The Thesis Committee for John Daniel Carlton
Certifies that this is the approved version of the following Thesis:

Expansion of Armatimonadetes through marine sediment sequencing reveals three classes with unique ecological roles

APPROVED BY
SUPERVISING COMMITTEE:

Brett J. Baker, Supervisor

Jordan Casey

Kristin Nielsen

Expansion of Armatimonadetes through marine sediment sequencing reveals three classes with unique ecological roles

by

John Daniel Carlton

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Marine Science

The University of Texas at Austin

May 2023

Dedication

This thesis is dedicated to anyone and everyone who wants to make the world a better place however they can.

Acknowledgements

This thesis would not have been possible without the full support and expertise of my advisor, Dr. Brett Baker, and my mentor, Dr. Valerie De Anda. They have contributed countless hours to my education and development as a scientist, professional, and person over the last three and a half years. In particular, I thank Dr. Baker for taking a chance on me, first as an undergraduate student with an uncertain future. I would surely be on a different path without the grace and opportunity he extended me.

My research would not have been possible without the help of the other skilled scientists who guided me within Dr. Baker's research group. Several of their contributions to my project helped make it the success that it is. I am grateful to Dr. Xianzhe Gong and Dr. Mirna Vázquez-Rosas-Landa for advancing my knowledge and skills; Maggie Langwig and Dr. Kiley Seitz for helping to lay the foundation for this work before I started this project; and Emily Aguilar-Pine for being an excellent student and contributing beyond expectations of her in all things. Finally, I am grateful to all the other members of the Baker Lab who have given me knowledge, smiles, and a home ever since I joined the lab.

My family and friends who supported me throughout my degree program have made the last few years better than I imagined. My mom, dad, and brother have been instrumental in my development in the best ways and are a rock I can always lean on. My friends see me on the good and bad days, and have never left my side when I needed them. I know time and distance may weaken our bonds, but I will always hold you in my heart dearly. This thesis is research, but behind every section are the memories we made together. Thank you all.

Abstract

Expansion of Armatimonadetes through marine sediment sequencing reveals three classes with unique ecological roles

John Daniel Carlton, M.S. Marine Sci

The University of Texas at Austin, 2023

Supervisor: Brett J. Baker

Marine sediments comprise one of the largest environments on the planet, and their microbial inhabitants are significant players in global carbon and nutrient cycles. With the advent of improved sampling techniques, recent scientific studies have shown the complexity of these communities and identified novel microorganisms from the ocean floor. Here we add to our understanding of understudied microbes by obtaining 77 metagenome-assembled genomes from the bacterial phylum Armatimonadetes in the Guaymas Basin, Gulf of California, and the Bohai Sea, coastal China. Seven of these MAGs are not classified at the class level. Thus, we propose to name these organisms Zipacnadia. Searches of public databases revealed that the 77 Armatimonadetes described in this study (including Zipacnadia) are globally distributed in hypoxic and anoxic environments and are dominant members of deep-sea sediments (up to 1.95% of the GB metagenomic raw reads). The Armatimonadetes described here also have unique metabolic capabilities for this phylum. They have pathways to reduce CO_2 to acetate via the Wood-Ljungdahl pathway (WLP) and generate energy through the oxidative branch

of the Embden-Meyerhof-Parnas pathway using CO2 as an electron sink, maintaining the redox balance via WLP. Some of these organisms may also have an autotrophic lifestyle not previously identified in Armatimonadetes. Furthermore, these Armatimonadetes may play a role in sulfur and nitrogen cycling, using the intermediate compounds hydroxylamine and sulfite. The description of the Armatimonadetes identified in this study enhances our understanding of the diversity and metabolic potential of anoxic habitats worldwide.

Table of Contents

List of Figures

Introduction

Microorganisms outnumber other forms of life and drive biogeochemical cycling on the planet [1]. Ocean floor microbial communities are among the most complex on Earth [2] and play a major role in global carbon and nutrient cycling. However, microbial biodiversity in marine sediments is largely unknown despite their importance. This is due to the difficulty and expense of obtaining these samples and the challenges associated with replicating environmental conditions in a laboratory [3–5]. Metagenomics has provided insights into marine microbial communities by bypassing the need for culturing [6] and has transformed our understanding of biodiversity [1], microbial metabolism [7], and the evolution of life [8]. However, significant gaps remain in our understanding of ocean floor microbes as we have yet to characterize many dominant community members in these systems.

In marine sediments, two microbial guilds responsible for the terminal degradation of organic matter are strictly anaerobic acetogenic bacteria and methanogenic archaea. Both specialized microbial groups, respectively, reduce CO_2 to acetate and methane by the reductive acetyl-CoA pathway (also called the Wood–Ljungdahl pathway, WLP, in acetogenic bacteria) [9]. Acetogenesis and methanogenesis are processes linked to proton (H+) or sodium (Na+) ion pumps that drive ATP synthases in the membrane. The processes utilize H_2 as a major electron donor and CO_2 as the electron acceptor for energy conservation. In contrast to methanogenesis, acetogenic bacteria can obtain energy through substrate-level phosphorylation and autotrophy, depending on the other metabolic machinery present. Acetogenic bacteria can grow by converting one

carbon (C1) substrates (e.g., H_2-CO_2, CO, and formate) and fermentation substrates (e.g., methoxylated aromatic compounds, sugars, and amino acids, alcohols) to acetate. This metabolic versatility makes acetogenic bacteria an essential player in anaerobic food webs worldwide [10]. Acetogenesis is found in several archaeal lineages [7, 11, 12], but it has historically only been identified in two bacterial phyla, Firmicutes [13, 14] and Spirochaetes [15]. Recent metabolic reconstructions from environmental genomes indicate acetogenesis also occurs in Chloroflexi [16]. These recent studies highlight that acetogenesis is likely more widespread in the bacterial tree of life than previously thought.

Here, we evaluated the ocean floor microbial diversity of two contrasting marine sediment environments, the Guaymas Basin (GB) and the Bohai Sea (BS). GB is a geologically active region of hydrothermal vents located at a depth of approximately 2 000 meters in the Gulf of California. In GB, hydrothermal plumes provide an abundance of reduced electron donors for microbial growth, such as H_2, H_2S, Fe^{2+}, and NH_4^+ [17, 18]. GB has high sedimentation rates and organic-rich sediments, and thus can support diverse and active microbial communities [19]. BS is a shallower marine site (average of 18 meters deep) located along China's northern coast and is connected to several bays and the Yellow Sea. This region is characterized by anthropogenic inputs resulting in high levels of contaminants and eutrophication, primarily driven by agriculture and industry [20, 21].

Our metagenomic characterization of these marine sediment environments allowed us to identify 77 metagenome-assembled genomes (MAGs) belonging to the

Armatimonadetes phylum (previously known as Candidate division OP10). Armatimonadetes representatives were first discovered 20 years ago through 16S rRNA gene surveys in Yellowstone National Park [22, 23]. However, it was not until a decade later that these organisms were designated a new phylum [24]. Difficulties defining this phylum have been caused by a limited number of cultivated Armatimonadetes strains, making it challenging to characterize their metabolisms [25, 26] and determine their phylogenetic position. Armatimonadetes are known to be aerobic oligotrophs that degrade complex carbon compounds, and no strict anaerobic acetogenic members of this phylum have been described. Here, we characterize novel members within the Armatimonadetes that are potentially capable of performing acetogenesis, participating in nitrogen and sulfur cycling, and mediating key processes in the anaerobic carbon cycle in marine sediments.

Results

Reconstruction of Armatimonadetes Genomes From Marine Sediments

Sixty-nine MAGs (previously classified as a novel phylum [5]) were recovered from three hydrothermally impacted sediment cores from GB (described in Langwig-De Anda et al., 2021 [27]). In addition, eight MAGs were obtained from three sediment cores from BS in Bohai Bay and Midline sites (described in Gong et al., 2022 [28]). Based on the presence of single-copy marker genes inferred by CheckM v1.0.11 [29], these 77 MAGs have an average completeness of 77.35% and an average contamination of 3.59%. The MAGs range in size from 1.28 to 7.02 Mbp, with a median of 3.24 Mbp. The estimated average genome size is 3.9 Mbp. In addition, the MAGs have a high GC-content [30] ranging from 55-70% (Supplementary Table 1).

PHYLOGENETIC RELATIONSHIP AND TAXONOMIC AFFILIATION

Using several phylogenomic approaches (see methods, Phylogenetic analyses), we determined that the MAGs obtained in this study fall within the Armatimonadotes (Armatimonadota) phylum. GTDB-Tk v1.4.0 [31, 32] indicates that most GB and all BS MAGs are within classes UBA5377 and UBA10988, respectively, except for seven MAGs (2 BS, 5 GB) that do not fall within any known Armatimonadetes class (Supplementary Table 1). Because these MAGs were first identified in the Gulf of California/ Sea of Cortez, located in Mexico [5], we propose naming these seven class-level-unclassified MAGs "Zipacnadia" after the Mayan mythological figure Zipacna who personified the geological activity of the Earth's crust [33].

