Degradation of marine and terrestrial organic matter in anoxic marine sediments of the Helgoland mud area

DISSERTATION

zur

Erlangung des Grades eines

Doktors der Naturwissenschaften

- Dr. rer. nat. -

dem Fachbereich Biologie/Chemie

der Universität Bremen vorgelegt von

Mara Denice Maeke

Bremen

Januar 2025

The research presented in this thesis was conducted from April 2020 to January 2025 in the working group of Microbial Ecophysiology at the University of Bremen. This thesis was prepared under the framework of the International Max Planck Research School of Marine Microbiology (MarMic) and partially funded by the DFG under Germany's excellence Strategy, no. EXC-2077-390741603.

Gutachter

1. Gutachter: Prof. Dr. Michael W. Friedrich (Universität Bremen)
2. Gutachter: Dr. Tristan Wagner (Institut de Biologie Structurale Grenoble, Frankreich)
3. Gutachter: Prof. Dr. William Orsi (Ludwig-Maximilians-Universität München)

Tag des Promotionskolloquiums

07.03.2025

Table of Content

Summary

Large amounts of marine and terrestrial organic carbon are deposited in marine sediments. Microorganisms inhabiting these sediments remineralize organic carbon by decomposing large complex molecules into smaller organic molecules and eventually into dissolved inorganic carbon. The process of remineralization is facilitated by a complex community, including primary fermenters thriving on complex organic compounds, secondary fermenters utilizing fermentation products, and respiratory microorganisms coupling the oxidation of fermentation products to an electron acceptor. Despite ongoing research, bacteria and archaea involved in the process of organic matter remineralization remain mostly uncharacterized, owing to the lack of isolated representatives in most of the bacterial and archaeal groups. While novel data-integrative studies attempted to shed light on the phylogenetic diversity, only a fraction of these microorganisms has been characterized.

This dissertation aims to characterize bacteria and archaea involved in the organic matter remineralization to further unravel complex microbial interactions in anoxic marine sediments by utilizing a combination of enrichments, stable isotope probing (SIP), metagenomics and data-mining. In **Chapter II,** I describe the application of a novel developed small-scale ^{18}O-RNA-SIP strategy to identify active microorganisms in the *in situ* sediments and, thereby, detected the dominance of autotrophic sulfur compound cycling organisms of the class Desulfobulbia involved in the ultimate steps of organic matter degradation. Active heterotrophic secondary fermenters, such as Fusobacteriia, Clostridia, Bacteroidia and respiratory microorganisms, such as Sva1033 detected in these sediments were most likely fueled by fermentation products derived from complex organic matter degrading taxa, such as an undescribed group of Thermoplasmatota involved in protein degradation (**Chapter III**). The class EX4484-6 was so far only represented by five genomes of one order and our efforts utilizing a newly developed data-mining strategy led to the increased representation, including three novel orders. Due to its low abundance in coastal marine sediments globally, we classified the EX4484-6 as a rare biosphere member. By functional comparison of the class members with other Thermoplasmatota, we detected high functional novelty within this group, indicated by high numbers of unclassified genes not detected in other Thermoplasmatota members. Other detected active primary fermenters include undescribed members of four orders of the Clostridia, acting on methoxylated aromatic compounds (MACs), deriving from terrestrial organic matter (**Chapter IV**). While these Clostridia acted on a large variety of MACs, their

metabolic potential suggested a general C_1-carbon utilization, indicating these members as important members in carbon transformations in marine sediments.

The results presented in this dissertation offer valuable new insights into previously undetected microorganisms involved in the remineralization of organic matter in anoxic marine sediments. Among these, not only abundant members but also rare biosphere members contributed to carbon remineralization, indicating that low-abundant taxa in marine sediments can become essential for organic carbon degradation. The detection of diverse understudied microorganisms and their novel functional potential hints towards the vast untapped microbial diversity and functionality that is yet to be explored.

Zusammenfassung

Große Mengen organischen Kohlenstoffs marinen und terrestrischen Ursprungs sind in marinen Sedimenten abgelagert. Mikroorganismen, die diese Sedimente bewohnen, remineralisieren organischen Kohlenstoff, in dem sie große komplexe Moleküle in kleinere organische Moleküle und schließlich inorganischen Kohlenstoff abbauen. Der Prozess der Remineralisierung wird durch eine komplexe Gemeinschaft bestehend aus primären Fermentern, die große komplexe organische Stoffe nutzen, sekundären Fermentern, die Fermentationsprodukte nutzen und respiratorischen Organismen, die die Oxidation der Fermentationsprodukte mit der Reduktion von Elektronenakzeptoren koppeln, durchgeführt. Trotz der bestehenden Forschung sind viele Bakterien und Archaeen, die in der Remineralisierung von organischen Materialien involviert sind, aufgrund fehlender Isolation weitestgehend nicht charakterisiert. Während neue Daten integrierenden Studien versuchen mehr Licht in die phylogenetische Diversität zu bringen, konnten bisher nur wenige der Mikroorganismen charakterisiert werden.

