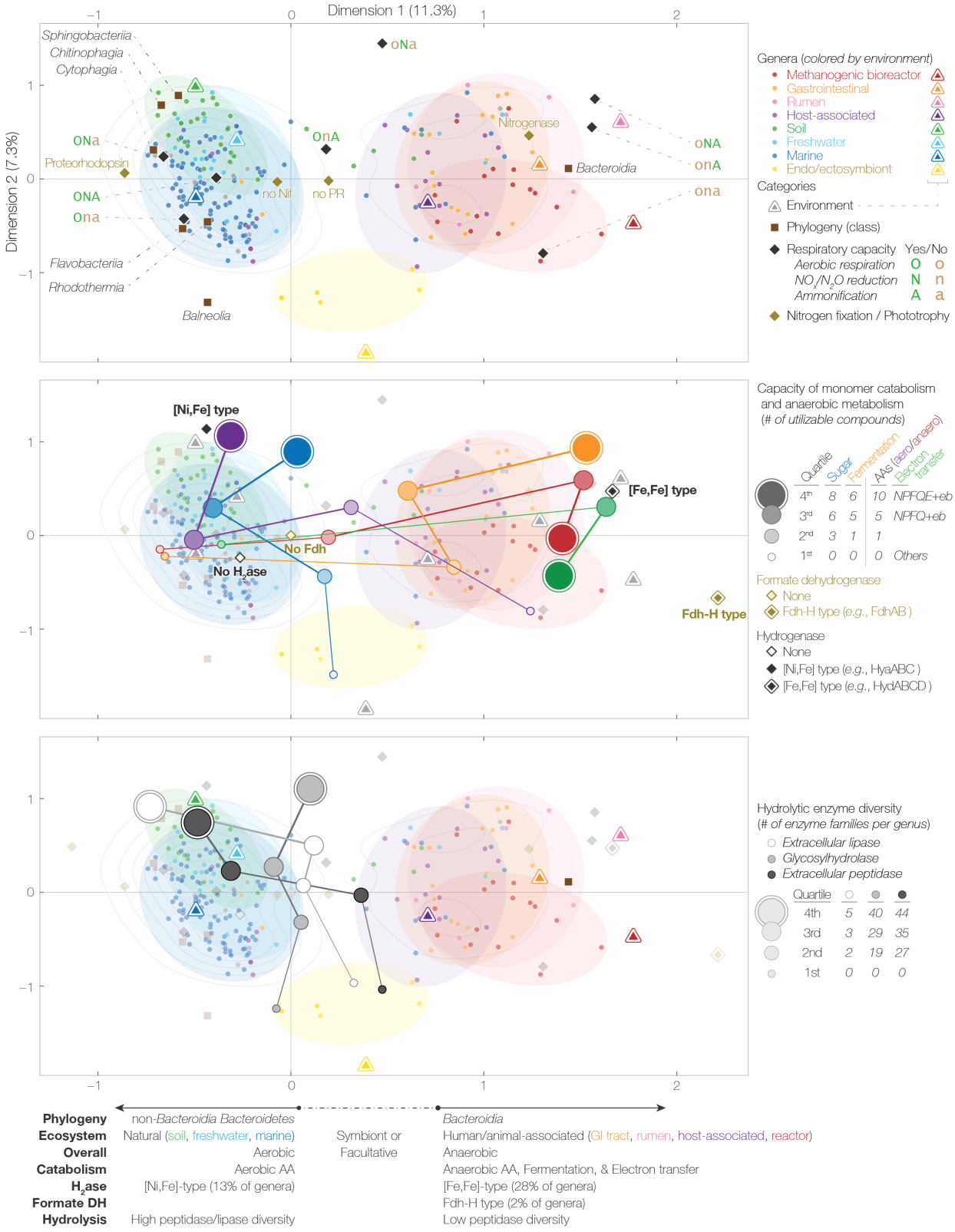

Figure E.2. Phylum-level multiple correspondence analysis of *Bacteroidetes* genera metabolic capacities. Each genus is colored by their most associated habitat: bioreactor (red), gastrointestinal (orange), rumen (pink), host-associated (purple), marine (blue), freshwater (cyan), soil (green), and

Figure E.2. (cont.) ecto/endosymbionts (yellow). MCA yields the correlation of these genera's traits as vectors and labels are shown at their respective coordinates (*i.e.*, weight and direction). The top panel shows phylogeny, respiration, and habitat. The middle panel shows the correlation of genera with different levels of sugar degradation (blue), anaerobic (red) and aerobic (purple) amino acid degradation, fermentation (orange), and electron transfer capacity (green) with lines connecting them in increasing capacity. This panel also displays hydrogen (black) and formate (brown) metabolism. In a similar fashion, the bottom panel shows different levels of diversity in extracellular lipase (white), glycosylhydrolase (gray), and extracellular peptidase (black). Levels of sugar degradation, fermentation, and hydrolytic enzymes are defined by quartiles ($<1^{st}$, $<2^{nd}$, $<3^{rd}$, and $\leq 4^{th}$ quartile), amino acid degradation by percentage of 20 standard amino acids, and electron transfer capacity by the ability to perform electron-bifurcating (eb) electron transduction.

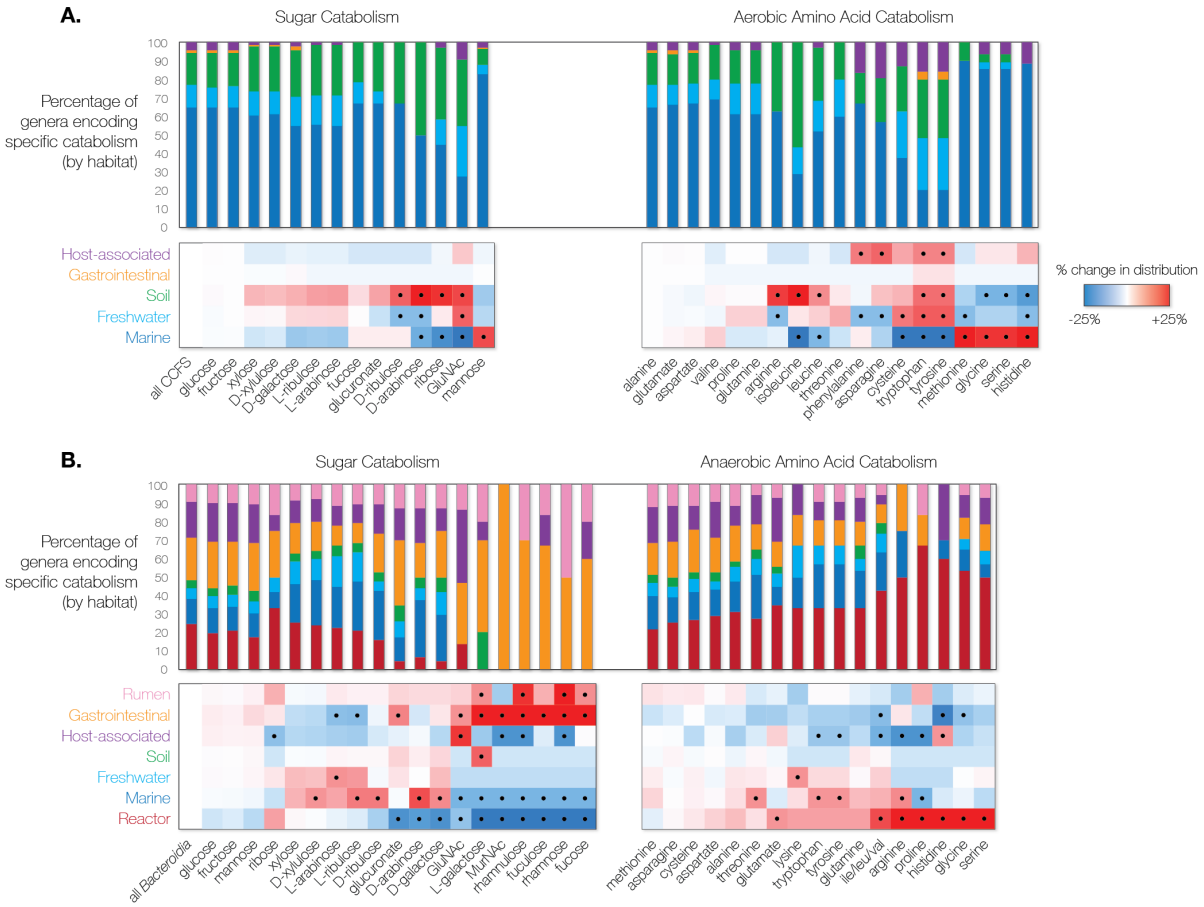

Figure E.3. Distribution of sugar and amino acid degradation pathways in CCFS and *Bacteroidia* from differing habitats. (A, top) For each given sugar/AA-catabolizing pathway, the percentages of CCFS genera affiliated with differing habitats among all CCFS genera harboring the pathway are shown. As a reference, the distribution of CCFS genera in different habitats is shown on the left-hand side. (A, bottom) Corresponding to these distributions, the change in the habitat percentages are shown below (increase=red, decrease=blue). A black dot is shown for those that with at least 10% change. (B) The same analysis is shown for *Bacteroidia*.

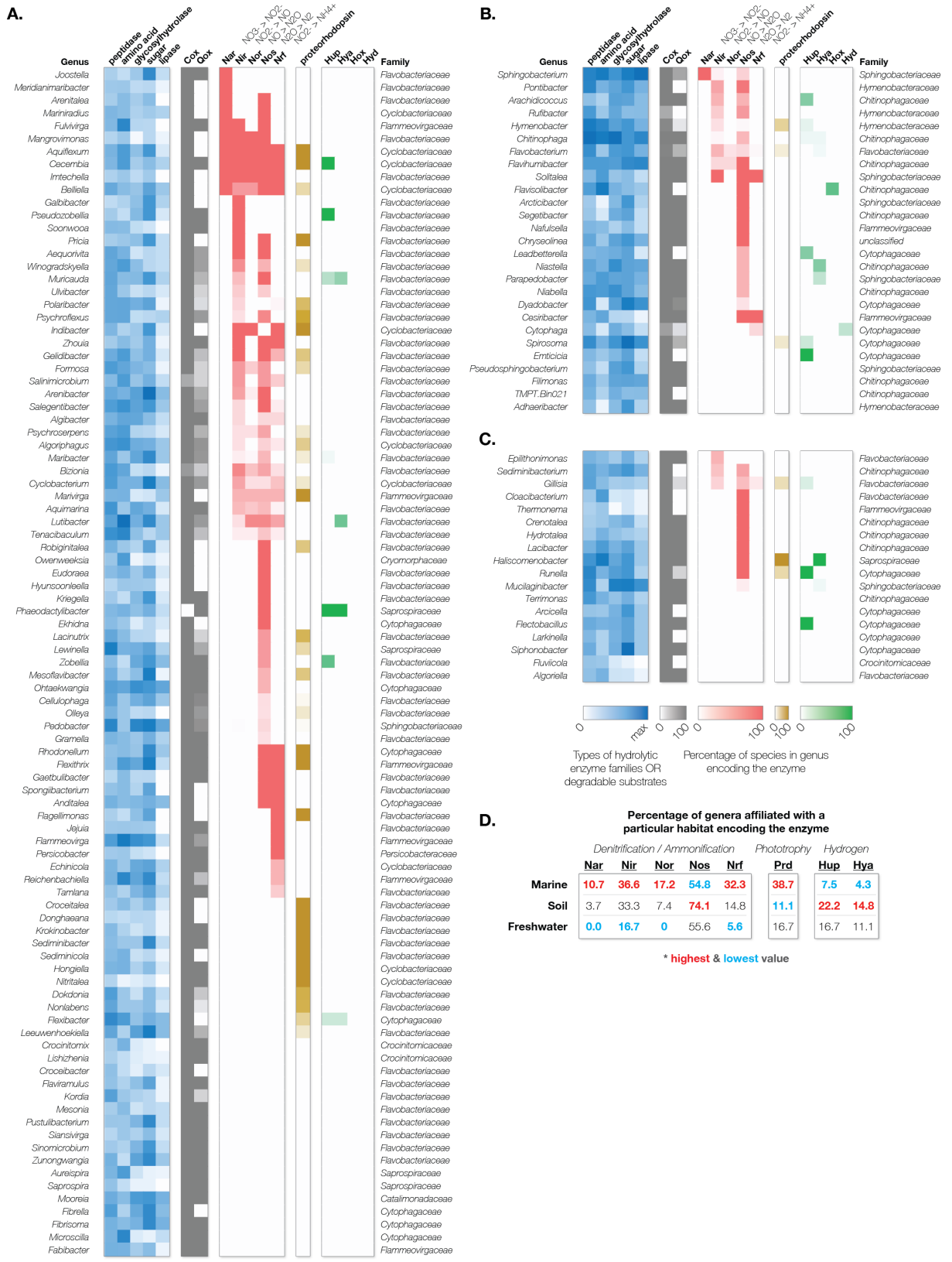

Figure E.4. Distribution of supplemental energy metabolism in aerobic environmental *Bacteroidetes*. The capacities for amino acid and sugar degradation (blue); protein, carbohydrate, lipid hydrolysis (blue);

Figure E.4. (cont.) aerobic (black) and nitrogen specie (red) respiration (Cox – cytochrome c oxidase, Qox – quinol oxidase, Nar – nitrate reductase, Nir- nitrite reductase, Nor – nitric oxide reductase, Nos – nitrous oxide reductase, and Nrf – ammonifying nitrite reductase); phototrophy (*i.e.*, proteorhodopsin) (brown); and hydrogen metabolism (green) are shown for marine (A), soil (B), and freshwater (C) *Chitinophagia*, *Cytophagia*, *Flavobacteriia*, and *Sphingobacteria* are shown. (D) In addition, the overall percentage of genera encoding denitrification and ammonification enzymes, proteorhodopsin, and hydrogen metabolism (Hup – uptake hydrogenase, Hya – membrane-bound quinol-dependent hydrogenase).