The 37-marker gene phylogenetic reconstruction, which includes 319 publicly available representatives closely related to Armatimonadetes (Firmicutes, Actinobacteria, Chloroflexi, Ca. *Eremiobacteraeota*) (Supplementary Table 2), confirms the taxonomic relationship of the 77 MAGs within the Armatimonadetes phylum but does not recover the class structure predicted by GTDB-Tk (Figure 1). Instead, Zipacnadia is monophyletic with the UBA10988 class. Similarly, the 16S rRNA gene phylogeny did not recover the GTDB-Tk-predicted class structure, where Zipacnadia is again placed within the same monophyletic group as UBA10988 16S rRNA gene sequences. The 16S rRNA gene phylogeny also revealed 34 unclassified, environmentally derived 16S rRNA genes recovered by past studies closely related to Zipacnadia, UBA10988, and UBA5377 (Figure 2). These unclassified, publicly available sequences are from globally distributed aquatic sediments in human-derived, freshwater, and marine systems (Figure 2),

suggesting these three understudied classes are globally distributed and may play important ecological roles outside of marine systems (Supplementary Table 8). A comparison of average amino acid identities (AAI) between our MAGs and 95 publicly available Armatimonadetes genomes (Supplementary Table 4) revealed that the Armatimonadetes MAGs from deep and coastal marine sediments are distinct from previously described Armatimonadetes and share up to 48.56% and 51.46% genome-wide amino acid similarity to one another respectively (Supplementary File 1 and Supplementary Table 4). When comparing the MAGs reconstructed here, the seven Zipacnadia MAGs share more similarity (58.7%) than the genomes within the two defined classes, UBA10988 and UBA5377 (49.3 and 49.5%, respectively). With the recovery of additional Armatimonadetes genomes, delineating the Zipacnadia MAGs into a novel class or several classes will likely become more evident. For this study, we followed the GTDB-Tk-predicted taxonomy in characterizing Zipacnadia as a novel class of the Armatimonadota phylum.

ECOLOGICAL SETTING

To estimate the abundance of the marine sediment MAGs (using an in-house pipeline MetaGaia, Supplementary Table 3), we mapped all the metagenomic reads against the genomic assemblies. This revealed that the relative abundance of MAGs from BS was low, averaging 0.0000487% across all sampling locations, though their abundance increased with depth (up to 10 times that of the shallowest samples in the same site). MAGs from GB sediments were obtained from two unique sites. First was

Megamat (Alvin cores Meg19 and Meg22 taken within close proximity to each other), an alkane-rich site named for the sizeable microbial mat discovered at this location [34]. Second was Aceto balsamico (AB), named after high acetate porewater concentrations at this site, reaching >800 µM. Megamat is low in methane (<1mM) and high in sulfate (~26mM), while AB has higher methane concentrations (5.8-8.8 mM) and low sulfate and sulfide (<1mM). The low abundance of our Armatimonadetes MAGs recovered from AB (0.4%), and their high abundance in Megamat (1%) suggests a preferential niche for sulfate-rich environments. In these ecosystems, methanogenesis may not be predominant or is restricted to substrates that cannot be metabolized by sulfate-reducing bacteria, such as methylamines [35].

METABOLIC INFERENCE

Several approaches were used to characterize the metabolic capabilities of the 77 Armatimonadetes described in this study (see methods). First, we clustered these bacteria based on their protein composition. Hierarchical clustering of the 77 MAGs based on the presence/absence profile of 17 930 protein domains from the Pfam v3.0 database identified five metabolic clusters (Supplementary Table 1). These clusters were consistent with the phylogenetic position of the marine sediment MAGs (Figure 1).

We also searched the MAGs for involvement in key biogeochemical processes using MEBS (Multigenomic Entropy Based Scores) [36]. MEBS searches for protein sequences involved in nitrogen, iron, oxygen, carbon (primarily methane-related), and sulfur cycling. The normalized entropy scores of the Armatimonadetes MAGs, as well as

319 publicly available references (Supplementary Table 2), were computed with a set of previously precomputed scores from 2,107 non-redundant genomes [36] (Supplementary Figure 1 and Supplementary Table 6). We used three projection methods and four clustering algorithms to analyze the consistency of the clusters. The non-supervised clustering from this analysis suggested that most of the MAGs from deep-sea sediments (58 UBA5377 and 4 Zipacnadia) are similar and contain fermentative anaerobic pathways (Supplementary Table 6). In contrast, all MAGs obtained from BS sediments (6 UBA10988 and 2 Zipacnadia) and some from GB (6 UBA5377 and 5 Zipacnadia) share protein content with organisms that generate energy by the oxidation or reduction of inorganic sulfur or nitrogen molecules (Supplementary Figure 2).

To further investigate the metabolic capacity of the novel Aratimonadetes genomes, we compared their predicted proteomes with a combination of functional databases and protein phylogenies. We manually reconstructed the metabolic pathways of the MAGs obtained in this study. These metabolic inferences agree with the MEBS clustering (Supplementary Figure 2) and suggest that the Armatimonadetes MAGs from marine sediment are mainly anaerobic acetogens with different energy production pathways.

Acetogenic Pathways to Conserve Energy

Most (48/77) MAGs code acetyl-CoA ligase (AcdAB) or both phosphate acetyltransferase (Pta) and acetate kinase (AckA) for acetate production (Figure 4 and Figure 5) and ATP generation. UBA5377 codes the two-step acetate formation pathway

with Pta and Ack (Figure 4), which is common in acetate-forming bacteria [37]. In contrast, UBA10988 and Zipacnadia MAGs (Figure 4) code AcdAB. This indicates these organisms are capable of the reversible one-step conversion of acetyl-CoA to acetate coupled to ATP synthesis and the use of acetate as a source of carbon and energy in the absence of other substrates. This is unique because AcdAB-mediated conversion of acetyl-CoA into acetate has mainly been found in Archaea [38]. These MAGs also encode an integral membrane multisubunit ferredoxin–NAD+ oxidoreductase, called the Rnf complex, that catalyzes the electron transfer from reduced ferredoxin (Fd^{2-}) to NAD+, generating a chemiosmotic gradient for H^+ or Na^+ [13].

Carbon Fixation and Energy Production

The WLP is at least partially present in all three classes of MAGs. Class UBA10988 codes a complete WLP (4/6 MAGs), while UBA5377 and Zipacnadia have only a few members that code a complete pathway (3/71) and several others that are almost complete (Figure 4). Though all UBA5377 and most (6/7) Zipacnadia do not code for fchA, this enzyme is unnecessary because of the bifunctional folD in these genomes (Supplementary Table 7). Like other acetogenic bacteria, these MAGs can likely oxidize a large variety of organic substrates (e.g., hexoses, pentoses, formate, and formaldehyde) and inorganic substrates such as hydrogen (H_2) or carbon monoxide (CO), which can be coupled to the reduction of CO_2.

Some acetogenic bacteria utilize the six-subunit Rnf complex for energy conservation [13]. This complex reduces one NAD+ to NADH and simultaneously

moves protons or sodium across the cell membrane to generate ATP through ATP synthase. In the UBA10988 MAGs, 2/6 code all six subunits of the Rnf complex, while the remaining 4/6 UBA10988 code at least four subunits. In contrast, Zipacnadia MAGs encode an incomplete Rnf complex, where no MAGs code all six subunits. The presence of an Rnf complex is the key indicator of autotrophy in UBA10988, suggesting these organisms can grow autotrophically via acetogenesis using the WLP and Rnf complex (Figure 5). This metabolic structure resembles the previously described autotrophic acetogen isolate *Clostridium ljungdahli-i* [13]. In contrast, organisms without an Rnf complex, like UBA5377 and possibly Zipacnadia, are not capable of autotrophic growth and must rely on heterotrophy to drive acetogenesis, usually by consuming glucose [39].

In autotrophic acetogenic bacteria, electrons from H_2 and CO_2 are derived from hydrogen oxidation, catalyzed by hydrogenases such as HydABCD [13]. After searching Armatimonadetes MAGs for potential hydrogenases, we identified NiFe group 4b, 1a, 4g, and 3b hydrogenase. NiFe group 4b hydrogenase can be associated with formate dehydrogenase, carbon monoxide dehydrogenase, and glutamate synthase, which can act as electron-input sources. We identified all of these associated complexes in all three classes (Figures 4 and 5) (Supplementary Table 7), suggesting these organisms can pair diverse electron inputs to their hydrogenase [40]. NiFe group 1a hydrogenase is thought to pair H_2 oxidation with sulfate, metal, or organohalide reduction, and is encoded in all six UBA10988 MAGs, 5/7 Zipacnadia, and 5/64 UBA5377 [40]. NiFe group 4g hydrogenase was identified in 16/64 UBA5377 MAGs, and this group is predicted to utilize reduced ferredoxins from the TCA cycle to synthesize hydrogen and translocate

protons [40]. These NiFe group 4g hydrogenases are phylogenetically closely related to NiFe group 4e (Supplementary Figure 5, red stars). NiFe group 4e hydrogenases are sometimes associated with the Ech complex, an alternative autotrophic acetogenesis mechanism similar to the Rnf complex. However, we found no evidence of an Ech complex in the 77 MAGs described in this study. Finally, UBA10988 and UBA5377 uniquely code for NiFe group 3b hydrogenase, which reversibly couples the oxidation of NADPH to the fermentative evolution of H_2.

A Nuo complex, better known as complex I in the electron transport chain, is partially coded in 70/77 Armatimonadetes MAGs (Nuo, Supplementary Table 7)—no MAGs code for a complete Nuo complex. If the Nuo complex is present in these organisms, it would provide an alternative pathway for creating a chemiosmotic gradient that could be utilized by an F-type ATP synthase identified in the MAGs (25/77 complete, 52/77 partial, Supplementary Table 7).

Central Carbon Metabolism

The Armatimonadetes MAGs recovered here are predicted to have largely incomplete TCA cycles that closely resemble other acetate-producing bacteria [41]. Malate dehydrogenase (35/77) replaces malate:quinone oxidoreductase in the Armatimonadetes MAGs, and citrate (Re)-synthase (15/77) is more common than citrate synthase (5/77). The presence of citrate (Re)-synthase in these organisms provides evidence for an anaerobic lifestyle because it is inactivated by oxygen [42]. This contrasts citrate synthase, which functions in comparably oxygen-rich aerobic organisms.