Diese Dissertation hat das Ziel Bakterien und Archaeen, die organischen Kohlenstoff remineralisieren, zu charakterisieren und komplexe mikrobielle Interaktionen aufzudecken. Dabei wird eine Kombination aus Anreicherungen, dem Beproben stabiler Isotope (SIP), Metagenomanalysen und der Sammlung und Analyse von weiteren Daten angewendet. In **Kapitel II** beschreibe ich die Anwendung eines neu entwickelten SIP, das mit kleinen Sedimentmengen durchgeführt wurde und ^{18}O als Substrat nutzt, um aktive Mikroorganismen *in situ* zu detektieren. Dabei wurden Mitglieder der Klasse Desulfobulbia als dominante Organismen entdeckt, die im finalen Schritt des Abbaus organischen Materials involviert sind. Aktive heterotrophe sekundäre Fermenter, wie die Fusobacteriia, Clostridia und Bacteroidia und respiratorische Organismen wie Sva1033 wurden im Sediment entdeckt und mit Fermentationsprodukten von anderen Taxa versorgt, die im Abbau von komplexen organischen Materialien beteiligt waren. Als solche wurde, zum Beispiel eine unbeschriebene Klasse innerhalb der Thermoplasmatota entdeckt, die im Abbau von Proteinen involviert ist (**Kapitel III**). Die Repräsentation der EX4484-6 Klasse war bisher auf fünf Genome aus einer Ordnung beschränkt und konnte durch unsere Entwicklung einer Methode zur Datensuche und -analyse verbessert werden. Hierdurch wird die Klasse nun durch drei neue Ordnungen ergänzt. Aufgrund der geringen Abundanz dieser Gruppe in globalen küstennahen marinen Sedimenten, konnten wir die EX4484-6 Klasse der seltenen Biosphäre zuordnen. Durch funktionelle Vergleiche der Mitglieder mit anderen Thermoplasmatota, haben wir neue Funktionalität innerhalb der neuen Klasse entdeckt, die sich durch eine hohe Zahl an unklassifizierten Genen

auszeichnet, die nicht in anderen Mitgliedern der Thermoplasmatota auftreten. Andere aktive primäre Fermenter waren unbeschriebene Mitglieder aus vier Ordnungen der Clostridia, die methoxylierte aromatische Verbindungen (MACs) terrestrischen Ursprungs nutzen (**Kapitel IV**). Diese Clostridia konnten eine große Vielfalt an verschiedenen MACs nutzen. Das metabolische Potential zeigte jedoch auch ein generelles Potential für die Nutzung andere C_1-Kohlenstoffe auf und indiziert damit, dass diese Organismen auch anderen Kohlenstoff in marinen Sedimenten transformieren können.

Die Ergebnisse, die in dieser Dissertation präsentiert werden, geben einen neuen Einblick in zuvor unentdeckte Mikroorganismen, die in der Remineralisierung von organischen Materialien in marinen Sedimenten involviert sind. Dabei wurden nicht nur abundante Organismen, sondern auch Organismen der seltenen Biosphäre entdeckt, die zur Remineralisierung der organischen Kohlenstoffverbindungen beitragen. Damit zeigt diese Dissertation, dass auch wenig abundante Taxa in marinen Sedimenten wichtig für den Abbau organischen Kohlenstoffs sein können. Das Auffinden von diversen kaum untersuchten Mikroorganismen und ihres neuen funktionellen Potentials deutet auf die enorme mikrobielle Diversität hin, die noch immer darauf wartet, entdeckt zu werden.

Chapter I

Introduction

1.1 Organic matter deposition in marine sediments

The marine water column and sediments are estimated to contain a significant portion of fixed carbon on earth. In the marine water column, organic matter is present as particulate organic matter (POM) and dissolved organic matter (DOM). The global amount of dissolved organic matter (DOM) found in the marine water column is estimated to be about 660 petagrams (Pg, 10^{12} kg) carbon (Hansell et al. 2009). With this, the amount of carbon in the water column equals that in the atmosphere. Even more carbon was estimated to be stored in marine sediments containing about 2,322 Pg carbon in the top meter below the surface (Atwood et al. 2020). Therefore, marine sediments contain large amounts of organic matter, potentially available for microbial degradation. Organic matter in marine sediments is of marine (autochthonous) and terrestrial (allochthonous) origin. Marine organic matter is derived from photosynthetic organisms inhabiting the surface ocean layer (Middelburg 2018). During photosynthesis, microorganisms assimilate atmospheric carbon dioxide into their biomass. Phytoplankton debris and detritus form necromass, which is exported from surface layers into the marine water column and sediments (Lomstein et al. 2012). Besides photosynthetic organisms, other autotrophic and heterotrophic microorganisms inhabiting the water column contribute to the export of marine organic matter (Kawasaki et al. 2011, Lomstein et al. 2012, Braun et al. 2017). Terrestrial organic matter (TOM), similar to marine organic matter, is derived from the photosynthesis of plants and enters the marine system mainly through river discharge (Schlesinger and Melack 1981, Blair and Aller 2012).