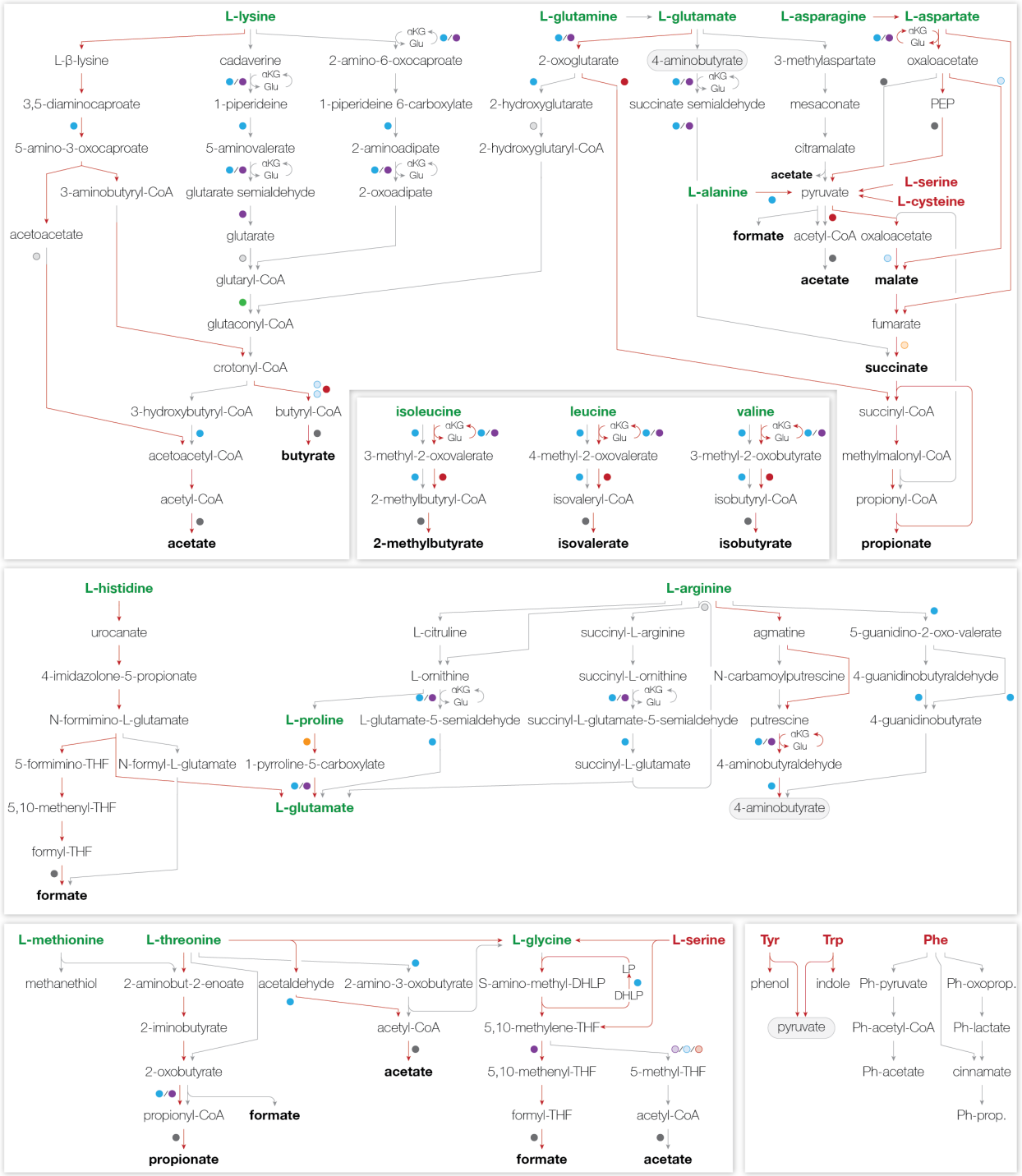

Figure E.5. Pathways for anaerobic degradation of proteinogenic amino acids. All known pathways are shown except for Stickland fermentation, which is thought to be restricted to *Clostridia*. Each reduction/oxidation reaction is labeled with the corresponding electron carrier(s): NAD(H) (blue), NADP(H) (purple), ferredoxin (red), and electron transfer flavoprotein (green); open if reduction reaction and closed if oxidation. Amino acids are colored by whether they can be degraded syntrophically (green) or primarily through fermentation (red). In addition, pathways that ADurb.Bin008 encode are colored red.

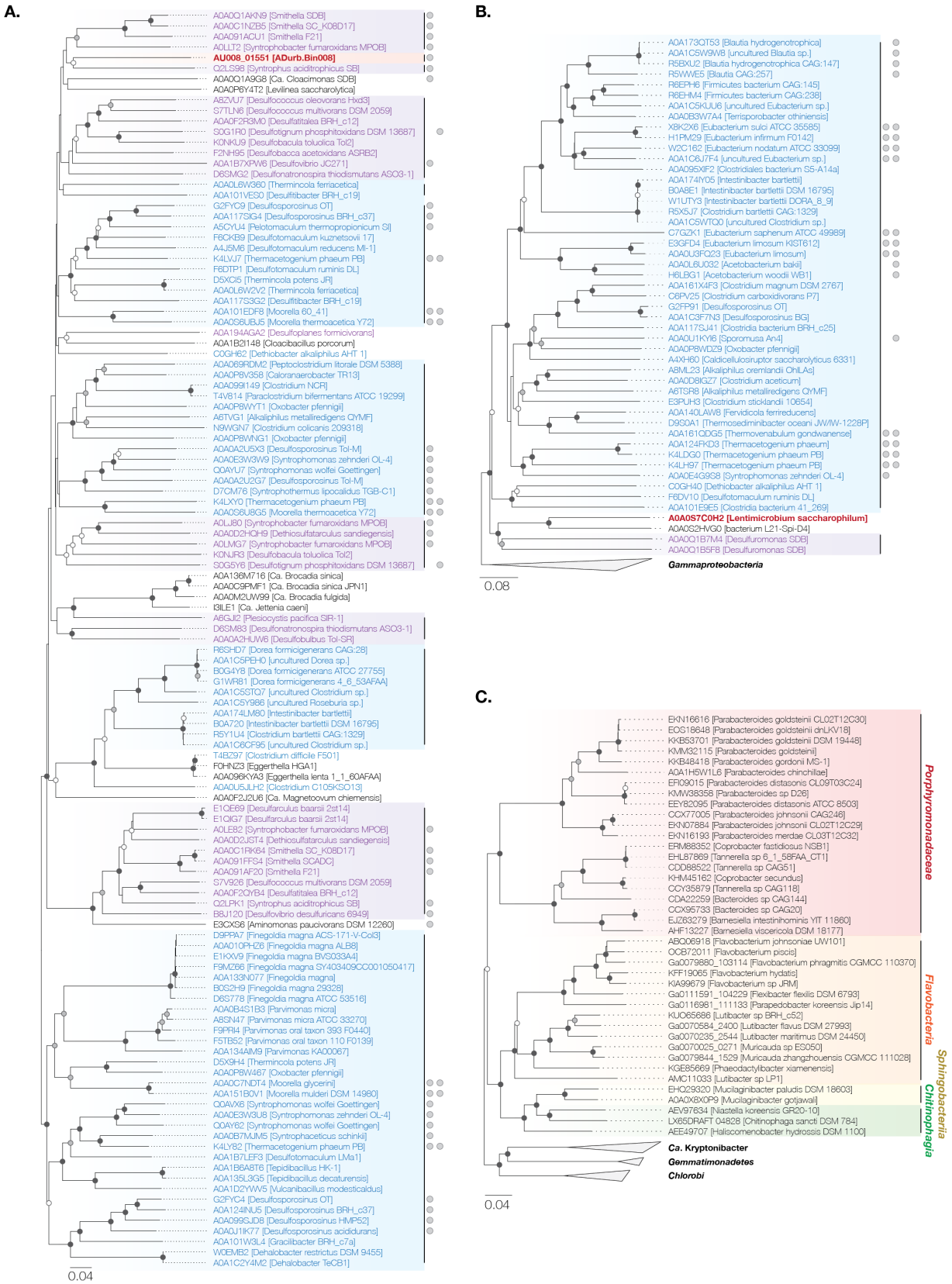

Figure E.6. Phylogeny of formate dehydrogenases and Ni,Fe Hya uptake hydrogenases.

Figure E.6. (cont.) (A) Formate dehydrogenase catalytic subunits with at least 40% gene similarity to that of ADurb.Bin008 are collected from the UniProt database and phylogenetically compared using peptide sequence alignment and tree construction by ClustalW2. Bootstrap values (>50% white, >75% gray, and >90% black) from 1000 iterations are shown. Genes affiliated with *Firmicutes* (blue), *Deltaproteobacteria* (purple), syntrophic bacteria (left gray circle), and homoacetogenic genera (right gray circle) are the majority. (B) The same phylogenetic analysis is performed for *Lentimicrobium* putative formate dehydrogenase with unknown function. (C) The phylogenetic relationship of *Bacteroidetes* Ni,Fe Hya uptake hydrogenase catalytic subunits (HyaB) found in *Bacteroidia* and CCFS are shown. The gene IDs are either Genbank/UniProt accession numbers or IMG locus tags.