UBA5377 largely lacks either citrate synthase (4/64) or citrate (Re)-synthase (5/64), while UBA10988 and Zipacnadia both lack malate dehydrogenase. Although the TCA cycle is incomplete, reducing power from the TCA cycle can still be generated in the Armatimonadetes MAGs using isocitrate dehydrogenase (IDH), 2-oxoglutarate:ferredoxin oxidoreductase (Kor), succinate dehydrogenase (Sdh), and malate dehydrogenase (Mdh) (Figure 4). Most MAGs recovered in this study also have incomplete pathways for the Embden–Meyerhof–Parnas (EMP), pentose phosphate pathway (PPP), and reverse ribulose monophosphate pathway (Figure 4, Supplementary Table Table 7). Over half (42/77) code complete Pyruvate:ferredoxin oxidoreductase (Por), which converts acetyl-CoA to pyruvate, potentially linking the WLP of autotrophic CO_2 fixation to the TCA cycle. These results suggest that when growing heterotrophically, Armatimonadetes (especially UBA5377, which lacks autotrophic machinery) may degrade sugars into pyruvate that can be further oxidized to acetate.

Chemolithotrophy

The presence of anaerobic sulfite reductase (Asr) in most of the MAGs (65 of 77) indicates that these organisms may be capable of reducing sulfite to hydrogen sulfide [43]. The annotated domain from the putative Asr (Supplementary Table 7) suggests this complex may also act on other compounds, such as hydroxylamine (NH_2OH) or selenium trioxide (SeO_3) (Figure 5) [43]. UBA10988 also codes HydABGD, a sulfhydrogenase complex that reduces polysulfide to hydrogen sulfide (H_2S).

Several MAGs contain genes for hydroxylamine-utilizing enzymes (Figure 4). Half of UBA10988 MAGs (3/6) and one Zipacnadia MAG code hydroxylamine dehydrogenase (Hao). However, protein phylogenies suggest these hao genes are more closely related to cytochrome c552 (Hao-like), which is an electron transporter involved in dissimilatory nitrite reduction to ammonium (Supplementary Figure 6) [44]. This is in contrast to cytochrome c554 (true Hao), which would reduce hydroxylamine (NH_2OH) to nitrite (NO_2^-) or nitrous oxide (NO). Following these results, UBA10988 and Zipacnadia do not appear to encode true Hao. Armatimonadetes MAGs also code hydroxylamine dehydrogenase (Hcp) (6/7 Zipacnadia, 3/6 UBA10988, 21/64 UBA5377), which reduces hydroxylamine to ammonium. The phylogeny of Hcp proteins (Supplementary Figure 7) supports their original annotation as hydroxylamine dehydrogenase. However, we could not identify complete ammonia oxidation or nitrate/ite reduction pathways; thus, the greater function of this gene needs to be clarified [45].

Degradation of Carbohydrates and Proteins

Our analysis of peptidases and carbohydrate-active enzymes (CAZYmes) revealed several metabolic strategies for degrading complex molecules. We detected at least one extracellular peptidase in 7 UBA10988, 57 UBA5377, and 3 Zipacnadia MAGs. All identified extracellular peptidases are shown in Supplementary Figure 10, and a comprehensive list of all detected peptidases is in Supplementary Table 10. Genes for extracellular subfamily S8A protease were common across each class, suggesting a subtilisin-like protease is excreted by these organisms to degrade polypeptides [46].

MAGs in UBA5377 encode S8B family peptidases, suggesting they produce a kexin-like protease for alternate polypeptide degradation through Lys-Arg and Arg-Arg cleavage [46]. Family C40 peptidases are present in 5/6 UBA10988 and 6/7 Zipacnadia MAGs, which play a role in cell wall component degradation [46].

CAZYmes (carbohydrate-active enzymes) for complex carbohydrate degradation were annotated using three methods and were only considered present if they were confirmed by at least two (see methods). Half of the MAGs (45/77) code CAZYmes, predicted to degrade carbohydrates such as cellulose, chitin, starch, xylan, mannan, pectin, and laminarin (Supplementary Table and Figure 9). UBA10988 and Zipacnadia have partial cellulose degradation pathways. Both classes code endoglucanase (Supplementary Table 7), capable of breaking cellulose to cellodextrin, and the cytoplasmic membrane-bound CAZyme family GH94 to further degrade cellodextrin to glucose. Most MAGs encode chitin-degrading machinery. For example, all three classes code family GH23, suggesting these organisms can initially degrade chitin to chitodextrin (Supplementary Table 9). Inside the cytoplasm, degradation of diacetylchitobiose to glucosamine may be carried out by hexosaminidase based on the presence of *HEX/HEXA_B/nagZ* (45/77 MAGs).

Extracellular CAZYmes were detected in 4 UBA10988, 39 UBA5377, and 2 Zipacnadia MAGs (Supplementary Table 9). Extracellular hydrolyzing CAZymes were uncommon, with only a few sequences for families GH62, GH121, and GH136. Interestingly, we only identified one extracellular CAZYme in UBA5377, GH62, indicating these organisms may degrade arabinofuranosyl to arabinofuranose [47] and

supply arabinofuranose to fermenters in the deep sea (Figure 6). Family GH141 is present in UBA10988 and UBA5377, which allows for the degradation of Lacto-N-tetraose sugars to lactose (Figure 5) [47]. Family GH136 is present in UBA10988 and Zipacnadia, and degrades xyloglucan to smaller oligosaccharides (Figure 5) [47].

Arsenate Detoxification

Due to its similarity to phosphate, arsenate can freely enter cells through phosphate transporters [48]. A minority of MAGs in each class possess a detoxification pathway to prevent the buildup of arsenic compounds in their cells. Arsenate within the cell could be transformed into arsenite by an arsenate reductase (ArsC) coded by 16/77 MAGs (Figure 4). Arsenite can be removed from the cell by an arsenical pump membrane protein (ArsB), present in 21/77 MAGs. A transcriptional regulator responsive to arsenate/arsenite (ArsR) was also found in 53/77 MAGs, likely regulating the expression of the detoxification proteins.

Environmental Interactions and Motility

MAGs from BS and GB differ in their ability to interact with other community members. In BS MAGs, we identified a type II bacterial secretion system that participates in biofilm formation. These coastal sediment MAGs code genes for polysaccharide biosynthesis proteins PslF, PslG, and PslH, which support biofilm matrix production. They also code for the adhesion factor PgaC/IcaA, which participates in exopolysaccharide biosynthesis [49–51]. Some GB MAGs (27/69) code PslG, which may help to disperse existing bacteria from biofilms [52]. We identified a type III bacterial

secretion system in GB MAGs, identified by the presence of other flagellar structures such as an L ring, P ring, M ring, hook biosynthesis, filament biosynthesis, and other minor basal body components (Supplementary Table 5). Furthermore, the GB MAGs in this study encode motor proteins MotA (40/69) and MotB (37/69), as well as additional genes involved in chemotaxis signal identification and response (Supplementary Table 5). Flagellar structures are common in other deep-sea bacteria, used to avoid temperature and pressure stress [53]. Similarly to previously described Armatimonadetes, the MAGs described in this study are likely gram-negative, encoding genes for lipopolysaccharide biosynthetic enzymes, CMP-KDO synthetase, KDO 8-P synthase, 3-deoxy-D-manno-octulosonic-acid transferase, and glucosamine N-acyltransferases (KdsABCD, KdtA, LpxABDL) (Supplementary Table 5).

Discussion

In this study, we broadly expand the genomic diversity of Armatimonadetes bacteria. We describe three classes, UBA10988, UBA5377, and Zipacnadia, from coastal and deep-sea sediments. UBA10988 and UBA5377 were previously identified, but undescribed [31], and Zipacnadia is identified for the first time in this study. Prior to this work, Armatimonadetes bacteria were known to be a primarily aerobic, heterotrophic phylum [25, 54]. The MAGs recovered here expand the diversity of Armatimonadetes to include likely obligate anaerobes with chemolithotrophic potential, some of which are capable of autotrophic growth. These new Armatimonadetes members are broadly distributed worldwide and abundant in deep-sea hydrothermal environments. UBA5377

genomes are among the most dominant microbial populations in GB sediments (Supplementary Table 3). They are particularly abundant at intermediate depths of anoxic sediments with methane concentrations less than 1mM, CO_2 ~10mM, temperatures >30℃, and high sulfate concentrations ~26 mM, similar to the average sulfate concentration in the ocean (28 mM). Autotrophic acetogenesis may lead these organisms to compete with methanogenic archaea for hydrogen (H_2) and carbon dioxide (CO_2). The large number of UBA5377 MAGs in Megamat, paired with low methane levels, may suggest that these acetogenic organisms are outcompeting methanogens and thus play a vital role in the final degradation of organic matter in the deep sea. In addition, their flagellar motility may contribute to ecological success in the deep sea, allowing these organisms to navigate the varied microcosms of an active sea floor. UBA5377 and Zipacnadia MAGs are dominant in the deep sea, making up about 1% of the total microbial diversity in GB, suggesting that understanding their roles is crucial to interpreting the ecology of the ocean floor.

Based on our detailed genomic characterization, the recovered Armatimonadetes appear to be acetogenic and fermentative obligate anaerobes. Amino acid comparisons suggest these MAGs comprise five distinct metabolic clusters (Figure 1). However, phylogenetic analyses indicate these organisms are members of three classes, UBA5377, UBA10988, and Zipacnadia. Our detailed metabolic annotations agree with the three-class view, where the three classes have several key differences in pathways likely to be essential to these bacteria. UBA10988 MAGs are defined by autotrophic growth through the WLP and an Rnf complex that bolsters the proton gradient exploited by ATP

synthase. Zipacnadia may also share this capability, as it partially codes for the WLP and Rnf complex, but neither was found complete in this class. Conversely, UBA5377 is defined by the total lack of an Rnf complex, necessitating a heterotrophic lifestyle more reliant on the class's hydrogenases and complex I-like structure to generate a proton gradient. All three classes are likely able to transform inorganic carbon (CO_2) into acetate, making a carbon pool accessible to other organisms (Figure 6). Acetate produced by Zipacnadia can become a carbon source, electron donor, or other chemical intermediates for community members.