Organic matter in the ocean undergoes a process of remineralization. During remineralization, organic matter is decomposed into smaller organic molecules and eventually into dissolved inorganic carbon (Emerson 1985). While labile organic matter has turnover times of hours to days, more complex recalcitrant organic matter can persist for thousands and tens of thousands of years (Hansell et al. 2009, Hansell 2013, Follett et al. 2014). Remineralization within the water column is dependent on the export efficiency of organic matter. Export efficiency has been linked to the sinking rates of particles, which are promoted by particle aggregation (POM), and ecosystem structure (Keil and Kirchman 1994, Armstrong et al. 2001, Francois et al. 2002, Klaas and Archer 2002, Lutz et al. 2002). During sinking, more labile organic matter, such as proteins or nucleic acids, is hydrolyzed into smaller monomers or consumed by microbes, zooplankton, and nekton (Wakeham et al. 1997a, Wakeham et al. 1997b, Dauwe et al. 1999,

Iversen et al. 2010). Larger, more refractory molecules, such as cellulose, lignins and algaenan are less prone to microbial degradation. Proportionally, more of the refractory organic matter reaches the sediment. Microorganisms, protozoans, and metazoans process the organic carbon within the sediment (Hedges and Keil 1995).

Only 1% or less of the remaining low reactive organic matter is ultimately buried in marine sediments (Canfield 1994, Middelburg 2018). High sedimentation rates promote the burial of organic matter. Areas along continental shelves, such as the coasts of Namibia, Peru, Baja California, the Caribbean Sea, and the Baltic Sea, especially hold the highest concentrations of organic matter with up to 18,666 t carbon per km^2 (Berner 1982, Burdige 2007, Bianchi et al. 2018, Middelburg 2018, Atwood et al. 2020). Within these areas, additional carbon supplies through river discharges and land runoff or large phytoplankton blooms in coastal zones and upwelling regions add to the accumulation of organic matter (Burdige 2007, Atwood et al. 2020). The high accumulation of carbon and sediments can promote the preservation of labile organic matter. According to estimations, continental shelfs store between 256-274 Pg carbon (Atwood et al. 2020). Marine-derived organic matter consists of 50-60% proteins and amino acids, 20 % RNA, 3 % DNA, 10 % lipids and 10 % sugars (Neidhardt et al. 1990, Burdige 2007, Lengeler et al. 1999). In contrast, terrestrial organic matter is composed of more refractory polyaromatic and aromatic compounds found in living plant biomass, plant litter and soil organic matter (Burdige 2007). Overall, such plant material consists of about 30% lignin and 70% carbohydrates (Burdige 2007).

1.2 Organic matter degradation in marine sediments

The availability of organic matter in marine sediments shapes the microbial community. For all organic matter to be degraded, various microorganisms must interact by carrying out different metabolic capabilities (Arndt et al. 2013). The organic matter reaching the sea floor is mineralized in respiratory or fermentative pathways. Microorganisms respire oxygen in the uppermost sediment layers (Arndt et al. 2013, Jørgensen et al. 2019). These aerobic microorganisms can completely degrade organic compounds to CO_2, accounting for more than half of the carbon mineralization in marine sediments (Arndt et al. 2013). As oxygen penetration into the sediment varies between different sediment types, oxygen might be depleted within the first millimeters (Glud 2008), necessitating anaerobic degradation.

As a first step in anaerobic organic matter degradation, the polymeric organic matter is hydrolyzed by extracellular or membrane-bound hydrolytic enzymes to transform macromolecules into monomers and, with this, into more soluble molecules (Middelburg et al. 1993, Arnosti 2004). The resulting molecules, such as sugars, amino acids, lipids, or organic