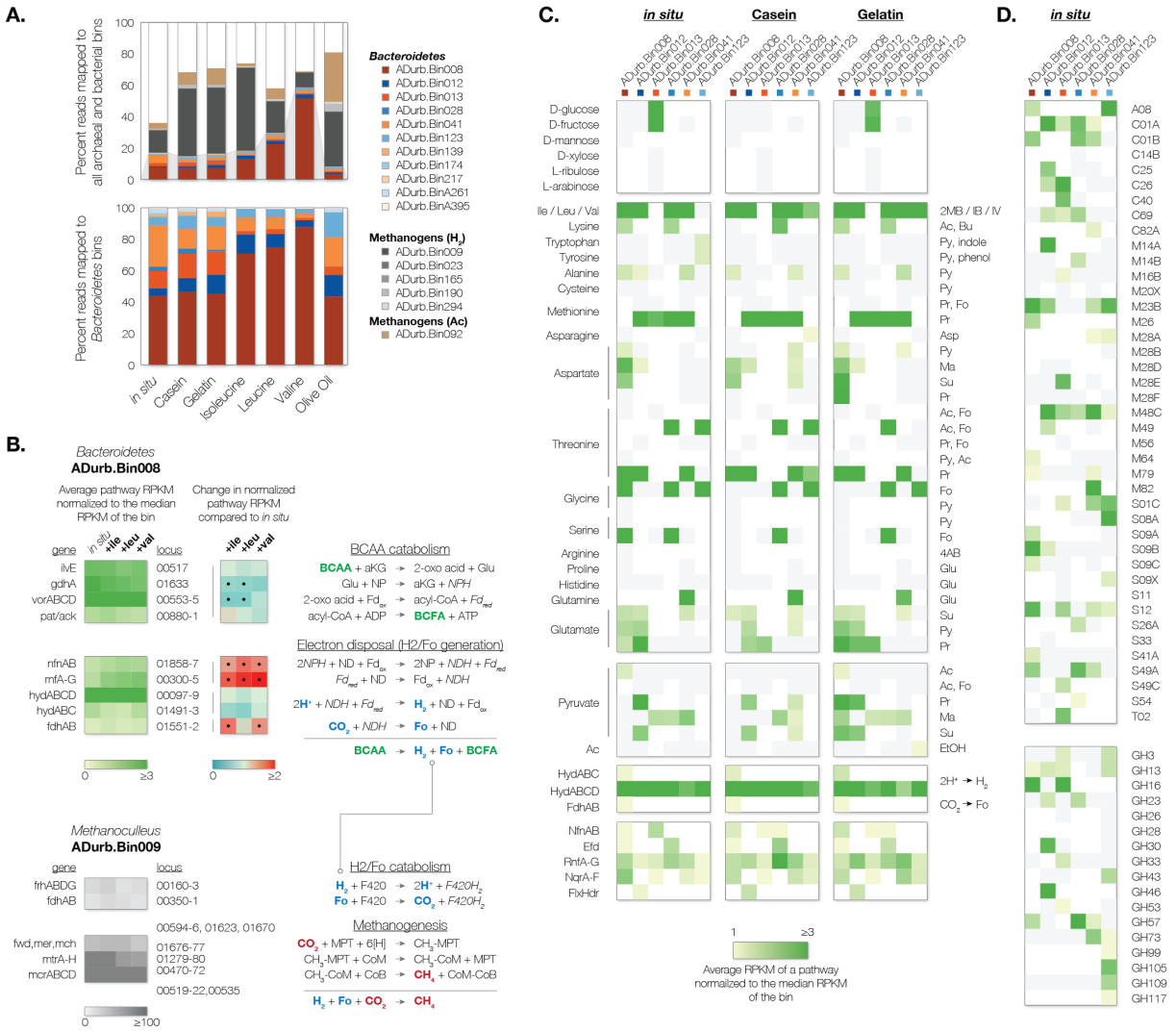

Figure E.7. Metatranscriptomic analysis of anaerobic digester *Bacteroidia*. (A) For metatranscriptomes recovered from anaerobic digester sludge and sludge incubated with specific substrates (all triplicate), the contribution of *Bacteroidetes* bins and *Euryarchaeota* methanogens to the total mapped metatranscriptome is shown (top). The relative contribution of each *Bacteroidetes* bin of all *Bacteroidetes* bins is shown as well (bottom). Model protein (casein and gelatin), branched-chain amino acid (BCAA: isoleucine, leucine, and valine), and model lipid (olive oil) were evaluated as substrates. (B) For anaerobic digester samples amended with BCAA, the expression levels of genes (reads per kilobase transcript per million reads (RPKM) normalized to the bin's median RPKM) supporting (i) ADurb.Bin008 BCAA degradation, H_2 generation, formate (Fo) generation, and essential energy-conserving electron transduction enzymes and (ii) ADurb.Bin009 H_2 catabolism, Fo catabolism, and methanogenesis are shown with their respective reactions and locus tags (without locus prefix; BWX49 for ADurb.Bin008 and BWX50 for ADurb.Bin009). For ADurb.Bin008, the ratio of each pathway's RPKM under BCAA-amended to *in situ* conditions are calculated. In addition, each pathway is marked with a black dot if the RPKM is significantly higher ($p<0.05$ and >25% increase; red) or lower ($p<0.05$ and >25% decrease; blue) than *in situ*. (C) For the six most active *Bacteroidetes in situ*, the average RPKM of metabolic pathways are colored if they meet the following criteria (otherwise gray): (i) the average RPKM of the genes involved in the pathway is greater than the median (excluding zeroes), (ii) RPKM for at least 2/3 of the genes are greater than the median, and (iii) genes unique to the pathway have RPKM greater than the median. (D) Similarly, the normalized RPKM of extracellular

Figure E.7. (cont.) peptidase and glycosylhydrolase families are shown. Abbreviations: Ac – acetate; Pr – propionate; Bu – butyrate; 2MB – 2-methylbutyrate; IB – isobutyrate; IV – isovalerate; Py – pyruvate; 4AB – 4-aminobutyrate; Ma – malate; Su – succinate.

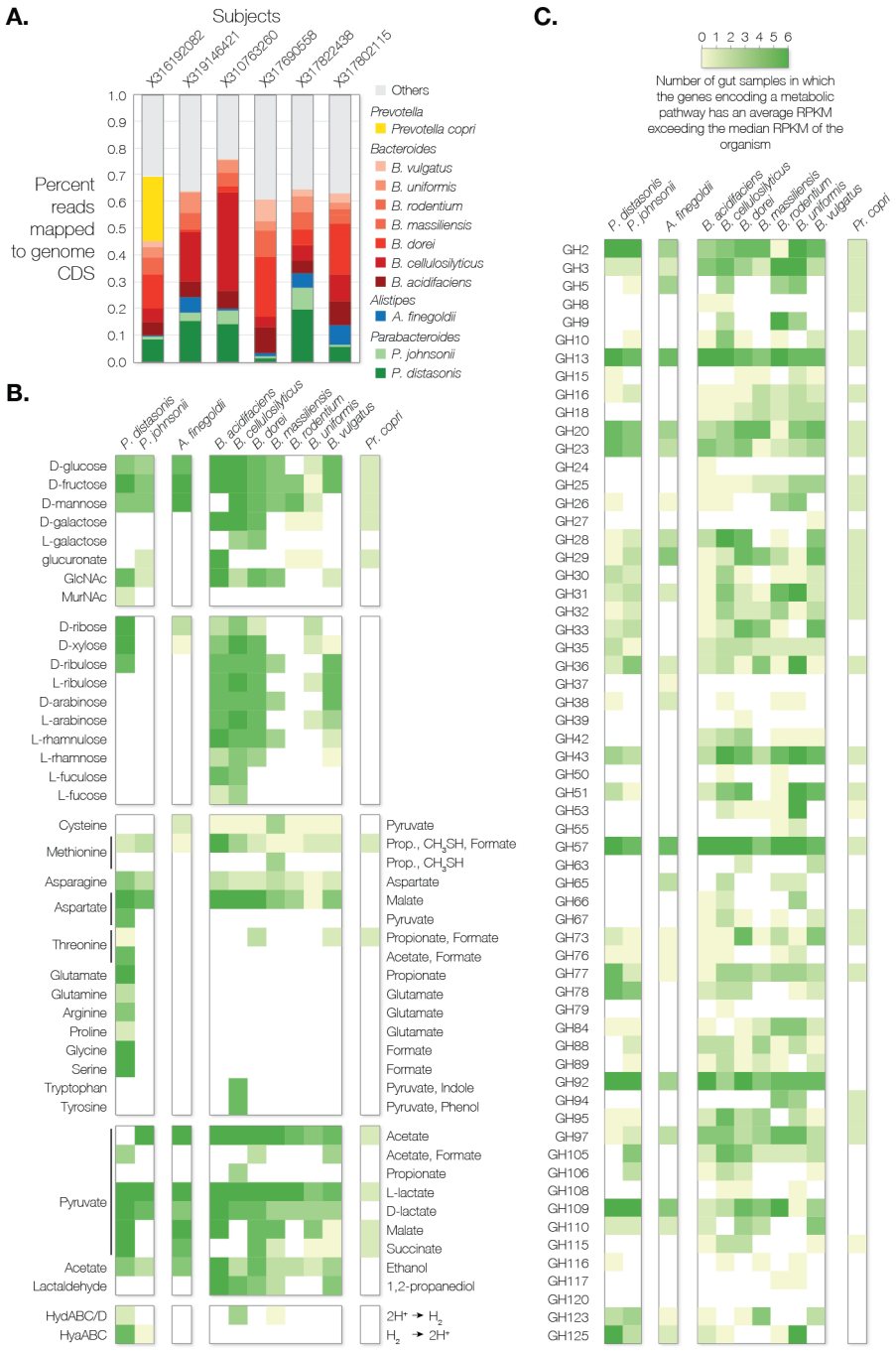

Figure E.8. Metatranscriptomic analysis of human gastrointestinal tract *Bacteroidia*. (A) Metatranscriptomes of gastrointestinal tract samples from six individuals (*279*) were mapped to all *Bacteroidia* genomes to determine the relative activity of each *Bacteroidetes* species in the gut. Among 107 *Bacteroidetes* genomes with at least one read mapped, the top contributor *Bacteroidetes* species were defined based on the following criteria: (i) are within the top ten *Bacteroidetes* in terms of total reads mapped per genome size in at least half of the samples or (ii) occupy at least 10% of the total reads mapped to *Bacteroidetes*. (B) For each of these *Bacteroidetes* species, metabolic pathways are colored based on the number of samples (0 to 6) in which the pathway met the following criteria: (i) the average RPKM of the genes involved in the pathway is greater than the median (excluding zeroes), (ii) RPKM for at least 2/3 of the genes are greater than the median, and (iii) genes unique to the pathway have RPKM greater

Figure E.8. (cont.) than the median. (C) Similarly, the frequencies at which glycosylhydrolase families are expressed at an RPKM greater than the median are also shown.

E.2 – Supplementary Tables

Phylum	Class	Family	Genus	Bin	Genome size (Mb)	Contigs	Coding Sequence	Completeness (%)	Contamination (%)	Coverage	Relative abundance (% total mapped metagenomic reads)	Relative activity (% total mapped metatranscriptome)
Bacteroidetes	*Bacteroidia*	AU008		ADurb.Bin008	3.06	68	2442	92.9	1.4	127.9	3.14	8.01
		AU012		ADurb.Bin012	2.64	154	2132	94.4	3.7	61.3	1.51	0.81
		AU013		ADurb.Bin013	2.26	122	1971	92.6	4.2	62.1	1.53	2.05
		AU028		ADurb.Bin028	2.92	376	2483	96.2	3.5	55.9	1.38	0.62
		AU041		ADurb.Bin041	3.08	254	2432	95.4	4.1	18.3	0.45	5.69
		Prolixibacteraceae		ADurb.Bin123	3.50	111	2784	97.3	2.3	13.9	0.34	1.06
		AU013		ADurb.Bin139	2.18	181	1948	92.8	5.2	8.2	0.20	0.29
		AU145		ADurb.Bin145	3.96	219	3307	99.3	5.7	11.6	0.29	0.18
		Porphyromonadaceae		ADurb.Bin174	2.95	93	2286	95.7	0.9	6.1	0.15	0.21
		AU217		ADurb.Bin217	3.10	194	2591	98.9	0.8	7.8	0.19	0.01
		Porphyromonadaceae		ADurb.BinA261	2.32	365	1932	93.8	4.2	3.0	0.07	1.46
		Porphyromonadaceae		ADurb.BinA395	2.37	228	1856	94.9	5.8	2.0	0.05	0.00
Euryarchaeota	*Methanomicrobia*	*Methanomicrobiaceae*	*Methanoculleus*	ADurb.Bin009	1.56	710	1712	68.1	3.5	192.5	4.73	9.73
	Methanomicrobia	*Methanomicrobiaceae*	*Methanoculleus*	ADurb.Bin190	1.57	308	1644	64.8	3.9	13.0	0.32	0.23
	Methanomicrobia	*Methanomicrobiaceae*	*Methanospirillum*	ADurb.Bin165	1.62	453	1581	81.3	4.6	11.6	0.28	0.16
	Ca. Methanofastidiosa			ADurb.Bin023	1.53	150	1617	87.3	2.0	67.0	1.65	0.80
	Methanomicrobia	*Methanosarcinaceae*	*Methanosaeta*	ADurb.Bin092	2.51	278	2445	84.8	7.5	14.2	0.35	3.78
Spirochaetes	*Spirochaetes*	*Spirochaetaceae*	SA-8	ADurb.Bin001	1.99	484	1751	91.7	2.4	478.0	11.75	13.01
Proteobacteria	*Deltaproteobacteria*	*Syntrophaceae*	*Smithella*	ADurb.Bin002	1.28	554	1261	46.2	1.9	311.7	7.66	0.97
Ca. Cloacimonetes			Ca. Cloacimonas	ADurb.Bin003	1.14	553	841	69.0	0.3	156.9	3.86	1.23
Ca. Fermentimonas				ADurb.Bin004	2.23	333	2116	90.7	3.3	374.7	9.21	7.45
Verrucomicrobia	Subdivision 3			ADurb.Bin006	4.80	200	3693	91.2	5.4	228.6	5.62	4.04

Table E.1. Binning results for *Bacteroidetes*, methanogens, and top 5 most abundant members from metagenome of full-scale anaerobic digester located in Urbana, Illinois. The relative abundance and activity of each bin was calculated based on the coverage of each bin in the October 2013 metagenome and metatranscriptome respectively divided by the total coverage of all recovered bin contigs and coding sequence.

	Glycosylhydrolase (CAZY)			**Peptidase** (MEROPS)			**Lipase** (LED)	
	Specific	Anaerobes	Aerobes	Specific	Anaerobes	Aerobes	Specific	Aerobes
Reactor	GH54	GH44/49/50		M (26/82), C85A, S01E	C26		abh26	
Rumen		GH50			C26			
Gastrointestinal	GH101	GH50		M (03B/43A) S37	A28A			
Host-associated	GH104	GH4	GH12	S26A	A28A	M(12B/43B/50B)		abh08.09, abh30.01
Marine	GH47/82/*86*		GH113	*C82B*	C26	M(12B/20F/43B/50B/57), S15		abh08.09, abh30, abh30.01
	GH*86*		GH113	*C82B*		M(10B/12B/20F/36/43B/50B/57/90), S15		abh07.02, abh30, abh30.01
Freshwater		GH49/50		M42	C26	A02A, M(12B/17/43B/50B/57)		abh08.07, abh09.01, abh30, abh30.01
			GH12/113	A28B		A(01A/A02A), M(10A/10B/12B/20B/20F/36/43B/50B/90)		abh08.09, abh30, abh30.01
Soil		GH44				M(12B/43B/57)		abh08.07, abh08.09, abh09.01, abh30, abh30.01
	GH87		GH12/113			A(01A/02A), M(10A/10B/12B/17/20F/43B/57/90)		abh07.02, abh08.07, abh09.01, abh30, abh30.01
Ubiquitous	GH (2, 3, 5, 10, 13, 16, 18, 20, 23, 26, 28, 29, 30, 31, 33, 35, 36, 43, 51, 65, 73, 78, 88, 92, 95, 97, 105, 109)			C (01A, 25), M (01, 03A, 14B, 16B, 23B, 28F, 48C) S (01C, 08A, 09B, 09C, 09X, 12, 13, 26A, 41A, 41B, 46, 49A, 54)			abH32.02	

Table E.2. Distribution of glycosylhydrolase, extracellular peptidase, and extracellular lipase families in *Bacteroidetes*. For each of the three types, the median number of families found in *Bacteroidetes* genera affiliated with specific environments divided into anaerobic (white), facultative (gray), and aerobic (dark gray) are shown. These are also marked if greater than or equal to the 2nd (blue) or 3rd (red) quartile. Hydrolytic enzyme families that were found in at least 5% of the genera associated with a particular niche (*i.e.*, environment and respiration capacity) and less than 5% in other genera are shown as specific to an environment or niche. Hydrolytic enzyme families found in at least 10 out of 11 niches are also shown.