All Armatimonadetes recovered here are predicted to utilize nitrogen and sulfur compounds (Figure 5). An anaerobic sulfite reductase (Asr) is present in the MAGs, suggesting they play an active role in sulfur cycling. In addition, these MAGs can reduce hydroxylamine (NH_2OH), an important intermediate in the nitrogen cycle, to ammonium (NH_4^+). Hydroxylamine is formed during nitrification and anaerobic ammonium oxidation, and is a precursor of nitrous oxide [55]. However, microbial hydroxylamine utilization needs to be studied more, and we know little about its activity in the environment currently.

Given the breadth of environments where Armatimonadetes were identified (16S rRNA gene analysis), it is likely that the MAGs recovered here are important players in anoxic environments globally. These marine sediment Armatimonadetes appear to fulfill a broad range of ecological roles, vastly expanding the previously known capabilities of Armatimonadetes, including complex carbohydrate degradation, carbon fixation, nitrogen reduction, and sulfur reduction.

Methods

Sample Collection

Guaymas Basin (GB) sediments samples (Aceto Balsamico, Megamat19, and Megamat22) were collected during Alvin submarine dives from sediments in the Gulf of California, Mexico, (27°N 0.388, 111°W 24.560) for more details see Langwig, et al. (2021) and Castelle, et al. (2021). Chinese sediment samples were collected from three sites in the Bohai Sea (BS) (BHB10: 38°45.00N, 118°9.12E; M3: 38°40.03N, 119°32.51E; and M8: 39°41.34N, 120°38.98E) during a cruise with the R/V Chuangxin Yi to Bohai Sea in August, 2018 [56]. These sediment samples were collected using a box-sampler, and a polyvinyl chloride (PVC) tube with 11 cm internal diameter was inserted into the box-sampler after carefully removing top water to take sediment-core samples. Sub-samples were taken through pre-drilled side-holes with intervals of 2 cm, and frozen at -80°C.

Genome Sequencing and Assembly

Genome sequencing and assembly of GB samples was carried out as described in Langwig, et al. 2021. Briefly, whole community DNA from ≥ 10 g of sediment was extracted using the DNeasy PowerSoil kit (Qiagen, Germantown, Maryland, USA) following the manufacturer's instructions. The DNA concentrations were quantified with a QUBIT 2.0 fluorometer (Thermo-Fisher, Singapore). Illumina HiSeq 4000 genome sequencing for Guaymas Basin samples was conducted at the Michigan State University RTSF Genomics Sequences, were trimmed and filtered using Sickle v1.33, and assembly

was performed using IDBA-UD v1.0.9 More details were described by Langwig et al., 2022 [27].

Genome sequencing and assembly for the BS samples is described previously in Gong et al., 2022 [28]. Whole community DNA from ≥ 1 g of sediment was extracted using the DNeasy PowerSoil kit (Qiagen, Germantown, Maryland, USA). DNA concentration was measured using the Qubit® dsDNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). OD values between 1.8~2.0 and DNA contents above 1μg were used to construct the library. DNA from the BS sample was sequenced with an Illumina HiSeq X™ Ten platform at Tianjin Novogene Bioinformatic Technology Co., Ltd (Tianjin, China). A total of 1μg DNA per sample was used as input material for the DNA preparations. Sequencing libraries were generated using NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NE, USA) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Briefly, the DNA sample was fragmented by sonication to a size of 350 bp, then DNA fragments were end-polished, A-tailed, and ligated with the full-length adaptor for Illumina sequencing with further PCR amplification. Finally, PCR products were purified (AMPure XP system), and libraries were analyzed for size distribution by Agilent2100 Bioanalyzer and quantified using real-time PCR before sequencing. BS-derived sequences were trimmed and quality controlled using Sickle v1.33, and assembly was performed using IDBA-UD v1.1.3.

RELATIVE ABUNDANCE

MAG relative abundance was calculated using MetaGaia (https://github.com/valdeanda/MetaGaia). Briefly, we used the files_prep.py script to link the depth file for each scaffold in the assembly, the size of each raw reads file, the length of each bin, and the sampling site of each bin. The output files were reformatted, and the abundance was calculated using the bin_abundancy.py script with the parameter: -n 10^9.

GENOME BINNING

Binning of individual GB assemblies, only scaffolds > 2000 bp, was performed from dives using Concoct v.0.4.0 [57] and Metabat v2.12.1 [58]. Concoct was used with default settings, and Metabat was run with the following parameters: –minCVSum 0 --saveCls -d -v --minCV 0.1 -m 2000. Results from these two binning tools were combined using DAS Tool v1.0 using default settings. CheckM v1.0.11 was used to determine MAG completeness and contamination. Genomes were only analyzed further if they were more than 50% complete and less than 10% contamination; 69 MAGs were identified as a potential novel group CP9 in a previous analysis [5] and included in this research.

The genome binning procedures for BS samples were similar to the GB samples. Scaffolds under 2000 bp were removed. Binning was carried out by DAS Tool, Concoct v0.4.0 [57] Metabat v2.12.1 [58], and MaxBin v2.2.7 [59]. The first three binning tools used the same settings as for GB samples. MaxBin v2.2.7 was run with default settings. CheckM v1.1.2 was used to determine MAG completeness and contamination. Genomes were only analyzed further if they were more than 50% complete and contained less than 10% contamination. 8 MAGs were identified as candidates within CP9 and included in

further analyses. In total, 77 MAGs were obtained from deep-sea GB and BS coastal sediments. Genome statistics are shown in Supplementary Table 1. The estimated genome size was calculated by solving the ratio of the assembled MAGs' size over the estimated completeness from CheckM to a theoretical complete genome.

PHYLOGENETIC ANALYSES

A set of 318 publicly available genomes were downloaded from NCBI to better resolve the phylogenetic relationships of the MAGs described in this study (Supplementary Table 2). Phylogenetic markers were then extracted from the GB, BS, and reference genomes using phylosift (v1.0.1) [60], with the 'phylosift search' and 'phylosift align' functions. The genomes were aligned using Geneious Prime v11.0.4+11 using MUSCLE v5.1 and MAFFT v7.490 using default settings, and were then masked (with a gap of 50%). A phylogenetic tree was generated through a maximum likeliness-based approach using IQTree v2.0.3 with 1 000 bootstrapping replicates using model LG+F+R10 [61]. The tree was visualized using the Interactive Tree of Life (iToL v5). Barrnap v0.7 [62] was used to extract 16S rRNA gene sequences from the MAGs in this study. These sequences were then compared to known 16S rRNA genes by using BLASTn [63], against the Silva database release 132 [64]. The 16S rRNA gene tree was created using RAxML v7.0.3 with standard parameters in the ARB software package [65]. The amino acid identity (AAI) profile was generated using the CompareM v0.1.1 option aai_wf [66]. MAGs were also classified using GTDB-Tk v1.4.0 with database version r95 [31] (Supplementary Table 1).

HIERARCHICAL CLUSTERING OF GENOMES BASED ON PROTEIN DOMAINS

Two unsupervised clustering approaches were performed as described in Langwig et al., 2022 [27] using MEBS v1.0 [36]. First, they were scanned against the Pfam v3.0 database to obtain a protein presence/absence profile using mebs.pl -comp option, then we hierarchically clustered Zipacnadia MAGs (mebs_clust.py) using Jaccard distance, Ward variance minimization, and a maximum distance threshold of 0.4 (options –distance –method and –cutoff, respectively). Second, the normalized MEBS scores from the 77 Armatimonadetes MAGs were clustered along with 2,107 publicly available genomes described in De Anda et al., 2017, and 319 references described in Supplementary Table 2. The clustering approach was performed with the F_MEBS_cluster.py script implemented in MEBS.

METABOLIC PREDICTIONS

Gene prediction for individual genomes was performed using Prodigal (v2.6.2) [67]. In addition, predicted genes of individual genomes were characterized using several databases: KofamKOALA [68], Interproscan (v5.31.70) [69], HydDB [70], dbCAN [71], and MEBSv1.1 [36].

Hydrogenases were identified through similar methods as described by Langwig et al. 2021 and De Anda et al. 2021 [7]. Briefly, hydrogenases were identified using DIAMOND v0.9.26.127 [72] against the reference hydrogenase database and then filtered to ensure an alignment length cutoff of > 40 amino acid residues and a sequence identity > 50%. Identified sequences were validated using the HydDB web server [70]. No FeFe- or Fe- hydrogenases were identified. We identified 113 NiFe hydrogenases.

These were used to construct a phylogenetic tree with a compiled database of known NiFe hydrogenases previously described [40]. The sequences were aligned through Geneious Prime v11.0.4+11 using the locally contained MUSCLE v5.1 and MAFFT v7.490 programs. A tree was created of the aligned hydrogenase sequences using IQ-Tree v2.0.3 on model LG+R10 with otherwise the same parameters as the 37-marker gene alignments and phylogeny and visualized in iToL v5.

Annotated proteins from all sources were mapped onto metabolic pathways using the KEGG Mapper tool [73], MetaCyc pathway information [41], and manual curation. Hits for key metabolic marker genes found were verified using BLASTP through the NCBI web server tool.