acids, are then used by primary fermenting and acetogenic microorganisms, which produce hydrogen, CO_2, and short-chain fatty acids, such as formate, acetate, propionate, and butyrate (Jørgensen et al. 2019). During fermentation, the organic substrates become both oxidized and reduced (Doelle 1975, Madigan et al. 2021). In fermentation, energy-rich compounds are formed, which contain phosphate bonds or coenzyme A, e.g., phosphoenolpyruvate or acetyl-CoA (Madigan et al. 2021). Substrates like acetyl-CoA contain thioester bonds, which can be transformed into an energy-rich acid anhydride by trans acetylation (e.g., acetyl phosphate). In substrate-level phosphorylation (SLP), the energy-rich phosphate bond is transferred to ADP to form ATP. In fermentations with lower energy yield, energy gained through decarboxylation is used to drive ATP synthesis by a proton-gradient (e.g., succinate decarboxylation to propionate by *Propionigenium modestum)* (Schink 2006). Fermentation products are excreted from the cell (Madigan et al. 2021). The ultimate mineralization of organic matter in anaerobic sediments is coupled to the reduction of terminal electron acceptors in the order of nitrate (NO_3^-), manganese (Mn(VI)), iron (Fe(III)) and sulfate (SO_4^{2-}), according to their available redox potential (Canfield and Thamdrup 2009).

Nitrate, similar to oxygen, has a high oxidation potential and is used as a primary electron acceptor after oxygen in marine anaerobic sediments (Zumft 1997, Devol 2015). Nitrate reduction is carried out primarily by heterotrophic bacteria (Zumft 1997, Devol 2015). Manganese and iron are found in their oxidized state in the form of oxides and oxyhydroxides in marine sediments (Stumm and Morgan 2013). These electron acceptors are predominantly present in the upper sediment layers and are depleted with increasing sediment depth (Jørgensen et al. 2019). Manganese and iron-reducing microorganisms use fermentation products and aromatic compounds as electron donors during microbial degradation (Burdige 1993). The resulting Mn^{2+} can be found in its dissolved form in the pore water; other states form organic complexes. The resulting Fe^{2+} precipitates with hydrogen sulfide to first form FeS and eventually the insoluble pyrite, or carbonates to form the insoluble phase siderite (Burdige 1993). The electron donors in both of these processes are oxidized to CO_2. As the penultimate electron acceptor, sulfate is abundant in marine seawater and penetrates meters into the marine sediments (Jørgensen et al. 2019). In the upper marine sediments, sulfate is continuously supplied and is a potential electron acceptor besides oxygen, nitrate, manganese and iron (Thamdrup et al. 1994). In lower anoxic sediment, the higher energy-yielding acceptors are depleted, and sulfate becomes the prevalent electron acceptor (Jørgensen et al. 2019). Most cultivated sulfate-reducing microorganisms use H_2 and fermentation products, besides other electron donors such as hydrocarbons and aromatic compounds (Jørgensen et al. 2019). In

dissimilatory sulfate reduction, sulfate is reduced to hydrogen sulfide; the organic carbon is oxidized to CO_2 or acetate (Laanbroek and Pfennig 1981, Nagpal et al. 2000, Grigoryan et al. 2008, Jørgensen et al. 2019). Upon sulfate depletion, methanogenic archaea become dominant (Jørgensen et al. 2019). Methanogenic archaea utilize H_2 or formate and CO_2, acetate, or methylated compounds, such as methanol, methylated amines or methoxylated aromatic compounds as substrates (Kurth et al. 2020). In the processes of hydrogenotrophic methanogenesis, CO_2 is reduced to CH_4 (Kurth et al. 2020). During acetoclastic methanogenesis, acetate is both reduced to CH_4 and oxidized to CO_2; similarly, methylated compounds can become reduced and oxidized during methylotrophic methanogenesis (Kurth et al. 2020).