REFERENCES

1. R. K. Thauer, A. K. Kaster, H. Seedorf, W. Buckel, R. Hedderich, Methanogenic *Archaea*: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* **6**, 579-591 (2008).
2. M. J. McInerney, J. R. Sieber, R. P. Gunsalus, Syntrophy in anaerobic global carbon cycles. *Current Opinion in Biotechnology* **20**, 623-632 (2009).
3. A. M. Buswell, Anaerobic Fermentations. *State of Illinois Department of Registration and Education* **Bulletin No. 32**, (1936).
4. T. Abbasi, S. M. Tauseef, S. A. Abbasi, in *Biogas Energy*. (Springer New York, New York, NY, 2012), pp. 11-23.
5. T. Narihiro, Y. Kamagata, in *Manual of Environmental Microbiology, Fourth Edition*. (American Society of Microbiology, 2016).
6. K. Schuchmann, V. Müller, A Bacterial Electron-bifurcating Hydrogenase. *The Journal of Biological Chemistry* **287**, 31165-31171 (2012).
7. D. Sondergaard, C. N. S. Pedersen, C. Greening, HydDB: A web tool for hydrogenase classification and analysis. *bioRxiv*, (2016).
8. K. L. James *et al.*, Pyrophosphate-Dependent ATP Formation from Acetyl Coenzyme A in Syntrophus aciditrophicus, a New Twist on ATP Formation. *MBio* **7**, (2016).
9. T. Shimoyama, S. Kato, S. Ishii, K. Watanabe, Flagellum mediates symbiosis. *Science* **323**, 1574-1574 (2009).
10. I. D. Puspite, Y. Kamagata, M. Tanaka, K. Asano, C. H. Nakatsu, Are Uncultivated Bacteria Really Uncultivable? *Microbes and Environments* **27**, 356-366 (2012).
11. A. J. Stams, C. M. Plugge, Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol* **7**, 568-577 (2009).
12. J. R. Sieber, M. J. McInerney, R. P. Gunsalus, Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annu Rev Microbiol* **66**, 429-452 (2012).
13. M. J. McInerney, M. P. Bryant, N. Pfennig, Anaerobic bacterium that degrades fatty-acids in syntrophic association with methanogens. *Archives of Microbiology* **122**, 129-135 (1979).
14. R. E. Speece, *Anaerobic biotechnology for industrial wastewaters*. (Archae Press, Nashville, TN, 1996).
15. R. E. Speece, S. Boonyakitsombut, M. Kim, N. Azbar, P. Ursillo, Overview of anaerobic treatment: thermophilic and propionate implications. *Water Environ Res* **78**, 460-473 (2006).
16. R. I. Amann, W. Ludwig, K. H. Schleifer, Phylogenetic identification and *in-situ* detection of individual microbial-cells without cultivation. *Microbiol Rev* **59**, 143-169 (1995).
17. J. T. Staley, A. Konopka, Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* **39**, 321-346 (1985).
18. D. Riviere *et al.*, Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *ISME J* **3**, 700-714 (2009).
19. M. C. Nelson, M. Morrison, Z. Yu, A meta-analysis of the microbial diversity observed in anaerobic digesters. *Bioresource Technology* **102**, 3730-3739 (2011).
20. F. Sanger, S. Nicklen, A. R. Coulson, DNA sequencing with chain-terminating inhibitors. *P Natl Acad Sci U S A* **74**, 5463-5467 (1977).
21. T. Hunkapiller, R. J. Kaiser, B. F. Koop, L. Hood, Large-scale and automated DNA sequence determination. *Science* **254**, 59 (1991).
22. R. D. Fleischmann *et al.*, Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269**, 496 (1995).
23. A. M. Maxam, W. Gilbert, A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* **74**, 560-564 (1977).

24. S. M. D. Goldberg *et al.*, A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proceedings of the National Academy of Sciences* **103**, 11240-11245 (2006).
25. D. R. Bentley *et al.*, Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. *Nature* **456**, 53-59 (2008).
26. G. Turcatti, A. Romieu, M. Fedurco, A.-P. Tairi, A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res* **36**, e25-e25 (2008).
27. M. J. McInerney *et al.*, The genome of *Syntrophus aciditrophicus*: Life at the thermodynamic limit of microbial growth. *P Natl Acad Sci U S A* **104**, 7600-7605 (2007).
28. E. Pierce *et al.*, The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). *Environ. Microbiol.* **10**, 2550-2573 (2008).
29. T. Kosaka *et al.*, The genome of *Pelotomaculum thermopropionicum* reveals niche-associated evolution in anaerobic microbiota. *Genome Res* **18**, 442-448 (2008).
30. T. Kosaka *et al.*, Reconstruction and Regulation of the Central Catabolic Pathway in the Thermophilic Propionate-Oxidizing Syntroph Pelotomaculum thermopropionicum. *J Bacteriol* **188**, 202-210 (2006).
31. O. Chertkov *et al.*, Complete genome sequence of *Aminobacterium colombiense* type strain (ALA-1). *Standards in genomic sciences* **2**, 280-289 (2010).
32. D. Oehler *et al.*, Genome-guided analysis of physiological and morphological traits of the fermentative acetate oxidizer *Thermacetogenium phaeum*. *BMC Genomics* **13**, 723 (2012).
33. M. K. Nobu *et al.*, The genome of *Syntrophorhabdus aromaticivorans* strain UI provides new insights for syntrophic aromatic compound metabolism and electron flow. *Environ Microbiol* **17**, 4861-4872 (2015).
34. B. Muller, L. Sun, A. Schnurer, First insights into the syntrophic acetate-oxidizing bacteria--a genetic study. *MicrobiologyOpen* **2**, 35-53 (2013).
35. J. R. Sieber *et al.*, The genome of *Syntrophomonas wolfei*: new insights into syntrophic metabolism and biohydrogen production. *Environ Microbiol* **12**, 2289-2301 (2010).
36. A. Schmidt, N. Muller, B. Schink, D. Schleheck, A proteomic view at the biochemistry of syntrophic butyrate oxidation in *Syntrophomonas wolfei*. *PLoS ONE* **8**, e56905 (2013).
37. J. R. Sieber *et al.*, Proteomic analysis reveals metabolic and regulatory systems involved in the syntrophic and axenic lifestyle of Syntrophomonas wolfei. *Front Microbiol* **6**, 115 (2015).
38. J. Handelsman, Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and molecular biology reviews : MMBR* **68**, 669-685 (2004).
39. G. W. Tyson *et al.*, Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43 (2004).
40. J. C. Venter *et al.*, Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74 (2004).
41. A. C. McHardy, H. G. Martin, A. Tsirigos, P. Hugenholtz, I. Rigoutsos, Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods* **4**, 63-72 (2007).
42. J. T. Simpson *et al.*, ABySS: a parallel assembler for short read sequence data. *Genome Res* **19**, 1117-1123 (2009).
43. D. R. Zerbino, E. Birney, Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821-829 (2008).
44. R. Luo *et al.*, SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18 (2012).
45. T. Namiki, T. Hachiya, H. Tanaka, Y. Sakakibara, MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* **40**, e155 (2012).

46. Y. Peng, H. C. Leung, S. M. Yiu, F. Y. Chin, IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420-1428 (2012).
47. D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676 (2015).
48. S. Nurk, D. Meleshko, A. Korobeynikov, P. Pevzner, metaSPAdes: a new versatile de novo metagenomics assembler. *ArXiv e-prints* **1604**, arXiv:1604.03071 (2016).
49. S. Karlin, C. Burge, Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet* **11**, 283-290 (1995).
50. A. Pati, L. S. Heath, N. C. Kyrpides, N. Ivanova, ClaMS: A Classifier for Metagenomic Sequences. *Standards in genomic sciences* **5**, 248-253 (2011).
51. K. R. Patil, L. Roune, A. C. McHardy, The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. *PLoS One* **7**, e38581 (2012).
52. Y. W. Wu, Y. H. Tang, S. G. Tringe, B. A. Simmons, S. W. Singer, MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**, 26 (2014).
53. M. Albertsen *et al.*, Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**, 533-538 (2013).
54. M. Imelfort *et al.*, GroopM: an automated tool for the recovery of population genomes from related metagenomes. *PeerJ* **2**, e603 (2014).
55. D. D. Kang, J. Froula, R. Egan, Z. Wang, MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
56. Y.-W. Wu, B. A. Simmons, S. W. Singer, MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2016).
57. N. Segata, D. Bornigen, X. C. Morgan, C. Huttenhower, PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* **4**, 2304 (2013).
58. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055 (2015).
59. K. C. Wrighton *et al.*, Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**, 1661-1665 (2012).
60. T. Woyke *et al.*, Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950-955 (2006).
61. J. C. Robidart *et al.*, Metabolic versatility of the Riftia pachyptila endosymbiont revealed through metagenomics. *Environ Microbiol* **10**, 727-737 (2008).
62. A. Gardebrecht *et al.*, Physiological homogeneity among the endosymbionts of Riftia pachyptila and Tevnia jerichonana revealed by proteogenomics. *ISME J* **6**, 766-776 (2012).
63. L. P. Partida-Martinez, C. Hertweck, Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* **437**, 884-888 (2005).
64. E. Pelletier *et al.*, "*Candidatus* Cloacamonas acidaminovorans": genome sequence reconstruction provides a first glimpse of a new bacterial division. *J Bacteriol* **190**, 2572-2579 (2008).
65. C. T. Brown *et al.*, Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208-211 (2015).
66. P. N. Evans *et al.*, Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434-438 (2015).
67. I. Vanwonterghem *et al.*, Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat Microbiol* **1**, 16170 (2016).

68. S. D. Perkins, N. B. Scalfone, L. T. Angenent, Comparative 16S rRNA gene surveys of granular sludge from three upflow anaerobic bioreactors treating purified terephthalic acid (PTA) wastewater. *Water Sci Technol* **64**, 1406-1412 (2011).
69. J. H. Wu, W. T. Liu, I. C. Tseng, S. S. Cheng, Characterization of microbial consortia in a terephthalate-degrading anaerobic granular sludge system. *Microbiol-Uk* **147**, 373-382 (2001).
70. L. Cheng *et al.*, Progressive degradation of crude oil n-alkanes coupled to methane production under mesophilic and thermophilic conditions. *PLoS One* **9**, e113253 (2014).
71. OECD, in *OECD Factbook 2010: Economic, Environmental and Social Statistics*. (OECD Publishing, 2010), vol. dx.doi.org/10.1787/factbook-2010-64-en.
72. J. B. van Lier, N. Mahmoud, J. Zeeman, in *Biological Wastewater Treatment Principles: Modeling and Design,* M. C. M. v. L. M. Henze, G.A. Ekama, and D. Brdjanovic, Ed. (IWA Publishing, London, 2008).
73. F. von Wintzingerode, B. Selent, W. Hegemann, U. B. Gobel, Phylogenetic analysis of an anaerobic, trichlorobenzene transforming microbial consortium. *Appl Environ Microb* **65**, 283-286 (1999).
74. M. A. Dojka, P. Hugenholtz, S. K. Haack, N. R. Pace, Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl Environ Microb* **64**, 3869-3877 (1998).
75. R. Y. Liu *et al.*, Microbial diversity in the anaerobic tank of a full-scale produced water treatment plant. *Process Biochem* **45**, 744-751 (2010).
76. B. Schink, A. J. M. Stams, *Syntrophism among Prokaryotes*. M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer, E. Stackebrandt, Eds., Prokaryotes: A Handbook on the Biology of Bacteria, Vol 2, Third Edition: Ecophysiology and Biochemistry (Springer, New York, 2006), pp. 309-335.
77. A. Brauman, J. A. Muller, J. L. Garcia, A. Brune, B. Schink, Fermentative degradation of 3-hydroxybenzoate in pure culture by a novel strictly anaerobic bacterium, *Sporotomaculum hydroxybenoicum* gen. nov., sp. nov. *Int J Syst Bacteriol* **48**, 215-221 (1998).
78. Y. L. Qiu *et al.*, *Pelotomaculum terephthalicum* sp. nov. and *Pelotomaculum isophthalicum* sp. nov.: two anaerobic bacteria that degrade phthalate isomers in syntrophic association with hydrogenotrophic methanogens. *Archives of Microbiology* **185**, 172-182 (2006).
79. Y. L. Qiu *et al.*, *Sporotomaculum syntrophicum* sp. nov., a novel anaerobic, syntrophic benzoate-degrading bacterium isolated from methanogenic sludge treating wastewater from terephthalate manufacturing. *Archives of Microbiology* **179**, 457-457 (2003).
80. B. E. Jackson, V. K. Bhupathiraju, R. S. Tanner, C. R. Woese, M. J. McInerney, *Syntrophus aciditrophicus* sp. nov., a new anaerobic bacterium that degrades fatty acids and benzoate in syntrophic association with hydrogen-using microorganisms. *Archives of Microbiology* **171**, 107-114 (1999).
81. D. O. Mountfort, W. J. Brulla, L. R. Krumholz, M. P. Bryant, *Syntrophus buswellii* gen. nov., sp. nov.: a benzoate catabolizer from methanogenic ecosystems. *Int J Syst Bacteriol* **34**, 216-217 (1984).
82. Y. L. Qiu *et al.*, *Syntrophorhabdus aromaticivorans* gen. nov., sp nov., the first cultured anaerobe capable of degrading phenol to acetate in obligate syntrophic associations with a hydrogenotrophic methanogen. *Appl Environ Microb* **74**, 2051-2058 (2008).
83. H. Mouttaki, M. A. Nanny, M. J. McInerney, Metabolism of hydroxylated and fluorinated benzoates by *Syntrophus aciditrophicus* and detection of a fluorodiene metabolite. *Appl Environ Microbiol* **75**, 998-1004 (2009).
84. F. Peters, Y. Shinoda, M. J. McInerney, M. Boll, Cyclohexa-1,5-diene-1-carbonyl-coenzyme A (CoA) hydratases of *Geobacter metallireducens* and *Syntrophus aciditrophicus*: Evidence for a common benzoyl-CoA degradation pathway in facultative and strict anaerobes. *J Bacteriol* **189**, 1055-1060 (2007).