HYDROXYLAMINE ANNOTATION

Putative hydroxylamine proteins hydroxylamine oxidoreductase (Hao) and hydroxylamine dehydrogenase (hcp) were subjected to phylogenetic analysis to ascertain potential function. A preliminary blastp was conducted on our genes against the NCBI non-redundant database to determine what likely references could be. The reference dataset of each protein sequence was subsequently obtained via different databases. First, 154 reviewed sequences containing the protein domain TIGR01703 were retrieved from Uniprot [74] for Hcp references, and 880 sequences annotated as PF02335 were retrieved from Interpro for Hao references [75]. Reference and our recovered sequences were aligned with MUSCLE v3.8.31 and Mafft v7.310 using default settings and masked at 50% using the default settings of Geneious Prime 2021.0.3. The phylogenetic tree was

generated using IQTree v2.1.4 with 1 000 bootstrapping replicates. Model WAG+R10 was used for Hao, resulting in 867 sequences in the tree. Model Q.pfam+R6 was used for Hcp, resulting in 191 sequences in the tree. The trees were visualized using iTOL v5.

CARBOHYDRATE-ACTIVE ENZYMES (CAZYME) IDENTIFICATION

We use HMMER, DIAMOND, and Hotpep tools within dbCAN v2.0.11 [71] to identify the CAZymes. We also included signalP [76] prediction and PSORT [77] subcellular localization search. We placed all annotations in a matrix, sorted by annotation tools that identified them. CAZymes annotated by at least two tools were accepted and compiled for analysis with the results of the localization searches.

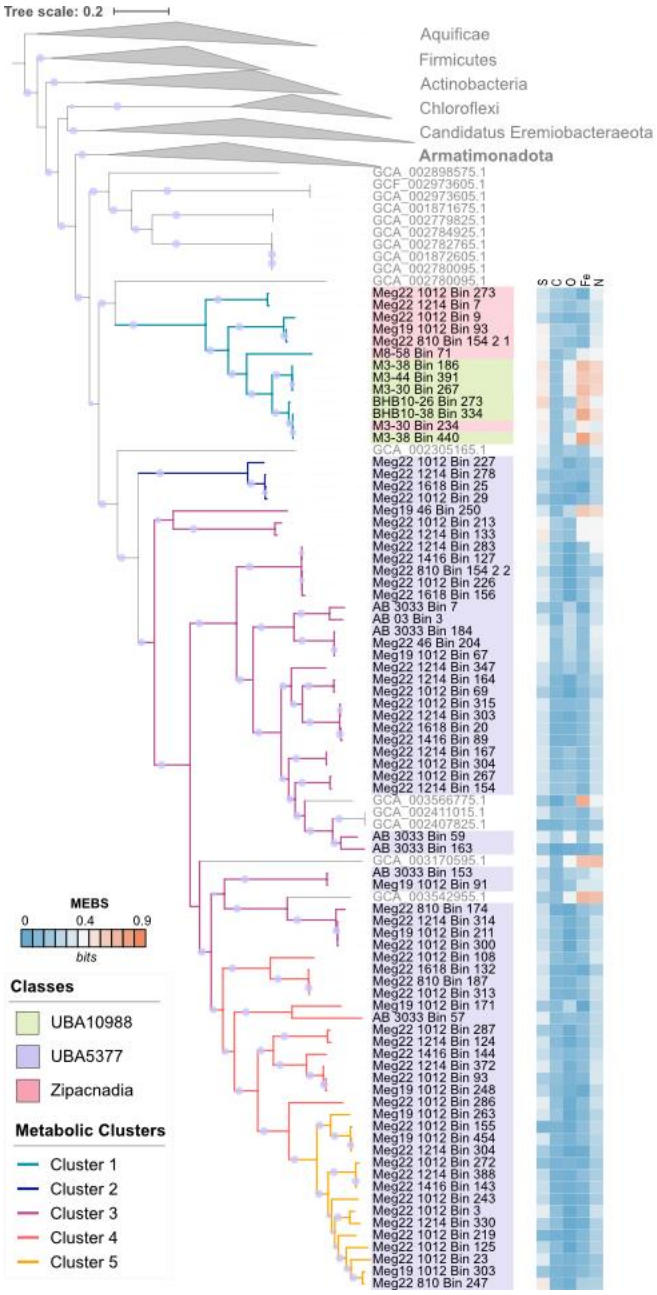

Figure 1: Phylogenetic (37-marker gene) Tree

Phylogeny of recovered genomes and references, based on 37 conserved marker genes. Bootstrap values over 70% are displayed with the proportional size of circles along internal branches. The three classes are distinct compared to NCBI database-available Armatimonadetes (100% bootstrap value). Aquificae was used as an outgroup to root the tree. The three identified classes: UBA10988, UBA5377, and Zipacnadia, are

highlighted within the tree in green, purple, and red. The color of tree branches shows the five metabolic clusters generated by PFAM clustering. Scores generated by MEBS are displayed by color blocks directly outside each corresponding branch label. An interactive version of this tree is available online at https://itol.embl.de/shared/2mUVQn1s5SIs8 as "Main Tree".

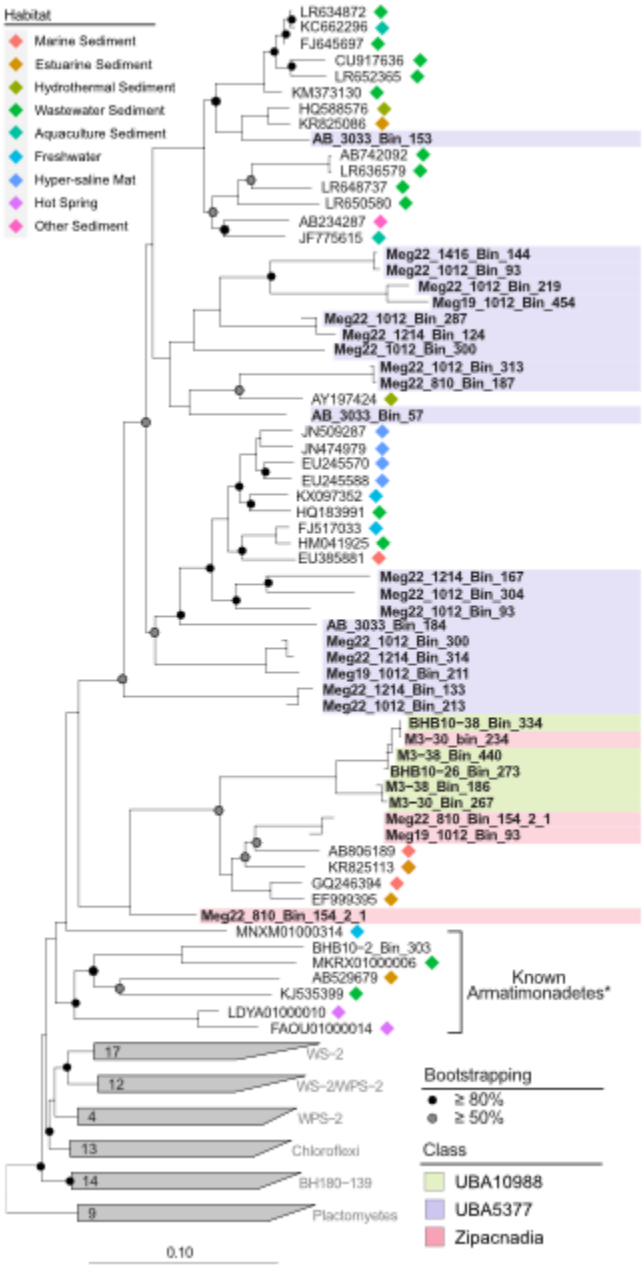

Figure 2: 16S rRNA Gene Tree

16S rRNA gene phylogeny of recovered MAGs. Two main branches make up the recovered Armatimonadetes. The deeper branch contains the shallow sea BS MAGs in classes UBA10988 and Zipacnadia alongside undescribed sequences isolated from the

estuary and sea sediments, while the UBA5377 MAGs are located in the newest branch among undescribed sequences. Bootstraps are displayed as gray and black circles on internal nodes with greater than 50% and 80% bootstrapping, respectively. Colored markers at the end of undescribed sequences note the environment they were collected from.