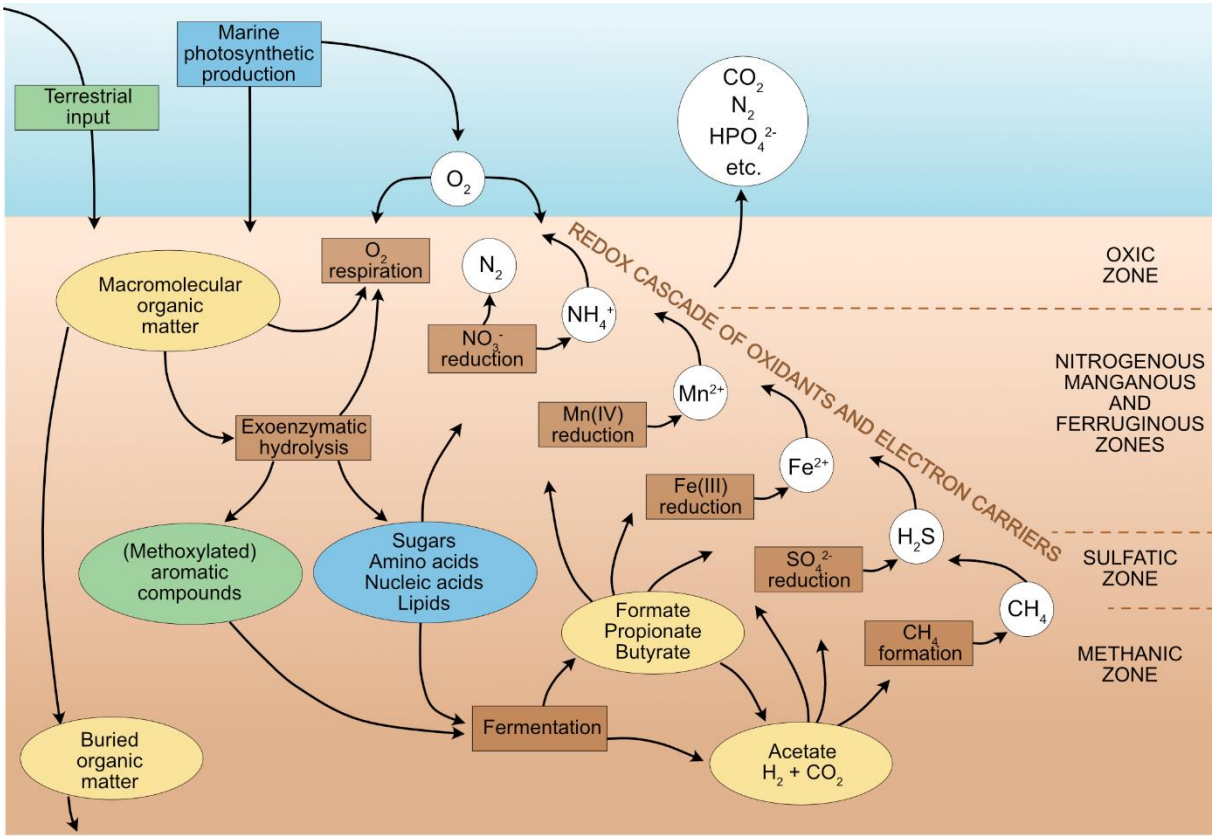

Figure 1 Organic matter degradation in marine sediments. Marine organic matter, derived from photosynthetic processes and terrestrial organic matter, is deposited in marine sediments. Extracellular enzymes hydrolyze macromolecular organic matter. Smaller molecular organic matter is used by fermenters, which release fermentation products, such as formate, propionate and butyrate. Microorganisms of the respiratory pathways use fermentation products. Ultimately, CO_2 and CH_4 are formed. The figure was modified after Jørgensen et al. (2022), initially published under the Creative Commons CC-BY license in Earth-Science Reviews, Elsevier (https://creativecommons.org/licenses/by/4.0/).

While microbial respiration and the secondary fermentation of short-chain fatty acids have been studied more intensively in marine sediments, many of the microorganisms involved in primary fermentation and the initial step of exoenzymatic hydrolysis of macromolecules remain rather unexplored (Arndt et al. 2013). Multiple groups of subseafloor bacteria and archaea, such as

Alphaproteobacteria, Gammaproteobacteria, Bacteroidetes, Chloroflexi, Firmicutes and Bathyarchaeia, were identified to contain enzymes involved in the exoenzymatic hydrolysis of macromolecules (Orsi et al. 2018). Due to their high abundance in the subseafloor, archaea are thought to play an important role in these fermentation processes (Biddle et al. 2006, Lipp et al. 2008, Arndt et al. 2013, Hoshino and Inagaki 2018, Orsi et al. 2020). Yet, many of these microorganisms remain unknown due to the lack of culturability. With advances in molecular techniques using the conserved 16S rRNA gene as a marker gene, first insights were gained into the diversity of non-culturable microorganisms (Pace et al. 1986, Ward et al. 1990, Hugenholtz et al. 1998). More recent surveys, including metagenomic and metatranscriptomic derived 16S rRNA genes, showed high phylogenetic novelty of both archaea and bacteria, especially in subseafloor sediments (Parkes et al. 2014, Lloyd et al. 2018). The number of yet uncultured organisms ranges between 7.3×10^{29} and 5.9×10^{29} cells globally (Lloyd et al. 2018). By computing rarefaction curves from the observed taxa recovery in the SILVA REF 114 database, species richness was estimated to reach ~1,350 different bacterial and archaeal phyla (Yarza et al. 2014). So far, databases, such as the Genome Taxonomy Database (GTDB), contain 194 phyla as of December 2024 (Yarza et al. 2014, Parks et al. 2018). Moreover, phylogenomic reconstructions showed a lack of isolated representatives in most of the bacterial and archaeal groups (Hug et al. 2016). By then, large-scale metagenomic studies, such as the genomic catalog of Earth's microbiomes (Nayfach et al. 2021), Ocean Microbiomics Database (Paoli et al. 2022), genomic catalog from global cold seeps (Han et al. 2023), catalog of glacier microbiomes (Liu et al. 2022), or the predecessor study of the GTDB (Parks et al. 2017) expanded the phylogenetic diversity. Despite these ongoing attempts to shed light on the phylogenetic diversity, only a fraction of microorganisms has been described. In recent years, identifying and characterizing undescribed microorganisms unraveled unexpected metabolic capabilities, for example, Bathyarchaeia and *Ca*. Thermoprofundales were the first non-extremophilic archaea identified to contain genes for protein fermentation in marine sediments (Lloyd et al. 2013) and Lokiarchaeia were found to encode eukaryotic signature proteins, building the bridge between the archaea and eukaryotes (Spang et al. 2015). Studies performed on the uncharacterized majority demonstrate that these uncharacterized microorganisms host millions of functional genes unknown to any cultured organism (Rodríguez del Río et al. 2024). Yet, the functional characterization of many taxa remains missing. Thus, gaining insights into such understudied organisms and processes is of utmost importance. In this dissertation, marine and terrestrial organic matter degradation processes are studied using the highly abundant complex organic matter: protein and lignin-derived aromatic compounds.