85. C. L. Chen *et al.*, Microbial community structure in a thermophilic anaerobic hybrid reactor degrading terephthalate. *Microbiology - SGM* **150**, 3429-3440 (2004).
86. C. L. Chen, J. H. Wu, W. T. Liu, Identification of important microbial populations in the mesophilic and thermophilic phenol-degrading methanogenic consortia. *Water Res* **42**, 1963-1976 (2008).
87. A. Lykidis *et al.*, Multiple syntrophic interactions in a terephthalate-degrading methanogenic consortium. *ISME J* **5**, 122-130 (2011).
88. Y. L. Qiu *et al.*, Identification and isolation of anaerobic, syntrophic phthalate isomer-degrading microbes from methanogenic sludges treating wastewater from terephthalate manufacturing. *Appl Environ Microb* **70**, 1617-1626 (2004).
89. M. K. Nobu *et al.*, Draft genome sequence of *Syntrophorhabdus aromaticivorans* strain UI, a mesophilic aromatic compound degrading syntroph. *Genome Announc.* **in press**, (2014).
90. J. W. Kung, J. Seifert, M. von Bergen, M. Boll, Cyclohexanecarboxyl-coenzyme A (CoA) and cyclohex-1-ene-1-carboxyl-CoA dehydrogenases, two enzymes involved in the fermentation of benzoate and crotonate in *Syntrophus aciditrophicus*. *J Bacteriol* **195**, 3193-3200 (2013).
91. S. Wischgoll *et al.*, Gene clusters involved in anaerobic benzoate degradation of *Geobacter metallireducens*. *Mol. Microbiol.* **58**, 1238-1252 (2005).
92. V. M. Markowitz *et al.*, The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic Acids Res* **38**, D382-D390 (2010).
93. M. Margulies *et al.*, Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380 (2005).
94. S. Bennett, Solexa Ltd. *Pharmacogenomics* **5**, 433-438 (2004).
95. C. Han, P. Chain, in *Proceedings of the 2006 international conference on bioinformatics & computational biology,* H. R. Arabnia, H. Valafar, Eds. (CSREA Press, Las Vegas, NV, 2006), pp. 141-146.
96. B. Ewing, P. Green, Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186-194 (1998).
97. B. Ewing, L. Hillier, M. C. Wendl, P. Green, Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**, 175-185 (1998).
98. D. Gordon, C. Abajian, P. Green, Consed: A graphical tool for sequence finishing. *Genome Res* **8**, 195-202 (1998).
99. J. Droge, A. C. McHardy, Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. *Briefings in bioinformatics* **13**, 646-655 (2012).
100. M. Goujon *et al.*, A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* **38**, W695-W699 (2010).
101. E. M. Zdobnov, R. Apweiler, InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847-848 (2001).
102. R. Adamczak, A. Porollo, J. Meller, Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins: Structure, Function, and Bioinformatics* **59**, 467-475 (2005).
103. I. Grissa, G. Vergnaud, C. Pourcel, The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**, 172 (2007).
104. J. Heider, G. Fuchs, Microbial anaerobic aromatic metabolism. *Anaerobe* **3**, 1-22 (1997).
105. K. Breese, G. Fuchs, 4-hydroxybenzoyl-CoA reductase (dehydroxylating) from the denitrifying bacterium *Thauera aromatica* - prosthetic groups, electron donor, and genes of a member of the molybdenum-flavin-iron-sulfur proteins. *Eur J Biochem* **251**, 916-923 (1998).
106. S. Schmeling *et al.*, Phenylphosphate synthase: a new phosphotransferase catalyzing the first step in anaerobic phenol metabolism in *Thauera aromatica*. *J Bacteriol* **186**, 8044-8057 (2004).
107. K. Schuhle, G. Fuchs, Phenylphosphate carboxylase: a new C-C lyase involved in anaerobic phenol metabolism in *Thauera aromatica*. *J Bacteriol* **186**, 4556-4567 (2004).

108. Z. He, J. Wiegel, Purification and characterization of an oxygen-sensitive reversible 4-hydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. *Eur J Biochem* **229**, 77-82 (1995).
109. X. Zhang, J. Wiegel, Reversible conversion of 4-hydroxybenzoate and phenol by *Clostridium hydroxybenzoicum*. *Appl Environ Microbiol* **60**, 4182-4185 (1994).
110. J. H. Wu *et al.*, Community and proteomic analysis of methanogenic consortia degrading terephthalate. *Appl Environ Microb* **79**, 105-112 (2013).
111. M. Boll, G. Fuchs, C. Meier, A. Trautwein, D. J. Lowe, EPR and Mössbauer studies of benzoyl-CoA reductase. *J Biol Chem* **275**, 31857-31868 (2000).
112. K. Breese, M. Boll, J. Alt-Morbe, H. Schagger, G. Fuchs, Genes coding for the benzoyl-CoA pathway of anaerobic aromatic metabolism in the bacterium *Thauera aromatica*. *Eur J Biochem* **256**, 148-154 (1998).
113. K. J. Gibson, J. Gibson, Potential early intermediates in anaerobic benzoate degradation by *Rhodopseudomonas palustris*. *Appl Environ Microbiol* **58**, 696-698 (1992).
114. M. Boll *et al.*, Nonaromatic products from anoxic conversion of benzoyl-CoA with benzoyl-CoA reductase and cyclohexa-1,5-diene-1-carbonyl-CoA hydratase. *J Biol Chem* **275**, 21889-21895 (2000).
115. J. W. Kung *et al.*, Identification and characterization of the tungsten-containing class of benzoyl-coenzyme A reductases. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17687-17692 (2009).
116. W. Buckel, R. K. Thauer, Energy conservation via electron bifurcating ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation. *Biochimica et biophysica acta* **1827**, 94-113 (2013).
117. R. K. Thauer, K. Jungermann, K. Decker, Energy-conservation in chemotrophic anaerobic bacteria. *Bacteriological Reviews* **41**, 100-180 (1977).
118. C. Wallrabenstein, B. Schink, Evidence of reversed electron-transport in syntrophic butyrate or benzoate oxidation by *Syntrophomonas wolfei* and *Syntrophus buswellii*. *Archives of Microbiology* **162**, 136-142 (1994).
119. K. Schuchmann, V. Muller, A bacterial electron-bifurcating hydrogenase. *J. Biol. Chem.* **287**, 31165-31171 (2012).
120. G. J. Schut, M. W. Adams, The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol* **191**, 4451-4457 (2009).
121. T. Ide, S. Baumer, U. Deppenmeier, Energy conservation by the H2:heterodisulfide oxidoreductase from *Methanosarcina mazei* Go1: identification of two proton-translocating segments. *J. Bacteriol.* **181**, 4076-4080 (1999).
122. A. K. Kaster, J. Moll, K. Parey, R. K. Thauer, Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2981-2986 (2011).
123. S. Wang, H. Huang, J. Kahnt, R. K. Thauer, A reversible electron-bifurcating ferredoxin- and NAD-dependent [FeFe]-hydrogenase (HydABC) in *Moorella thermoacetica*. *J. Bacteriol.* **195**, 1267-1275 (2013).
124. C. M. Plugge *et al.*, Complete genome sequence of *Syntrophobacter fumaroxidans* strain (MPOB(T)). *Standards in genomic sciences* **7**, 91-106 (2012).
125. H. S. Jeong, Y. Jouanneau, Enhanced nitrogenase activity in strains of *Rhodobacter capsulatus* that overexpress the rnf genes. *J Bacteriol* **182**, 1208-1214 (2000).
126. Y. Jouanneau, H. S. Jeong, N. Hugo, C. Meyer, J. C. Willison, Overexpression in *Escherichia coli* of the rnf genes from *Rhodobacter capsulatus* - characterization of two membrane-bound iron-sulfur proteins. *Eur J Biochem* **251**, 54-64 (1998).

127. B. Meyer *et al.*, Variation among *Desulfovibrio* species in electron transfer systems used for syntrophic growth. *J Bacteriol* **195**, 990-1004 (2013).
128. C. M. Plugge *et al.*, Global transcriptomics analysis of the *Desulfovibrio vulgaris* change from syntrophic growth with *Methanosarcina barkeri* to sulfidogenic metabolism. *Microbiol-Sgm* **156**, 2746-2756 (2010).
129. C. B. Walker *et al.*, The electron transfer system of syntrophically grown *Desulfovibrio vulgaris*. *J Bacteriol* **191**, 5793-5801 (2009).
130. F. Grein, A. R. Ramos, S. S. Venceslau, I. A. Pereira, Unifying concepts in anaerobic respiration: insights from dissimilatory sulfur metabolism. *Biochimica et biophysica acta* **1827**, 145-160 (2013).
131. C. A. Reddy, M. J. Wolin, M. P. Bryant, Characteristics of S-Organism isolated from *Methanobacillus omelianskii*. *J Bacteriol* **109**, 539-& (1972).
132. B. Schink, Energetics of syntrophic cooperation in methanogenic degradation. *Microbiology and Molecular Biology Reviews* **61**, 262-280 (1997).
133. B. Schink, A. J. M. Stams, in *The Prokaryotes,* E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt, F. Thompson, Eds. (Springer Berlin Heidelberg, 2013), chap. 59, pp. 471-493.
134. Y. Kamagata, Keys to Cultivating Uncultured Microbes: Elaborate Enrichment Strategies and Resuscitation of Dormant Cells. *Microbes Environ* **30**, 289-290 (2015).
135. T. Narihiro, Microbes in the Water Infrastructure: Underpinning Our Society. *Microbes Environ* **31**, 89-92 (2016).
136. Y. T. Liu, D. L. Balkwill, H. C. Aldrich, G. R. Drake, D. R. Boone, Characterization of the anaerobic propioate-degrading syntrophs *Smithella propionica* gen. nov., sp. nov. and *Syntrophobacter wolinii*. *Int J Syst Bacteriol* **49**, 545-556 (1999).
137. F. A. de Bok, A. J. Stams, C. Dijkema, D. R. Boone, Pathway of propionate oxidation by a syntrophic culture of *Smithella propionica* and *Methanospirillum hungatei*. *Appl Environ Microbiol* **67**, 1800-1804 (2001).
138. J. Dolfing, Syntrophic propionate oxidation via butyrate: a novel window of opportunity under methanogenic conditions. *Appl Environ Microbiol* **79**, 4515-4516 (2013).
139. M. S. Elshahed, V. K. Bhupathiraju, N. Q. Wofford, M. A. Nanny, M. J. McInerney, Metabolism of benzoate, cyclohex-1-ene carboxylate, and cyclohexane carboxylate by "*Syntrophus aciditrophicus*" strain SB in syntrophic association with H-2-using microorganisms. *Appl Environ Microb* **67**, 1728-1738 (2001).
140. J. G. Becker, G. Berardesco, B. E. Rittmann, D. A. Stahl, The role of syntrophic associations in sustaining anaerobic mineralization of chlorinated organic compounds. *Environ Health Perspect* **113**, 310-316 (2005).
141. Y. Gan, Q. Qiu, P. Liu, J. Rui, Y. Lu, Syntrophic oxidation of propionate in rice field soil at 15 and 30 degrees C under methanogenic conditions. *Appl Environ Microbiol* **78**, 4923-4932 (2012).
142. T. Lueders, B. Pommerenke, M. W. Friedrich, Stable-isotope probing of microorganisms thriving at thermodynamic limits: syntrophic propionate oxidation in flooded soil. *Appl Environ Microbiol* **70**, 5778-5786 (2004).
143. M. K. Nobu *et al.*, Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J* **9**, 1710-1722 (2015).
144. M. S. Elshahed, M. J. McInerney, Benzoate fermentation by the anaerobic bacterium *Syntrophus aciditrophicus* in the absence of hydrogen-using microorganisms. *Appl Environ Microbiol* **67**, 5520-5525 (2001).
145. I. Vanwonterghem *et al.*, Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *ISME J* **8**, 2015-2028 (2014).
146. S. D. Allison, J. B. Martiny, Colloquium paper: resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci U S A* **105 Suppl 1**, 11512-11519 (2008).