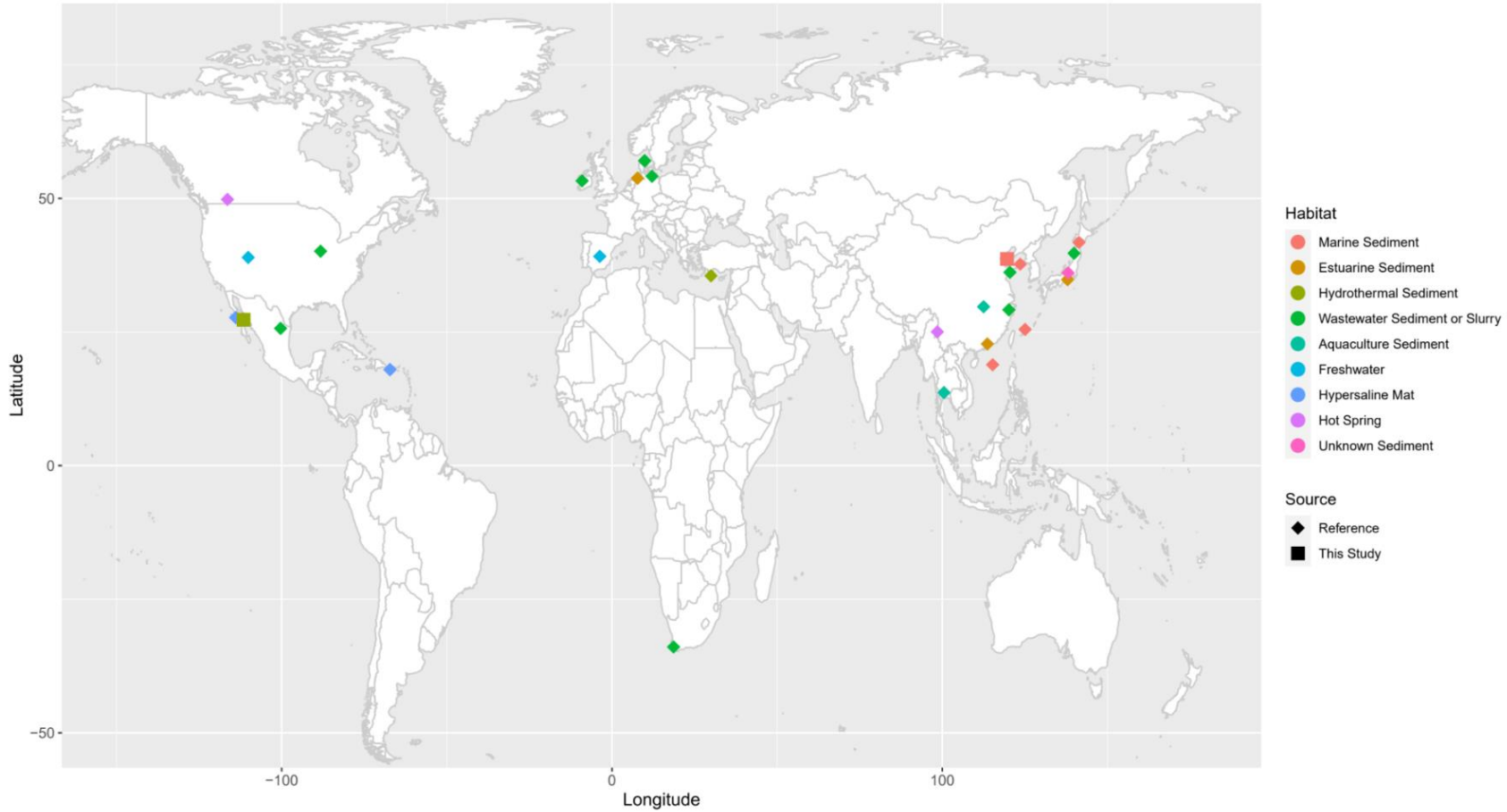

Figure 3: 16S Gene Samples Map

Global map of all Armatimonadetes sequences used in 16S rRNA gene phylogeny. Undescribed sequences from the SILVA database ("Reference") are marked by diamonds, colored to match a particular habitat type. The two locations of MAGs recovered in this study are marked by squares, also colored according to the surrounding habitat. Data collected to create the map is located in Supplementary Table 8

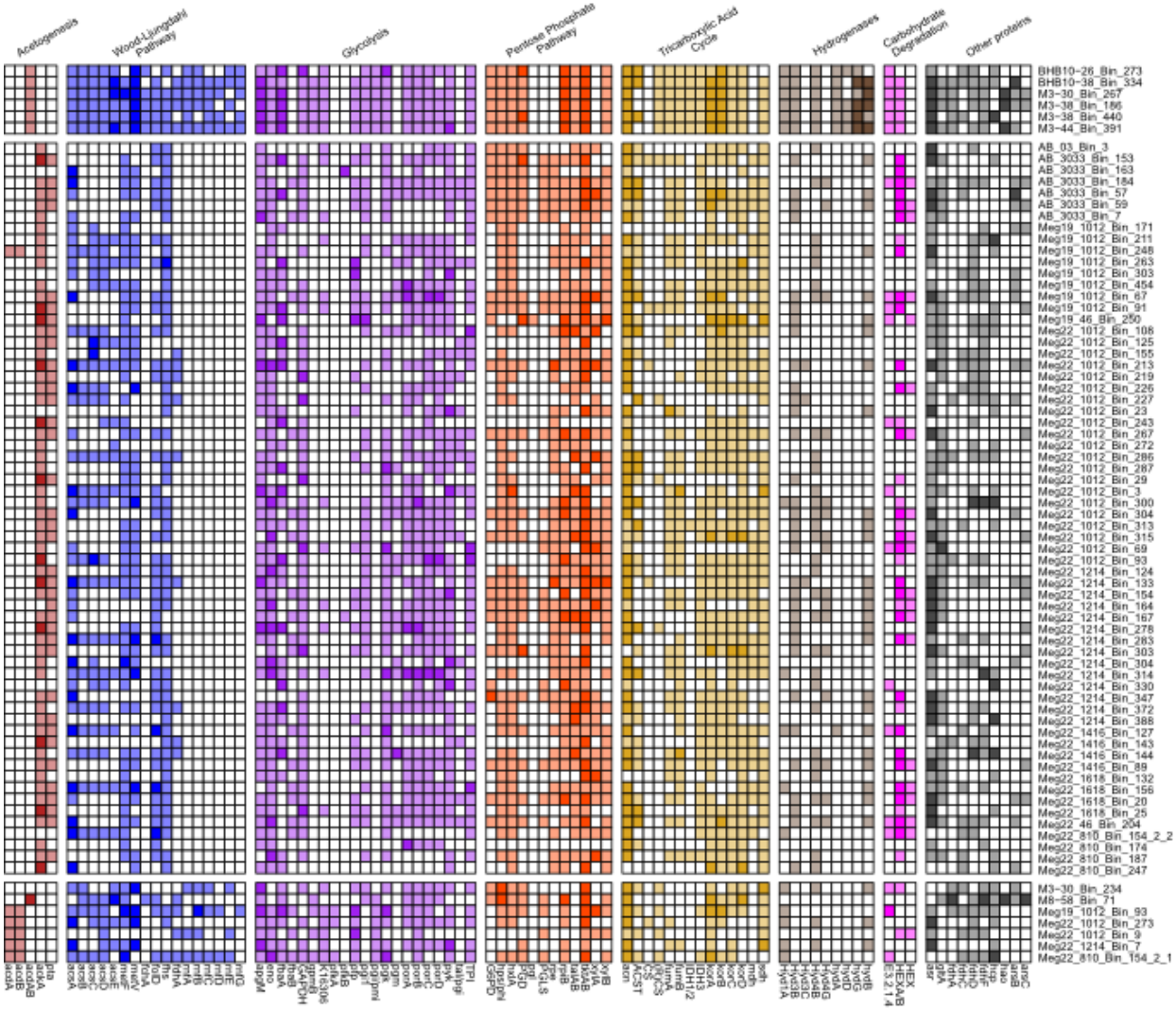

Figure 4: Metabolic Heatmap

A presence/absence chart of notable metabolic gene annotations among recovered MAGs for Acetogenesis, Wood-Ljungdahl Pathway (including Rnf Complex), electron transport chain, glycolysis, pentose phosphate pathway, tricarboxylic acid cycle, hydrogenases, carbohydrate-active enzymes, and other capabilities (such as sulfur and nitrogen utilization). Positive annotations for one copy are shown in the lighter shade of

each category. The darker shade corresponds to two or more copies identified in the genome. All annotations used in this figure are available in Supplementary Table 7.