1.3 Protein degradation

Proteins are biopolymers built from multiple amino acids and are a significant component of marine organic matter. Within all living cells, proteins are required for structural support, as enzymes and building blocks (Madigan et al. 2021). Before cellular uptake for intracellular degradation, proteins are cleaved extracellularly into peptides by secreted endoproteases (Fig. 2) (Hoppe et al. 2002). Further exopeptidases free amino acids (Nunn et al. 2003), which together with small peptides are taken up by heterotrophic archaea and bacteria. Within marine sediments, both bacteria and archaea were involved in the extracellular hydrolysis of proteins (Lloyd et al. 2013, Pelikan et al. 2021, Yin et al. 2022). Peptides and amino acids were imported into the microbial cells and further degraded (Pelikan et al. 2021, Yin et al. 2022). During amino acid degradation, amino acids are transaminated by glutamate dehydrogenases (Consalvi et al. 1991, Rahman et al. 1998, Yokooji et al. 2013) or other amino acid:2-oxoacid aminotransferases, such as alanine aminotransferases (Ward et al. 2000), aspartate aminotransferases (Ward et al. 2002), aromatic aminotransferases (Andreotti et al. 1994, Andreotti et al. 1995, Matsui et al. 2000, Ward et al. 2002) or alanine:glyoxylate aminotransferases (Sakuraba et al. 2004) to form corresponding 2-oxoacids. These 2-oxoacids are further degraded through 2-oxoacid oxidoreductases, which reduce ferredoxin, and form an acyl-CoA. Among the 2-oxoacid oxidoreductases, four different oxidoreductases with different substrate specificity are known: pyruvate:ferredoxin oxidoreductase (POR) (Blamey and Adams 1993, Smith et al. 1994, Ma et al. 1997), 2-oxoisovalerate:ferredoxin oxidoreductase (VOR) (Heider et al. 1996, Kletzin and Adams 1996, Ozawa et al. 2005), indolepyruvate:ferredoxin oxidoreductase (IOR) (Mai and Adams 1994, Siddiqui et al. 1997, Ozawa et al. 2012) and 2-oxoglutarate:ferredoxin oxidoreductase (KOR) (Mai and Adams 1996a, Schut et al. 2001). Hydrogenases subsequently reoxidize the reduced ferredoxin. Acyl-CoA is converted into its corresponding fatty acid through substrate-level phosphorylation (SLP) by an ADP-forming acyl-CoA synthetase (Schäfer et al. 1993, Mai and Adams 1996b, Glasemacher et al. 1997, Shikata et al. 2007). During this process, ATP is produced. In contrast to archaea, some proteolytic bacteria use the “Stickland reaction” to degrade amino acids (Nisman 1954, Orsi et al. 2020). These bacteria couple the oxidation of one amino acid to the reduction of another (Fonknechten et al. 2010).