147. J. P. Guyot, H. Macarie, A. Noyola, Anaerobic digestion of a petrochemical wastewater using the UASB process. *Applied biochemistry and biotechnology* **24**, 579-589 (1990).
148. C. Fajardo, J. P. Guyot, H. Macarie, O. Monroy, Inhibition of anaerobic digestion by terephthalic acid and its aromatic by products. *Water Sci Technol* **36**, 83-90 (1997).
149. R. Kleerebezem, J. Beckers, L. W. Hulshoff Pol, G. Lettinga, High rate treatment of terephthalic acid production wastewater in a two-stage anaerobic bioreactor. *Biotechnol Bioeng* **91**, 169-179 (2005).
150. M. K. Nobu, T. Narihiro, K. Kuroda, R. Mei, W. T. Liu, Chasing the elusive Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. *ISME J* **10**, 2478-2487 (2016).
151. K. R. Patil *et al.*, Taxonomic metagenome sequence assignment with structured output models. *Nat Methods* **8**, 191-192 (2011).
152. M. K. Nobu *et al.*, Phylogeny and physiology of candidate phylum 'Atribacteria' (OP9/JS1) inferred from cultivation-independent genomics. *ISME J* **10**, 273-286 (2016).
153. T. Narihiro, M. K. Nobu, N. K. Kim, Y. Kamagata, W. T. Liu, The nexus of syntrophy-associated microbiota in anaerobic digestion revealed by long-term enrichment and community survey. *Environ Microbiol* **17**, 1707-1720 (2015).
154. Y. Uyeno, Y. Sekiguchi, A. Sunaga, H. Yoshida, Y. Kamagata, Sequence-specific cleavage of small-subunit (SSU) rRNA with oligonucleotides and RNase H: a rapid and simple approach to SSU rRNA-based quantitative detection of microorganisms. *Appl Environ Microb* **70**, 3650-3663 (2004).
155. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
156. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
157. K. Schuchmann, V. Muller, Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nature reviews. Microbiology* **12**, 809-821 (2014).
158. Y. Sekiguchi, Y. Kamagata, K. Nakamura, A. Ohashi, H. Harada, Fluorescence in situ hybridization using 16S rRNA-targeted oligonucleotides reveals localization of methanogens and selected uncultured bacteria in mesophilic and thermophilic sludge granules. *Appl Environ Microbiol* **65**, 1280-1288 (1999).
159. R. I. Amann, L. Krumholz, D. A. Stahl, Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental-studies in microbiology. *J Bacteriol* **172**, 762-770 (1990).
160. H. Daims, A. Bruhl, R. Amann, K. H. Schleifer, M. Wagner, The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**, 434-444 (1999).
161. S. Lucker *et al.*, Improved 16S rRNA-targeted probe set for analysis of sulfate-reducing bacteria by fluorescence in situ hybridization. *J Microbiol Methods* **69**, 523-528 (2007).
162. Y. Ueno, K. Yamada, N. Yoshida, S. Maruyama, Y. Isozaki, Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**, 516-519 (2006).
163. R. Mondav *et al.*, Discovery of a novel methanogen prevalent in thawing permafrost. *Nat Commun* **5**, 3212 (2014).
164. K. Paul, J. O. Nonoh, L. Mikulski, A. Brune, "Methanoplasmatales," Thermoplasmatales-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl Environ Microbiol* **78**, 8245-8253 (2012).
165. B. Dridi, M. L. Fardeau, B. Ollivier, D. Raoult, M. Drancourt, *Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *Int J Syst Evol Microbiol* **62**, 1902-1907 (2012).

166. T. Iino *et al.*, *Candidatus* Methanogranum caenicola: a novel methanogen from the anaerobic digested sludge, and proposal of *Methanomassiliicoccaceae* fam. nov. and *Methanomassiliicoccales* ord. nov., for a methanogenic lineage of the class *Thermoplasmata*. *Microbes Environ* **28**, 244-250 (2013).
167. P. Hugenholtz, Exploring prokaryotic diversity in the genomic era. *Genome Biol* **3**, REVIEWS0003 (2002).
168. R. Chouari *et al.*, Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. *Environ Microbiol* **7**, 1104-1115 (2005).
169. M. A. Dojka, P. Hugenholtz, S. K. Haack, N. R. Pace, Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl Environ Microbiol* **64**, 3869-3877 (1998).
170. D. Wilkins, X. Y. Lu, Z. Shen, J. Chen, P. K. Lee, Pyrosequencing of mcrA and archaeal 16S rRNA genes reveals diversity and substrate preferences of methanogen communities in anaerobic digesters. *Appl Environ Microbiol* **81**, 604-613 (2015).
171. Y. Saito *et al.*, Presence of a novel methanogenic archaeal lineage in anaerobic digesters inferred from *mcrA* and 16S rRNA gene phylogenetic analyses. *J Water Envir Tech* **13**, 279-289 (2015).
172. T. W. Cheng *et al.*, Metabolic stratification driven by surface and subsurface interactions in a terrestrial mud volcano. *ISME J* **6**, 2280-2290 (2012).
173. A. Dhillon *et al.*, Methanogen diversity evidenced by molecular characterization of methyl coenzyme M reductase A (mcrA) genes in hydrothermal sediments of the Guaymas Basin. *Appl Environ Microbiol* **71**, 4592-4601 (2005).
174. F. D. Ciccarelli *et al.*, Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283-1287 (2006).
175. R. Sorek *et al.*, Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**, 1449-1452 (2007).
176. A. Ruepp *et al.*, The genome sequence of the thermoacidophilic scavenger Thermoplasma acidophilum. *Nature* **407**, 508-513 (2000).
177. T. C. Tallant, J. A. Krzycki, Methylthiol:coenzyme M methyltransferase from *Methanosarcina barkeri*, an enzyme of methanogenesis from dimethylsulfide and methylmercaptopropionate. *J Bacteriol* **179**, 6902-6911 (1997).
178. H. Fu, W. W. Metcalf, Genetic basis for metabolism of methylated sulfur compounds in *Methanosarcina* species. *J Bacteriol* **197**, 1515-1524 (2015).
179. J. J. Moran, C. H. House, J. M. Vrentas, K. H. Freeman, Methyl sulfide production by a novel carbon monoxide metabolism in *Methanosarcina acetivorans*. *Appl Environ Microbiol* **74**, 540-542 (2008).
180. E. Oelgeschläger, M. Rother, In vivo role of three fused corrinoid/methyl transfer proteins in *Methanosarcina acetivorans*. *Mol Microbiol* **72**, 1260-1272 (2009).
181. A. Bose, G. Kulkarni, W. W. Metcalf, Regulation of putative methyl-sulphide methyltransferases in *Methanosarcina acetivorans* C2A. *Mol Microbiol* **74**, 227-238 (2009).
182. G. M. LeClerc, D. A. Grahame, Methylcobamide:coenzyme M methyltransferase isozymes from *Methanosarcina barkeri*. Physicochemical characterization, cloning, sequence analysis, and heterologous gene expression. *J Biol Chem* **271**, 18725-18731 (1996).
183. T. C. Tallant, L. Paul, J. A. Krzycki, The MtsA subunit of the methylthiol:coenzyme M methyltransferase of *Methanosarcina barkeri* catalyses both half-reactions of corrinoid-dependent dimethylsulfide: coenzyme M methyl transfer. *J Biol Chem* **276**, 4485-4493 (2001).
184. G. Borrel *et al.*, Comparative genomics highlights the unique biology of *Methanomassiliicoccales*, a *Thermoplasmatales*-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genomics* **15**, 679 (2014).

185. G. Kulkarni, D. M. Kridelbaugh, A. M. Guss, W. W. Metcalf, Hydrogen is a preferred intermediate in the energy-conserving electron transport chain of Methanosarcina barkeri. *P Natl Acad Sci U S A* **106**, 15915-15920 (2009).
186. I. A. Berg *et al.*, Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* **8**, 447-460 (2010).
187. T. Sato, H. Atomi, T. Imanaka, Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* **315**, 1003-1006 (2007).
188. W. B. Whitman, E. Ankwanda, R. S. Wolfe, Nutrition and carbon metabolism of Methanococcus voltae. *J Bacteriol* **149**, 852-863 (1982).
189. T. L. Miller, C. Lin, Description of *Methanobrevibacter gottschalkii* sp. nov., *Methanobrevibacter thaueri* sp. nov., *Methanobrevibacter* woesei sp. nov. and *Methanobrevibacter wolinii* sp. nov. *Int J Syst Evol Microbiol* **52**, 819-822 (2002).
190. T. L. Miller, M. J. Wolin, *Methanosphaera stadtmaniae* gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. *Arch Microbiol* **141**, 116-122 (1985).
191. R. S. Tanner, R. S. Wolfe, Nutritional requirements of *Methanomicrobium mobile*. *Appl Environ Microbiol* **54**, 625-628 (1988).
192. G. Zellner *et al.*, *Methanocorpusculaceae* fam. nov., represented by *Methanocorpusculum parvum*, *Methanocorpusculum sinense* spec. nov. and *Methanocorpusculum bavaricum* spec. nov. *Arch Microbiol* **151**, 381-390 (1989).
193. C. R. Staples *et al.*, Expression and association of group IV nitrogenase NifD and NifH homologs in the non-nitrogen-fixing archaeon *Methanocaldococcus jannaschii*. *J Bacteriol* **189**, 7392-7398 (2007).
194. H. Li, H. Xu, D. E. Graham, R. H. White, Glutathione synthetase homologs encode alpha-L-glutamate ligases for methanogenic coenzyme F420 and tetrahydrosarcinapterin biosyntheses. *P Natl Acad Sci U S A* **100**, 9785-9790 (2003).
195. D. E. Graham, H. Xu, R. H. White, Identification of coenzyme M biosynthetic phosphosulfolactate synthase: a new family of sulfonate-biosynthesizing enzymes. *J Biol Chem* **277**, 13421-13429 (2002).
196. M. Graupner, H. Xu, R. H. White, Identification of the gene encoding sulfopyruvate decarboxylase, an enzyme involved in biosynthesis of coenzyme M. *J Bacteriol* **182**, 4862-4867 (2000).
197. D. E. Graham, S. M. Taylor, R. Z. Wolf, S. C. Namboori, Convergent evolution of coenzyme M biosynthesis in the *Methanosarcinales*: cysteate synthase evolved from an ancestral threonine synthase. *Biochem J* **424**, 467-478 (2009).
198. G. Lettinga, Anaerobic digestion and wastewater treatment systems. *Antonie van Leeuwenhoek* **67**, 3-28 (1995).
199. N. E. Blair, W. D. Carter Jr, The carbon isotope biogeochemistry of acetate from a methanogenic marine sediment. *Geochim Cosmochim Ac* **56**, 1247-1258 (1992).
200. K. Finster, G. M. King, F. Bak, Formation of methylmercaptan and dimethylsulfide from methoxylated aromatic compounds in anoxic marine and fresh water sediments. *FEMS Microbiol Ecol* **74**, 295-301 (1990).
201. H. Kadota, Y. Ishida, Production of volatile sulfur compounds by microorganisms. *Annu Rev Microbiol* **26**, 127-138 (1972).
202. R. P. Kiene, M. E. Hines, Microbial formation of dimethyl sulfide in anoxic *Sphagnum* peat. *Appl Environ Microbiol* **61**, 2720-2726 (1995).
203. S. H. Zinder, T. D. Brock, Methane, carbon dioxide, and hydrogen sulfide production from the terminal methiol group of methionine by anaerobic lake sediments. *Appl Environ Microb* **35**, 344-352 (1978).
204. C. T. Brown, A. C. Howe, Q. Zhang, A. B. Pyrkosz, T. H. Brom, A reference-free algorithm for computational normalization of shotgun sequencing data. *arXiv*, 1203.4802 [q-bio.GN] (2012).