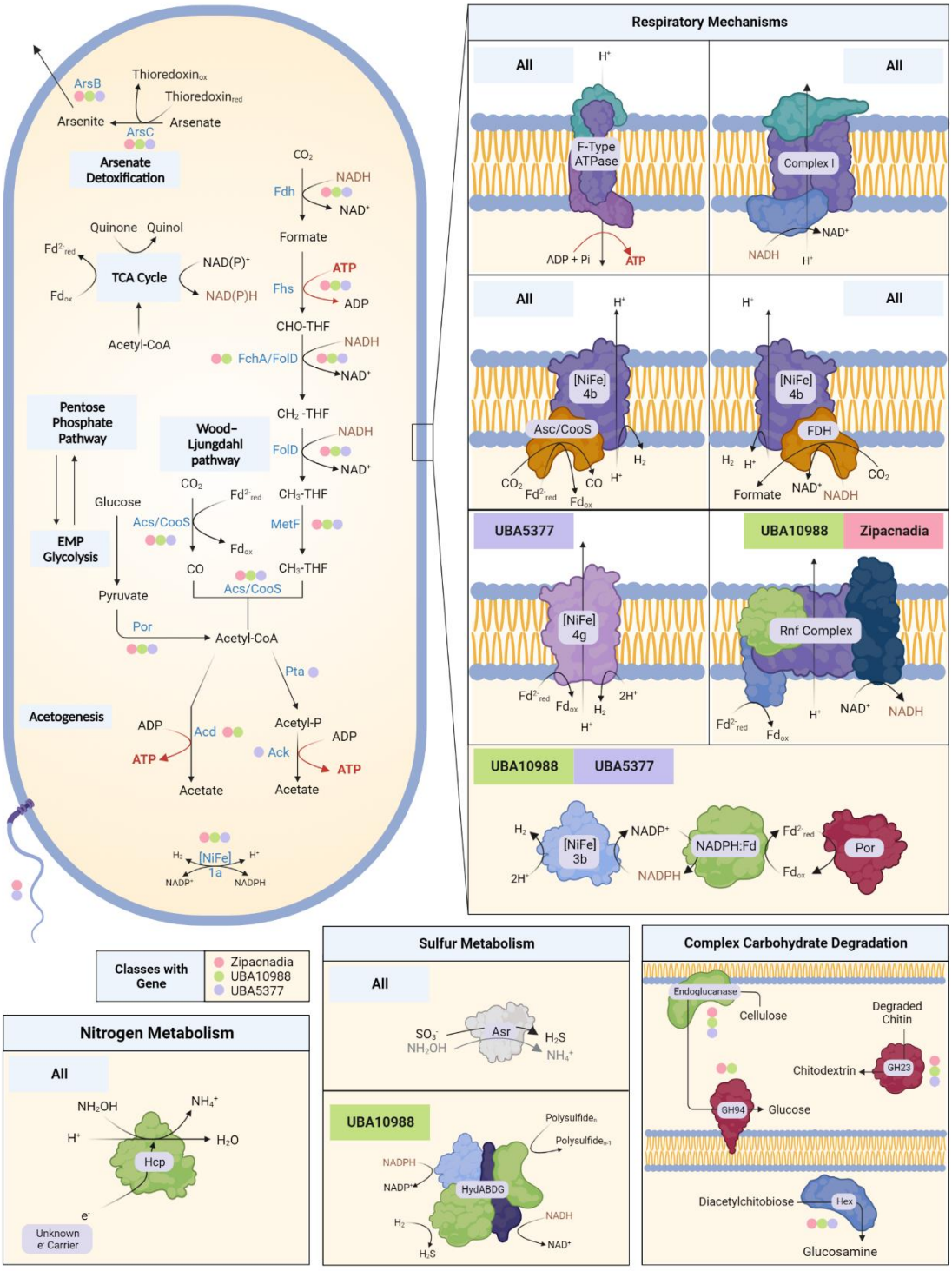

Figure 5: Cell Diagram and Metabolic Pathways

Overview of key metabolic pathways identified in the recovered genomes, using data shown in Figure 4. Features of the core metabolism are shown within the central representation of a cell. Pathways or reactions by class are labeled through text and colored dots. Presence in UBA10988 is noted by a green dot, UBA5377 a purple dot, and Zipacnadia a red dot. Components of energy generation are shown in the Respiratory Mechanisms box. These collected genomes are unique within Armatimonadetes with their redox mechanics, which we hypothesize act on nitrogen and sulfur species. The autotrophic growth allowed by the Rnf Complex in UBA10988 and Zipacnadia is a key capability of those recovered genomes. Unknown protein labels are placeholders for cellular machinery unable to be identified in our analyses. Created with BioRender.com.

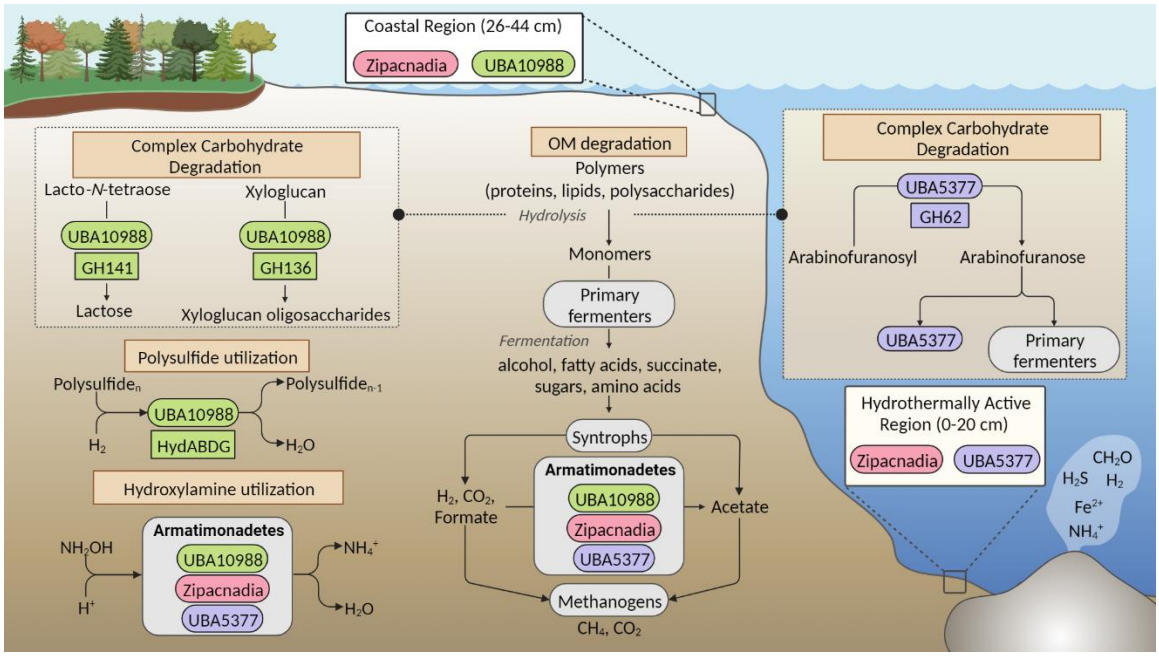

Figure 6: Roles in the Environment

Representation of proposed ecological roles of UBA10988, UBA5377, and Zipacnadia. Class distinctions are denoted by text and color; all unique environmental contributions are shown with the associated protein colored by class. All UBA10988 MAGs and 2/7 proposed Zipacnadia MAGS were located in the anoxic sediments of the Bohai Sea with depths of 26-44 cm below the water-sediment interface. UBA5377 MAGs and 5/7 proposed Zipacnadia were located in the anoxic sediments of the Guaymas Basin with depths of 0-20 cm below the water-sediment interface. All MAGS are anaerobic, functioning primarily on the intermediary compounds sulfite, nitrite, and hydroxylamine provided by sulfate and nitrate-reducing microbes in nearby sediments. Organic Matter (OM) includes buried sedimentary carbon metabolized by heterotrophs to produce CO_2 used directly in the WLP of all MAGs. All classes likely play a dominant role in supplying usable carbon through acetate secretion and further refractory or inorganic matter degradation. In particular, specific complex carbon degradation is confirmed by the presence of extracellular CAZymes (GH62, GH141, GH136), shown here. More distinctions in identified metabolic pathways and capabilities are clearly differentiated between groups in Figure 4. Created with BioRender.com.

Appendices

APPENDIX 1: MANUSCRIPT FUNDING, CONTRIBUTIONS, AND DATA AVAILABILITY

Funding and Support

This work was funded by the Simons Foundation (Award number 687165) and the National Science Foundation MCB – Systems and Synthetic Biology (grant number 1817354) provided to BJB. This research was also supported by the National Natural Science Foundation of China (grant numbers 91951202 and 42006134), Shandong University Foundation for Future Scholar Plan provided to XG. We thank the captain and crew of the R/V Chuangxin Yi for the help in the Bohai Sea sample collection.

Contributions

Valerie De Anda (VDA) and Brett J. Baker (BJB) conceptualized and supervised the project. John D. Carlton (JDC), VDA, and BJB administered the project. BJB and Xianzhe Gong (XG) acquired funding. BJB acquired resources. Marguerite V. Langwig (MVL), Kiley W. Seitz (KWS), and XG extracted DNA from environmental samples, and performed metagenomic sequence assemblies and binning. JDC, VDA., and BJB performed phylogenomic analyses. JDC, VDA, XG, and Mirna Vázquez-Rosas-Landa (MVRL) analyzed genomic data, conducted metabolic analyses and inferences, and created preliminary data representations. JDC, MVRL, and Emily J. Aguilar-Pine (EJA) collected metadata supporting visualization creation. JDC, VDA, and EJA created final visualizations. JDC authored the thesis. VDA, MVL, and BJB contributed text and edits to the thesis. All contributors approved the content of the thesis.

Data Availability

All sequence data and sample information are available at NCBI under BioProject ID PRJNA692327 and PRJNA743900 for samples from the Guaymas Basin and the Bohai Sea, respectively.

APPENDIX 2: DESCRIPTION OF SUPPLEMENTARY FILES

Supplementary Information. *PDF*. Supporting information includes Supplementary Figures 1-10

Supplementary File 1. *PDF*. AAI heatmap

Supplementary Table 1. *Excel file***.** Genomic statistics and basic information

Supplementary Table 2. *Excel file.* Publicly available genomes used as references for phylogenetic analysis (Figure 1)

Supplementary Table 3. *Excel file***.** Relative abundance of the collected MAGs across different sample sites in Guaymas Basin (GB) and Bohai Sea (BS). Details of each sample in GB and BS have been described by Langwig et al., 2021 and Gong et al., 2022, respectively.

Supplementary Table 4. *Excel file***.** Amino Acid comparison (the 77 collected MAGs 98 publicly available phylogenetically related reference genomes)

Supplementary Table 5. KEGG Annotations of the 77 collected MAGs

Supplementary Table 6. MEBS Clustering Information of the 77 collected genomes and references from phylogenetic analysis

Supplementary Table 7. *Excel file.* Annotations used to create Figures 4 and 5.

Supplementary Table 8. *Excel File.* Data used to create the 16S gene sequence map in Figure 3.

Supplementary Table 9. *Excel File*. CAZymes presence/absence and data.

Supplementary Table 10. *Excel File*. Extracellular peptidase presence/absence found through PSORT.

References

1. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. *Nat Microbiol* 2016; **1**: 16048.
2. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. *mSystems* 2018; **3**.
3. Parkes RJ, Cragg B, Roussel E, Webster G, Weightman A, Sass H. A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere:geosphere interactions. *Mar Geol* 2014; **352**: 409–425.
4. Orcutt BN, Sylvan JB, Knab NJ, Edwards KJ. Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol Mol Biol Rev* 2011; **75**: 361–422.
5. Baker BJ, Appler KE, Gong X. New Microbial Biodiversity in Marine Sediments. *Ann Rev Mar Sci* 2021; **13**: 161–175.
6. Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat Commun* 2018; **9**: 4999.
7. De Anda V, Chen L-X, Dombrowski N, Hua Z-S, Jiang H-C, Banfield JF, et al. Brockarchaeota, a novel archaeal phylum with unique and versatile carbon cycling pathways. *Nat Commun* 2021; **12**: 2404.
8. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 2015; **521**: 173–179.
9. Drake HL, Küsel K, Matthies C. Acetogenic Prokaryotes. *The Prokaryotes*. 2013. Springer, Berlin, Heidelberg, pp 3–60.
10. Müller V. Energy Conservation in Acetogenic Bacteria. *Appl Environ Microbiol* 2003; **69**: 6345.
11. He Y, Li M, Perumal V, Feng X, Fang J, Xie J, et al. Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nature Microbiology* 2016; **1**: 1–9.
12. Seitz KW, Lazar CS, Hinrichs K-U, Teske AP, Baker BJ. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *ISME J* 2016; **10**: 1696–1705.
13. Schuchmann K, Müller V. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol* 2014; **12**: 809–821.