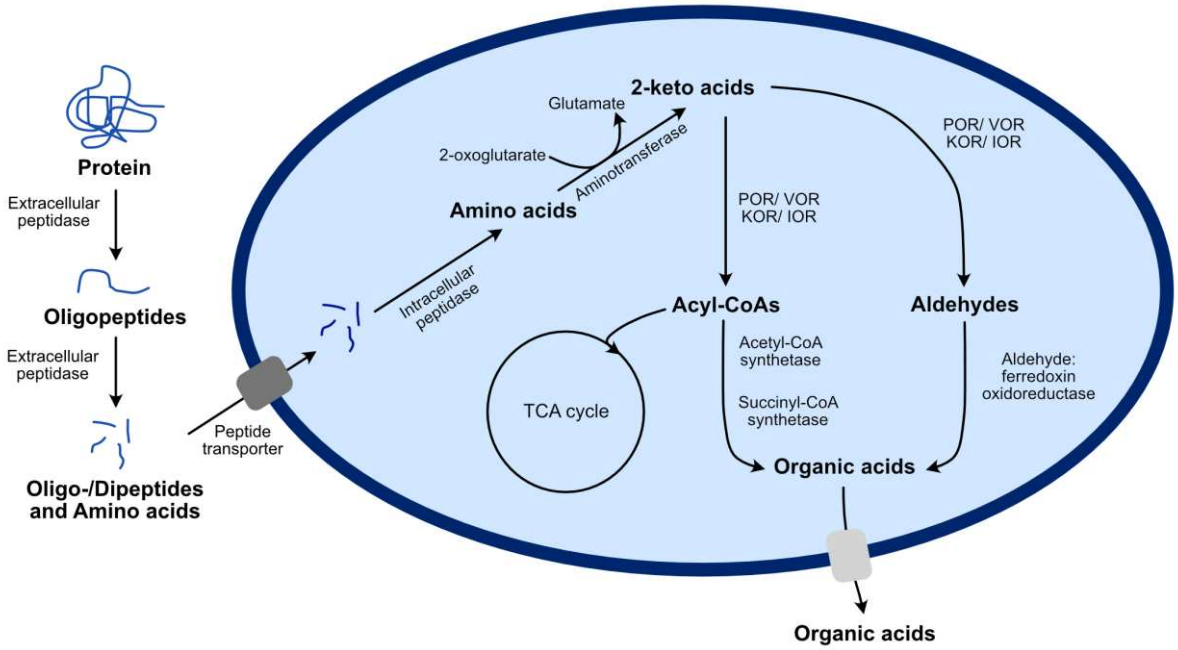

Figure 2 Pathway of protein degradation by archaea. Schematic overview of the intermediates and genes involved in protein degradation in anaerobic marine systems. The pathway was modified after Lloyd et al. (2013) and Yin et al. (2022). Electron transfer processes are not displayed in this overview. POR: pyruvate:ferredoxin oxidoreductase; VOR: oxoisovalerate:ferredoxin oxidoreductase; KOR: 2-oxoglutarate:ferredoxin oxidoreductase; IOR: indolepyruvate:ferredoxin oxidoreductase; TCA cycle: tricarboxylic acid cycle.

1.4 Lignin degradation

Lignin, as part of terrestrial organic matter, is the second most abundant biopolymer on earth (Burdige 2005, Gargulak et al. 2015). Lignin, along with cellulose and hemicellulose, is found in the secondary cell walls of plants, where they function as structural components (Boerjan et al. 2003). During biosynthesis, the lignin biopolymer is formed from phenolic polymers, namely *p*-coumaryl alcohol (*p*-hydroxyphenyl unit; H unit), coniferyl alcohol (guaiacyl; G unit) and sinapyl alcohol (syringyl; S unit) (Vanholme et al. 2010). As lignins constitute of 20 to 30% of vascular plant tissue, these biopolymers are often used as a biomarker for detecting the presence of terrestrial organic matter in marine sediments (T K Kirk and Farrell 1987, Bianchi et al. 2018).

Due to its size and recalcitrance (Zoghlami and Paës 2019), lignin cannot be used by most microorganisms directly; it needs to be broken down into monomers (T K Kirk and Farrell 1987). The microbial degradation of lignin has been mostly attributed to terrestrial fungi and bacteria. Within the kingdom of fungi, white-rot fungi, brown-rot fungi, and soft-rot fungi were shown to decompose lignin (Crawford and Crawford 1980, Sánchez 2009). These fungi use extracellular oxidative enzymes such as laccases and ligninolytic peroxidases to break down the lignin polymer (Bugg et al. 2011a). Within the kingdom of bacteria mostly Actinobacteria, Proteobacteria and Firmicutes contain lignin-degrading enzymes (Tian et al. 2014). However,

none of the bacteria analyzed, showed a complete enzymatic machinery to degrade whole lignin polymers (Brown and Chang 2014, Fisher and Fong 2014). Moreover, aerobic bacteria are also thought to modify the structure of lignin (Bugg et al. 2011b, Brown and Chang 2014, Fisher and Fong 2014). Like fungi, aerobic bacteria use laccases, extracellular peroxidases, and manganese superoxide dismutases to break down the lignin polymer (Bugg et al. 2011a, Rashid et al. 2015, Wang et al. 2021). In addition, dioxygenases were also detected, which take part in the breakdown of lignin components (Bugg et al. 2011a). The breakdown of different lignins leads to the formation of various aromatic compounds (Bugg et al. 2011a). For example, many lignin components result in the product vanillin. The breakdown of spruce wood lignin releases aromatic carboxylic acids, such as 3,4-dimethoxybenzoic acid, the breakdown of kraft lignin releases products such as 3,4,5-trimethoxybenzaldehyde or 4-hydroxy-3-methoxycinnamic acid (Bugg et al. 2011a).