205. J. Pell *et al.*, Scaling metagenome sequence assembly with probabilistic de Bruijn graphs. *P Natl Acad Sci U S A* **109**, 13272-13277 (2012).
206. A. Bankevich *et al.*, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477 (2012).
207. D. Hyatt *et al.*, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
208. T. Seemann, Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069 (2014).
209. J. G. Caporaso *et al.*, QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335-336 (2010).
210. E. Pruesse, J. Peplies, F. O. Glockner, SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823-1829 (2012).
211. C. Quast *et al.*, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590-596 (2013).
212. D. McDonald *et al.*, An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610-618 (2012).
213. W. Ludwig *et al.*, ARB: a software environment for sequence data. *Nucleic Acids Res* **32**, 1363-1371 (2004).
214. K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596-1599 (2007).
215. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
216. V. Miele, S. Penel, L. Duret, Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* **12**, 116 (2011).
217. C. Rinke *et al.*, Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431-437 (2013).
218. H. Juottonen *et al.*, Methanogen communities and Bacteria along an ecohydrological gradient in a northern raised bog complex. *Environ Microbiol* **7**, 1547-1557 (2005).
219. D. R. Nemergut *et al.*, The effects of chronic nitrogen fertilization on alpine tundra soil microbial communities: implications for carbon and nitrogen cycling. *Environ Microbiol* **10**, 3093-3105 (2008).
220. C. Griebler, T. Lueders, Microbial biodiversity in groundwater ecosystems. *Freshwater Biology* **54**, 649-677 (2009).
221. Y. H. Kong, R. Teather, R. Forster, Composition, spatial distribution, and diversity of the bacterial communities in the rumen of cows fed different forages. *FEMS Microbiol Ecol* **74**, 612-622 (2010).
222. J. Glockner *et al.*, Phylogenetic diversity and metagenomics of candidate division OP3. *Environ Microbiol* **12**, 1218-1229 (2010).
223. M. Kim, M. Morrison, Z. Yu, Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiol Ecol* **76**, 49-63 (2011).
224. T. Narihiro, M. K. Nobu, N. K. Kim, Y. Kamagata, W. T. Liu, The nexus of syntrophy-associated microbiota in anaerobic digestion revealed by long-term enrichment and community survey. *Environ. Microbiol.*, (2014).
225. Y. Marcy *et al.*, Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *P Natl Acad Sci U S A* **104**, 11889-11894 (2007).
226. J. Frias-Lopez *et al.*, Microbial community gene expression in ocean surface waters. *P Natl Acad Sci U S A* **105**, 3805-3810 (2008).
227. V. Muller, F. Imkamp, E. Biegel, S. Schmidt, S. Dilling, Discovery of a ferredoxin:NAD+-oxidoreductase (Rnf) in *Acetobacterium woodii* - a novel potential coupling site in acetogens. *Incredible Anaerobes: From Physiology to Genomics to Fuels* **1125**, 137-146 (2008).

228. M. K. Nobu *et al.*, The genome of Syntrophorhabdus aromaticivorans strain UI provides new insights for syntrophic aromatic compound metabolism and electron flow. *Environ Microbiol*, (2014).
229. M. Strous, B. Kraft, R. Bisdorf, H. E. Tegetmeyer, The binning of metagenomic contigs for microbial physiology of mixed cultures. *Frontiers in microbiology* **3**, (2012).
230. M. Taylor, in *Astronomical Data Analysis Software and Systems XIV*. (2005), vol. 347, pp. 29.
231. T. Yamada, Y. Sekiguchi, Cultivation of uncultured *Chloroflexi* subphyla: significance and ecophysiology of formary uncultured *Chloroflexi* 'subphylum I' with natural and biotechnological relevance. *Microbes and Environments* **24**, 205-216 (2009).
232. R. Chouari *et al.*, Novel major bacterial candidate division within a municipal anaerobic sludge digester. *Appl Environ Microb* **71**, 2145-2153 (2005).
233. R. K. Thauer, A. K. Kaster, H. Seedorf, W. Buckel, R. Hedderich, Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* **6**, 579-591 (2008).
234. J. R. Sieber, H. M. Le, M. J. McInerney, The importance of hydrogen and formate transfer for syntrophic fatty, aromatic and alicyclic metabolism. *Environ Microbiol* **16**, 177-188 (2014).
235. S. Schmidt, E. Biegel, V. Muller, The ins and outs of Na(+) bioenergetics in *Acetobacterium woodii*. *Biochimica et biophysica acta* **1787**, 691-696 (2009).
236. H. Huang, S. Wang, J. Moll, R. K. Thauer, Electron bifurcation involved in the energy metabolism of the acetogenic bacterium *Moorella thermoacetica* growing on glucose or H2 plus CO2. *J Bacteriol* **194**, 3689-3699 (2012).
237. J. Hugenholtz, L. G. Ljungdahl, Electron transport and electrochemical proton gradient in membrane vesicles of *Clostridium thermoautotrophicum*. *J Bacteriol* **171**, 2873-2875 (1989).
238. P. L. Tremblay, T. Zhang, S. A. Dar, C. Leang, D. R. Lovley, The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating ferredoxin:NAD+ oxidoreductase essential for autotrophic growth. *MBio* **4**, e00406-00412 (2012).
239. M. Calusinska, T. Happe, B. Joris, A. Wilmotte, The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiol-Sgm* **156**, 1575-1588 (2010).
240. M. Balk, J. Weijma, A. J. Stams, *Thermotoga lettingae* sp. nov., a novel thermophilic, methanol-degrading bacterium isolated from a thermophilic anaerobic reactor. *Int J Syst Evol Microbiol* **52**, 1361-1368 (2002).
241. C. Matthies, B. Schink, Reciprocal isomerization of butyrate and isobutyrate by the strictly anaerobic bacterium strain WoG13 and methanogenic isobutyrate degradation by a defined triculture. *Appl Environ Microb* **58**, 1435-1439 (1992).
242. M. Stieb, B. Schink, Anaerobic degradation of isovalerate by a defined methanogenic coculture. *Archives of Microbiology* **144**, 291-295 (1986).
243. R. S. Conrad, L. K. Massey, J. R. Sokatch, D- and L-isoleucine metabolism and regulation of their pathways in *Pseudomonas putida*. *J Bacteriol* **118**, 103-111 (1974).
244. A. I. Qatibi, J. L. Cayol, J. L. Garcia, Glycerol and propanediols degradation by *Desulfovibrio alcoholovorans* in pure culture in the presence of sulfate, or in syntrophic association with *Methanospirillum hungatei*. *FEMS Microbiol Lett* **85**, 233-240 (1991).
245. S. Baena *et al.*, *Aminobacterium colombiense* gen. nov. sp. nov., an amino acid-degrading anaerobe isolated from anaerobic sludge. *Anaerobe* **4**, 241-250 (1998).
246. M. L. Fardeau, B. K. Patel, M. Magot, B. Ollivier, Utilization of serine, leucine, isoleucine, and valine by *Thermoanaerobacter brockii* in the presence of thiosulfate or *Methanobacterium* sp. as electron acceptors. *Anaerobe* **3**, 405-410 (1997).
247. C. M. Plugge, M. Balk, E. G. Zoetendal, A. J. Stams, Gelria glutamica gen. nov., sp. nov., a thermophilic, obligately syntrophic, glutamate-degrading anaerobe. *Int J Syst Evol Microbiol* **52**, 401-407 (2002).

248. F. X. Wildenauer, J. Winter, Fermentation of isoleucine and arginine by pure and syntrophic cultures of *Clostridium sporogenes*. *FEMS Microbiol Lett* **38**, 373-379 (1986).
249. S. Wang, H. Huang, J. Kahnt, R. K. Thauer, *Clostridium acidurici* electron-bifurcating formate dehydrogenase. *Appl Environ Microbiol* **79**, 6176-6179 (2013).
250. H. J. M. Harmsen *et al.*, *Syntrophobacter fumaroxidans* sp. nov., a syntrophic propionate-degrading sulfate-reducing bacterium. *Int J Syst Bacteriol* **48**, 1383-1387 (1998).
251. A.-E. Rotaru *et al.*, Interspecies electron transfer via hydrogen and formate rather than direct electrical connections in cocultures of *Pelobacter carbinolicus*. *Appl Environ Microb* **78**, 7645-7651 (2012).
252. H. C. Dubourguier, E. Samain, G. Prensier, G. Albagnac, Characterization of 2 strains of *Pelobacter carbinolicus* isolated from anaerobic digesters. *Archives of Microbiology* **145**, 248-253 (1986).
253. N. H. Youssef, M. S. Elshahed, Diversity rankings among bacterial lineages in soil. *ISME J* **3**, 305-313 (2009).
254. B. Fernandez-Gomez *et al.*, Ecology of marine Bacteroidetes: a comparative genomics approach. *ISME J* **7**, 1026-1037 (2013).
255. F. O. Glockner, B. M. Fuchs, R. Amann, Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Appl Environ Microbiol* **65**, 3721-3726 (1999).
256. H. M. Wexler, Bacteroides: the Good, the Bad, and the Nitty-Gritty. *Clinical Microbiology Reviews* **20**, 593-621 (2007).
257. Y. Xia, Y. Kong, P. H. Nielsen, In situ detection of protein-hydrolysing microorganisms in activated sludge. *FEMS Microbiol Ecol* **60**, 156-165 (2007).
258. L. Gomez-Consarnau *et al.*, Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445**, 210-213 (2007).
259. Z. L. Sabree, S. Kambhampati, N. A. Moran, Nitrogen recycling and nutritional provisioning by Blattabacterium, the cockroach endosymbiont. *Proc Natl Acad Sci U S A* **106**, 19521-19526 (2009).
260. R. E. Ley, P. J. Turnbaugh, S. Klein, J. I. Gordon, Microbial ecology: Human gut microbes associated with obesity. *Nature* **444**, 1022-1023 (2006).
261. A. Schwiertz *et al.*, Microbiota and SCFA in Lean and Overweight Healthy Subjects. *Obesity* **18**, 190-195 (2010).
262. N. A. Moran, P. Tran, N. M. Gerardo, Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. *Appl Environ Microbiol* **71**, 8802-8810 (2005).
263. M. Hess *et al.*, Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* **331**, 463-467 (2011).
264. F. Thomas, J.-H. Hehemann, E. Rebuffet, M. Czjzek, G. Michel, Environmental and Gut Bacteroidetes: The Food Connection. *Frontiers in Microbiology* **2**, 93 (2011).
265. S. C. Leahy *et al.*, Genome sequencing of rumen bacteria and archaea and its application to methane mitigation strategies. *Animal : an international journal of animal bioscience* **7 Suppl 2**, 235-243 (2013).
266. B. J. Baker, C. S. Lazar, A. P. Teske, G. J. Dick, Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* **3**, 14 (2015).
267. I. Brettar, R. Christen, M. G. Hofle, Aquiflexum balticum gen. nov., sp. nov., a novel marine bacterium of the Cytophaga-Flavobacterium-Bacteroides group isolated from surface water of the central Baltic Sea. *Int J Syst Evol Microbiol* **54**, 2335-2341 (2004).

268. V. B. Borisov, R. B. Gennis, J. Hemp, M. I. Verkhovsky, The cytochrome bd respiratory oxygen reductases. *Biochimica et biophysica acta* **1807**, 1398-1413 (2011).
269. P. Lu *et al.*, The cytochrome bd-type quinol oxidase is important for survival of Mycobacterium smegmatis under peroxide and antibiotic-induced stress. *Scientific Reports* **5**, 10333 (2015).
270. A. Das, R. Silaghi-Dumitrescu, L. G. Ljungdahl, D. M. Kurtz, Jr., Cytochrome bd oxidase, oxidative stress, and dioxygen tolerance of the strictly anaerobic bacterium Moorella thermoacetica. *J Bacteriol* **187**, 2020-2029 (2005).
271. A. D. Baughn, M. H. Malamy, The strict anaerobe Bacteroides fragilis grows in and benefits from nanomolar concentrations of oxygen. *Nature* **427**, 441-444 (2004).
272. F. A. de Bok, E. H. Roze, A. J. Stams, Hydrogenases and formate dehydrogenases of Syntrophobacter fumaroxidans. *Antonie van Leeuwenhoek* **81**, 283-291 (2002).
273. R. B. Canani *et al.*, Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. *World journal of gastroenterology : WJG* **17**, 1519-1528 (2011).
274. N. G. Kowlgi, L. Chhabra, D-lactic acidosis: an underrecognized complication of short bowel syndrome. *Gastroenterol Res Pract* **2015**, 476215 (2015).
275. J. R. Sheedy *et al.*, Increased d-lactic Acid intestinal bacteria in patients with chronic fatigue syndrome. *In Vivo* **23**, 621-628 (2009).
276. L. Sun *et al.*, Lentimicrobium saccharophilum gen. nov., sp. nov., a strictly anaerobic bacterium representing a new family in the phylum Bacteroidetes, and proposal of Lentimicrobiaceae fam. nov. *Int J Syst Evol Microbiol*, (2016).
277. P. L. Tremblay, D. R. Lovley, Role of the NiFe hydrogenase Hya in oxidative stress defense in Geobacter sulfurreducens. *J Bacteriol* **194**, 2248-2253 (2012).
278. A. L. Zbell, S. E. Maier, R. J. Maier, Salmonella enterica serovar Typhimurium NiFe uptake-type hydrogenases are differentially expressed in vivo. *Infection and immunity* **76**, 4445-4454 (2008).
279. E. A. Franzosa *et al.*, Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* **111**, E2329-2338 (2014).
280. H. Zhang *et al.*, Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci U S A* **106**, 2365-2370 (2009).
281. A. L. Zbell, R. J. Maier, Role of the Hya hydrogenase in recycling of anaerobically produced H2 in Salmonella enterica serovar Typhimurium. *Appl Environ Microbiol* **75**, 1456-1459 (2009).
282. V. M. Markowitz *et al.*, IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res* **40**, D123-D129 (2012).
283. A. Marchler-Bauer, S. H. Bryant, CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32**, W327-331 (2004).
284. A. Marchler-Bauer *et al.*, CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43**, D222-226 (2015).
285. T. N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785-786 (2011).
286. Y. Yin *et al.*, dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**, W445-451 (2012).
287. N. D. Rawlings, A. J. Barrett, R. Finn, Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **44**, D343-350 (2016).
288. M. Fischer, J. Pleiss, The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Res* **31**, 319-321 (2003).
289. R. Ihaka, R. Gentleman, R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**, 299-314 (1996).
290. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Use R! (Springer-Verlag, New York, 2009).