14. Esposito A, Tamburini S, Triboli L, Ambrosino L, Chiusano ML, Jousson O. Insights into the genome structure of four acetogenic bacteria with specific reference to the Wood-Ljungdahl pathway. *Microbiologyopen* 2019; **8**.

15. Graber JR, Breznak JA. Physiology and Nutrition of Treponema primitia, an H2/ CO2-Acetogenic Spirochete from Termite Hindguts. *Appl Environ Microbiol* 2004; **70**: 1307.

16. Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, et al. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* 2013; **1**: 1–17.

17. De la Lanza-Espino G, Soto LA. Sedimentary geochemistry of hydrothermal vents in Guaymas Basin, Gulf of California, Mexico. *Appl Geochem* 1999; **14**: 499–510.

18. Simoneit BRT, Lonsdale PF. Hydrothermal petroleum in mineralized mounds at the seabed of Guaymas Basin. *Nature* 1982; **295**: 198–202.

19. Calvert SE. Origin of Diatom-Rich, Varved Sediments from the Gulf of California. *J Geol* 1966; **74**: 546–565.

20. Qiao Y, Feng J, Cui S, Zhu L. Long-term changes in nutrients, chlorophyll a and their relationships in a semi-enclosed eutrophic ecosystem, Bohai Bay, China. *Mar Pollut Bull* 2017; **117**: 222–228.

21. Hu L, Zhang G, Zheng B, Qin Y, Lin T, Guo Z. Occurrence and distribution of organochlorine pesticides (OCPs) in surface sediments of the Bohai Sea, China. *Chemosphere* 2009; **77**: 663–672.

22. Hugenholtz P, Pitulle C, Hershberger KL, Pace NR. Novel division level bacterial diversity in a Yellowstone hot spring. *J Bacteriol* 1998; **180**: 366–376.

23. Hugenholtz P, Goebel BM, Pace NR. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 1998; **180**: 4765–4774.

24. Tamaki H, Tanaka Y, Matsuzawa H, Muramatsu M, Meng XY, Hanada S, et al. Armatimonas rosea gen. nov., sp. nov., of a novel bacterial phylum, Armatimonadetes phyl. nov., formally called the candidate phylum OP10. *Int J Syst Evol Microbiol* 2011; **61**.

25. Lee KCY, Dunfield PF, Stott MB. The Phylum Armatimonadetes. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds). *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. 2014. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 447–458.

26. Tanaka Y, Tamaki H, Tanaka K, Tozawa E, Matsuzawa H, Toyama T, et al. 'Duckweed-Microbe Co-Cultivation Method' for Isolating a Wide Variety of Microbes Including Taxonomically Novel Microbes. *Microbes Environ* 2018; **33**.

27. Langwig MV, De Anda V, Dombrowski N, Seitz KW, Rambo IM, Greening C, et al. Large-scale protein level comparison of Deltaproteobacteria reveals cohesive metabolic groups. *ISME J* 2021.

28. Gong X, del Río ÁR, Xu L, Chen Z, Langwig MV, Su L, et al. New globally distributed bacterial phyla within the FCB superphylum. *Nat Commun* 2022; **13**: 1–12.

29. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; **25**: 1043–1055.

30. Bohlin J, Eldholm V, Pettersson JHO, Brynildsrud O, Snipen L. The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes. *BMC Genomics* 2017; **18**: 1–11.

31. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2019.

32. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 2020; **38**: 1079–1086.

33. Tarn N. Popol Vuh: The Definitive Edition of the Mayan Book of the Dawn of Life and the Glories of Gods and Kings. 1986. JSTOR.

34. Teske A, De Beer D, McKay LJ, Tivey MK, Biddle JF, Hoer D, et al. The Guaymas Basin hiking guide to hydrothermal mounds, chimneys, and microbial mats: Complex seafloor expressions of subsurface hydrothermal circulation. *Front Microbiol* 2016; **7**: 75.

35. Thauer RK, Kaster A-K, Seedorf H, Buckel W, Hedderich R. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* 2008; **6**: 579–591.

36. De Anda V, Zapata-Peñasco I, Poot-Hernandez AC, Eguiarte LE, Contreras-Moreira B, Souza V. MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: unraveling the sulfur cycle. *Gigascience* 2017; **6**: 1–17.

37. Musfeldt M, Schönheit P. Novel type of ADP-forming acetyl coenzyme A synthetase in hyperthermophilic archaea: heterologous expression and characterization of isoenzymes from the sulfate reducer Archaeoglobus fulgidus and the methanogen Methanococcus jannaschii. *J Bacteriol* 2002; **184**.

38. Reeves RE, Warren LG, Susskind B, Lo HS. An energy-conserving pyruvate-to-acetate pathway in Entamoeba histolytica. Pyruvate synthase and a new acetate thiokinase. *J Biol Chem* 1977; **252**: 726–731.

39. Schuchmann K, Müller V. Energetics and Application of Heterotrophy in Acetogenic Bacteria. *Appl Environ Microbiol* 2016; **82**: 4056–4069.

40. Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, et al. Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. *ISME J* 2016; **10**: 761–777.

41. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, et al. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res* 2020; **48**: D445–D453.

42. Dittbrenner S, Chowdhury AA, Gottschalk G. The stereospecificity of the (R)-citrate synthase in the presence of p-chloromercuribenzoate. *Biochem Biophys Res Commun* 1969; **36**: 802–808.

43. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* 2020; **48**: D265–D268.

44. Einsle O, Messerschmidt A, Stach P, Bourenkov GP, Bartunik HD, Huber R, et al. Structure of cytochrome c nitrite reductase. *Nature* 1999; **400**: 476–480.

45. Soler-Jofra A, Pérez J, van Loosdrecht MCM. Hydroxylamine and the nitrogen cycle: A review. *Water Res* 2021; **190**: 116723.

46. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* 2014; **42**: D503–9.

47. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014; **42**: D490–5.

48. Thomas DJ, Styblo M, Lin S. The cellular metabolism and systemic toxicity of arsenic. *Toxicol Appl Pharmacol* 2001; **176**: 127–144.

49. Itoh Yoshikane, Rice John D., Goller Carlos, Pannuri Archana, Taylor Jeannette, Meisner Jeffrey, et al. Roles of pgaABCD Genes in Synthesis, Modification, and Export of the Escherichia coli Biofilm Adhesin Poly-β-1,6-N-Acetyl-d-Glucosamine. *J Bacteriol* 2008; **190**: 3670–3680.

50. Harmsen M, Yang L, Pamp SJ, Tolker-Nielsen T. An update on Pseudomonas aeruginosa biofilm formation, tolerance, and dispersal. *FEMS Immunol Med Microbiol* 2010; **59**: 253–268.

51. Lazarevic V, Abellan F-X, Möller SB, Karamata D, Mauël C. Comparison of ribitol and glycerol teichoic acid genes in Bacillus subtilis W23 and 168: identical function, similar divergent organization, but different regulation. *Microbiology* 2002; **148**: 815–824.

52. Yu S, Su T, Wu H, Liu S, Wang D, Zhao T, et al. PslG, a self-produced glycosyl hydrolase, triggers biofilm disassembly by disrupting exopolysaccharide matrix. *Cell Res* 2015; **25**: 1352–1367.

53. Myka KK, Allcock DJ, Eloe-Fadrosh EA, Tryfona T, Haag AF, Lauro FM, et al. Adaptations of Cold- and Pressure-Loving Bacteria to the Deep-Sea Environment: Cell Envelope and Flagella. In: Chénard C, Lauro FM (eds). *Microbial Ecology of Extreme Environments*. 2017. Springer International Publishing, Cham, pp 51–80.

54. Lee C-YK. The First Insights into the Phylogeny, Genomics, and Ecology of the Novel Bacterial Phylum Armatimonadetes. 2015. University of Waikato.

55. Arp DJ, Stein LY. Metabolism of inorganic N compounds by ammonia-oxidizing bacteria. *Crit Rev Biochem Mol Biol* 2003; **38**.

56. Gong X, Chen Z, Deng Y, Zhao D, Gao P, Zhang L, et al. Contrasting archaeal and bacterial community assembly processes and the importance of rare taxa along a depth gradient in shallow coastal sediments. 2022.

57. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014; **11**: 1144–1146.

58. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 2015; **3**: e1165.

59. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2016; **32**: 605–607.

60. Darling AE, Jospin G, Lowe E, Matsen FA 4th, Bik HM, Eisen JA. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2014; **2**: e243.

61. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019; **47**: W256–W259.

62. Seemann T. barrnap. Github.

63. Mount DW. Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc* 2007; **2007**: db.top17.

64. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 2012; **28**: 1823–1829.

65. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al. ARB: a software environment for sequence data. *Nucleic Acids Res* 2004; **32**: 1363–1371.

66. Parks D. CompareM. Github.

67. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**: 119.

68. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 2020; **36**: 2251–2252.

69. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014; **30**: 1236–1240.

70. Søndergaard D, Pedersen CNS, Greening C. HydDB: A web tool for hydrogenase classification and analysis. *Sci Rep* 2016; **6**: 34212.

71. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 2018; **46**: W95–W101.

72. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015; **12**: 59–60.

73. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci* 2020; **29**: 28–35.

74. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* 2023; **51**: D523–D531.

75. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* 2021; **49**: D344–D354.

76. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 2011; **8**: 785–786.

77. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 2010; **26**: 1608–1615.