291. S. Lê, J. Josse, F. Husson, FactoMineR: An R Package for Multivariate Analysis. *2008* **25**, 18 (2008).
292. R. Mei, T. Narihiro, M. K. Nobu, W. T. Liu, Effects of heat shocks on microbial community structure and microbial activity of a methanogenic enrichment degrading benzoate. *Lett Appl Microbiol* **63**, 356-362 (2016).
293. V. K. Tyagi, S.-L. Lo, Sludge: A waste or renewable source for energy and resources recovery? *Renewable and Sustainable Energy Reviews* **25**, 708-728 (2013).
294. K. Kuroda *et al.*, A Single-Granule-Level Approach Reveals Ecological Heterogeneity in an Upflow Anaerobic Sludge Blanket Reactor. *PLoS One* **11**, e0167788 (2016).
295. S. Wischgoll *et al.*, Gene clusters involved in anaerobic benzoate degradation of *Geobacter metallireducens*. *Mol Microbiol* **58**, 1238-1252 (2005).
296. J. W. Kung *et al.*, Identification and characterization of the tungsten-containing class of benzoyl-coenzyme A reductases. *P Natl Acad Sci U S A* **106**, 17687-17692 (2009).
297. S. Bräuer, H. Cadillo-Quiroz, R. J. Ward, J. Yavitt, S. Zinder, *Methanoregula boonei* gen. nov., sp. nov., an acidiphilic methanogen isolated from an acidic peat bog. *Int J Syst Evol Micr* **61**, 45-52 (2010).
298. Y. Kamagata *et al.*, Characterization of three thermophilic strains of Methanothrix (Methanosaeta) thermophila sp. nov. and rejection of Methanothrix (Methanosaeta) thermoacetophila. *Int J Syst Bacteriol* **42**, 463-468 (1992).
299. D. D. Sommer, A. L. Delcher, S. L. Salzberg, M. Pop, Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* **8**, 64 (2007).
300. C. S. Miller, B. J. Baker, B. C. Thomas, S. W. Singer, J. F. Banfield, EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol.* **12**, R44 (2011).
301. J. G. Caporaso *et al.*, PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266-267 (2010).
302. L. Kall, A. Krogh, E. L. Sonnhammer, A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology* **338**, 1027-1036 (2004).
303. C. L. Dupont *et al.*, Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* **6**, 1186-1199 (2012).
304. A. Poehlein *et al.*, An ancient pathway combining carbon dioxide fixation with the generation and utilization of a sodium ion gradient for ATP synthesis. *PLoS ONE* **7**, e33439 (2012).
305. W. E. Balch, S. Schoberth, R. S. Tanner, R. S. Wolfe, *Acetobacterium*, a new genus of hydrogen-oxidizing, carbon dioxide-reducing, anaerobic bacteria. *Int J Syst Bacteriol* **27**, 355-361 (1977).
306. M. D. Kane, J. A. Breznak, *Acetonema longum* gen. nov. sp. nov., an H2/CO2 acetogenic bacterium from the termite, *Pterotermes occidentis*. *Arch Microbiol* **156**, 91-98 (1991).
307. W. H. Lorowitz, M. P. Bryant, *Peptostreptococcus productus* strain that grows rapidly with CO as the energy source. *Appl Environ Microbiol* **47**, 961-964 (1984).
308. F. E. Fontaine, W. H. Peterson, E. McCoy, M. J. Johnson, G. J. Ritter, A new type of glucose fermentation by *Clostridium thermoaceticum*. *J. Bacteriol.* **43**, 701-715 (1942).
309. F. Rieu-Lesme, B. Morvan, M. D. Collins, G. Fonty, A. Willems, A new H2/CO2-using acetogenic bacterium from the rumen: description of *Ruminococcus schinkii* sp. nov. *FEMS Microbiol Lett* **140**, 281-286 (1996).
310. J. M. Bruno-Barcena, M. S. Chinn, A. M. Grunden, Genome sequence of the autotrophic acetogen *Clostridium autoethanogenum* JA1-1 strain DSM 10061, a producer of ethanol from carbon monoxide. *Genome Announc* **1**, (2013).
311. J. S. Liou, D. L. Balkwill, G. R. Drake, R. S. Tanner, *Clostridium carboxidivorans* sp. nov., a solvent-producing clostridium isolated from an agricultural settling lagoon, and reclassification of

the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. *Int J Syst Evol Microbiol* **55**, 2085-2091 (2005).
312. A. Schnurer, B. Schink, B. H. Svensson, *Clostridium ultunense* sp. nov., a mesophilic bacterium oxidizing acetate in syntrophic association with a hydrogenotrophic methanogenic bacterium. *Int J Syst Evol Micr* **46**, 1145-1152 (1996).
313. H. Nagarajan *et al.*, Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of *Clostridium ljungdahlii*. *Microbial cell factories* **12**, 118 (2013).
314. R. S. Tanner, L. M. Miller, D. Yang, *Clostridium ljungdahlii* sp. nov., an acetogenic species in clostridial rRNA homology group I. *Int J Syst Bacteriol* **43**, 232-236 (1993).
315. B. Möller, R. Oßmer, B. H. Howard, G. Gottschalk, H. Hippe, *Sporomusa*, a new genus of gram-negative anaerobic bacteria including *Sporomusa sphaeroides* spec. nov. and *Sporomusa ovata* spec. nov. *Archives of Microbiology* **139**, 388-396 (1984).
316. A. Poehlein, G. Gottschalk, R. Daniel, First insights into the genome of the gram-negative, endospore-forming organism *Sporomusa ovata* strain H1 DSM 2662. *Genome Announc* **1**, (2013).
317. S. Manzoor, E. Bongcam-Rudloff, A. Schnurer, B. Muller, First genome sequence of a syntrophic acetate-oxidizing bacterium, *Tepidanaerobacter acetatoxydans* strain Re1. *Genome Announc* **1**, (2013).
318. M. Westerholm, S. Roos, A. Schnurer, *Tepidanaerobacter acetatoxydans* sp. nov., an anaerobic, syntrophic acetate-oxidizing bacterium isolated from two ammonium-enriched mesophilic methanogenic processes. *Syst Appl Microbiol* **34**, 260-266 (2011).
319. S. Hattori, Y. Kamagata, S. Hanada, H. Shoun, *Thermacetogenium phaeum* gen. nov., sp. nov., a strictly anaerobic, thermophilic, syntrophic acetate-oxidizing bacterium. *Int J Syst Evol Micr* **50**, 1601-1609 (2000).
320. J. A. Leigh, F. Mayer, R. S. Wolfe, *Acetogenium kivui*, a new thermophilic hydrogen-oxidizing acetogenic bacterium. *Archives of Microbiology* **129**, 275-280 (1981).
321. J. R. Graber, J. R. Leadbetter, J. A. Breznak, Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Appl Environ Microbiol* **70**, 1315-1320 (2004).
322. S. W. Ragsdale, Enzymology of the Wood-Ljungdahl pathway of acetogenesis. *Ann. N. Y. Acad. Sci.* **1125**, 129-136 (2008).
323. J. Mock, S. Wang, H. Huang, J. Kahnt, R. K. Thauer, Evidence for a hexaheteromeric methylenetetrahydrofolate reductase in *Moorella thermoacetica*. *J Bacteriol* **196**, 3303-3314 (2014).
324. G. Wohlfarth, G. Geerligs, G. Diekert, Purification and properties of a NADH-dependent 5,10-methylenetetrahydrofolate reductase from *Peptostreptococcus productus*. *Eur J Biochem* **192**, 411-417 (1990).
325. P. M. Vignais, A. Colbeau, Molecular biology of microbial hydrogenases. *Curr. Issues Mol. Biol.* **6**, 159-188 (2004).
326. S. Wang *et al.*, NADP-specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in *Clostridium autoethanogenum* grown on CO. *J. Bacteriol.* **195**, 4373-4386 (2013).
327. A. J. Shaw, D. A. Hogsett, L. R. Lynd, Identification of the [FeFe]-hydrogenase responsible for hydrogen generation in *Thermoanaerobacterium saccharolyticum* and demonstration of increased ethanol yield via hydrogenase knockout. *J Bacteriol* **191**, 6457-6464 (2009).
328. Y. Zheng, J. Kahnt, I. H. Kwon, R. I. Mackie, R. K. Thauer, Hydrogen formation and its regulation in *Ruminococcus albus*: involvement of an electron-bifurcating [FeFe]-hydrogenase, of a non electron-bifurcating [FeFe]-hydrogenase and of a putative hydrogen-sensing [FeFe]-hydrogenase. *J Bacteriol*, (2014).

329. G. Wohlfarth, G. Geerligs, G. Diekert, Purification and characterization of NADP(+)-dependent 5,10-methylenetetrahydrofolate dehydrogenase from *Peptostreptococcus productus* marburg. *J Bacteriol* **173**, 1414-1419 (1991).
330. S. W. Ragsdale, L. G. Ljungdahl, D. V. DerVartanian, Isolation of carbon monoxide dehydrogenase from Acetobacterium woodii and comparison of its properties with those of the Clostridium thermoaceticum enzyme. *J. Bacteriol.* **155**, 1224-1237 (1983).
331. G. Bruant, M. J. Levesque, C. Peter, S. R. Guiot, L. Masson, Genomic analysis of carbon monoxide utilization and butanol production by *Clostridium carboxidivorans* strain P7. *PLoS One* **5**, e13033 (2010).
332. H. Roh *et al.*, Complete genome sequence of a carbon monoxide-utilizing acetogen, *Eubacterium limosum* KIST612. *J Bacteriol* **193**, 307-308 (2011).
333. S. Hattori, A. S. Galushko, Y. Kamagata, B. Schink, Operation of the CO dehydrogenase/acetyl coenzyme A pathway in both acetate oxidation and acetate formation by the syntrophically acetate-oxidizing bacterium *Thermacetogenium phaeum*. *J Bacteriol* **187**, 3471-3476 (2005).
334. M. Hakobyan, H. Sargsyan, K. Bagramyan, Proton translocation coupled to formate oxidation in anaerobically grown fermenting *Escherichia coli*. *Biophysical chemistry* **115**, 55-61 (2005).
335. M. Stieb, B. Schink, Anaerobic oxidation of fatty acids by *Clostridium bryantii* sp. nov., a sporeforming, obligately syntrophic bacterium. *Archives of Microbiology* **140**, 387-390 (1985).
336. M. Stieb, B. Schink, Anaerobic degradation of isobutyrate by methanogenic enrichment cultures and by a *Desulfococcus multivorans* strain. *Archives of Microbiology* **151**, 126-132 (1989).
337. A. Jimeno, J. J. Bermudez, M. Canovasdiaz, A. Manjon, J. L. Iborra, Degradation of isovlaerate and 2-methylbutyrate in an anaerobic trickling filter. *Biological Wastes* **34**, 241-250 (1990).
338. R. S. Conrad, L. K. Massey, J. R. Sokatch, D-isoleucine and L-isoleucine metabolism and regulation of their pathways in *Pseudomonas putida*. *J. Bacteriol.* **118**, 103-111 (1974).
339. Y. Sekiguchi, Y. Kamagata, K. Nakamura, A. Ohashi, H. Harada, *Syntrophothermus lipocalidus* gen. nov., sp. nov., a novel thermophilic, syntrophic, fatty-acid-oxidizing anaerobe which utilizes isobutyrate. *Int J Syst Evol Micr* **50**, 771-779 (2000).
340. T. E. Hanson, F. R. Tabita, A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress. *P Natl Acad Sci U S A* **98**, 4397-4402 (2001).
341. H. Ashida *et al.*, A functional link between RuBisCO-like protein of *Bacillus* and photosynthetic RuBisCO. *Science* **302**, 286-290 (2003).
342. F. R. Tabita, S. Satagopan, T. E. Hanson, N. E. Kreel, S. S. Scott, Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J Exp Bot* **59**, 1515-1524 (2008).
343. L. ElAntak *et al.*, The cytochrome c3-[Fe]-hydrogenase electron-transfer complex: structural model by NMR restrained docking. *Febs Lett* **548**, 1-4 (2003).
344. A. Rodrigue, A. Chanal, K. Beck, M. Muller, L. F. Wu, Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial Tat pathway. *J Biol Chem* **274**, 13223-13228 (1999).