Only a few studies so far have focused on the lignin degrading microorganisms in marine systems (Wang et al. 2021, Li et al. 2022, Peng et al. 2023). In these systems, the metabolic capability of abundant groups, such as Pseudomonadota, Bacillota and Bacteroidia showed their involvement in lignin degradation (Wang et al. 2021, Li et al. 2022, Peng et al. 2023). So far, only aerobic organisms have been shown to depolymerize lignin and most of the microbial groups identified perform aerobic degradation of lignin components (Li et al. 2022, Peng et al. 2023). A few anaerobic bacteria were identified as being involved in degrading secondary aromatic ring structures derived from lignin components. Among these, the families Arcobacteraceae, Cohaesibacteraceae, Desulfovibrionaceae, Lachnospiraceae, Marinifilaceae and Vibrionaceae were identified (Peng et al. 2023). However, close interactive connections between aerobic and anaerobic microorganisms were observed, suggesting an interplay between these organisms to fully degrade aromatic lignin-derived compounds (Peng et al. 2023). Most studies on the degradation of terrestrial organic matter focused on the degradation capabilities of microbial soil communities. Much less is known about microbial TOM degradation in marine sediments. By detecting the organisms and processes involved in the anaerobic degradation of terrestrial organic matter, the processes involved in the breakdown of lignin in the anoxic marine biosphere can be identified.

1.4.1 Anaerobic Methoxylated Aromatic Compound (MAC) degradation

To understand degradation processes during lignin breakdown in fully anoxic sediments, intermediate products of lignin polymer breakdown can be used as model compounds. The three main building blocks of lignin differ in their functional groups. Both the S unit and G unit contain methoxy groups (-OCH_3), classifying these building blocks as so-called **M**ethoxylated

Aromatic **C**ompounds (MAC). Methoxy groups can make up to 3 wt % of the biopolymer lignin (Lee et al. 2019). For the degradation of MACs, the methyl group from the MAC is removed first (Kato et al. 2015). Second, the remaining aromatic ring structure is cleaved to form acetate, hydrogen and carbon dioxide. Ultimately, the degradation of MACs leads to the formation of methane or carbon dioxide (Grbić-Galić 1983, Kato et al. 2015).

MACs are transported into the cell by passive diffusion, transport channels, or proteins specific for aromatic compounds (Kamimura et al. 2017). Passive transporters, porins or substrate-specific channels facilitate transport across the outer membrane of gram-negative bacteria (Kamimura et al. 2017). Aromatic compounds are mainly translocated by transporters of the major facilitator superfamily (MFS), the ATP-binding cassette (ABC transporters), the tripartite ATP-independent periplasmic (TRAP) transporters and the ion transporter (IT) superfamily (Kamimura et al. 2017).

1.4.1.1 Demethoxylation

During anaerobic degradation, organisms using the methoxy groups of MACs utilize a four-component aromatic *O*-demethylase system, which requires tetrahydrofolate to perform demethylation (Kaufmann et al. 1997, 1998). Firstly, the C-O bond in the methylated substrate is cleaved by a substrate-specific methyltransferase I (MTI, *mtvB*). Due to their substrate specificity, multiple MTI-coding genes might be present in single genomes (Khomyakova and Slobodkin 2023). The methyl group is then transferred by a cobalamin-binding corrinoid protein to methyltransferase II (MTII, *mtvA*), which transfers the methyl group to a terminal acceptor. Lastly, the cobalamin-binding corrinoid protein (*mtvC*) is reactivated by an activating enzyme (Kaufmann et al. 1997).

Methoxydotrophic acetogenesis

Methoxy group utilizing microorganisms are called methoxydotrophs. The main pathway used by anaerobic methoxydotrophic bacteria and archaea is methoxydotrophic acetogenesis (Fig. 3a and 3b) (Khomyakova and Slobodkin 2023). Hereby, one methyl group derived from a methoxy group of MACs is oxidized to CO_2 in the methyl branch of the Wood Ljungdahl pathway (WLP), forming six reducing equivalents. The reducing equivalents are required to reduce three molecules of CO_2 to CO in the carbonyl branch, which, combined with an additional three methyl groups from MACs, form three molecules of acetyl-CoA and subsequently acetate (Drake and Daniel 2004, Pierce et al. 2008, Kremp et al. 2018).

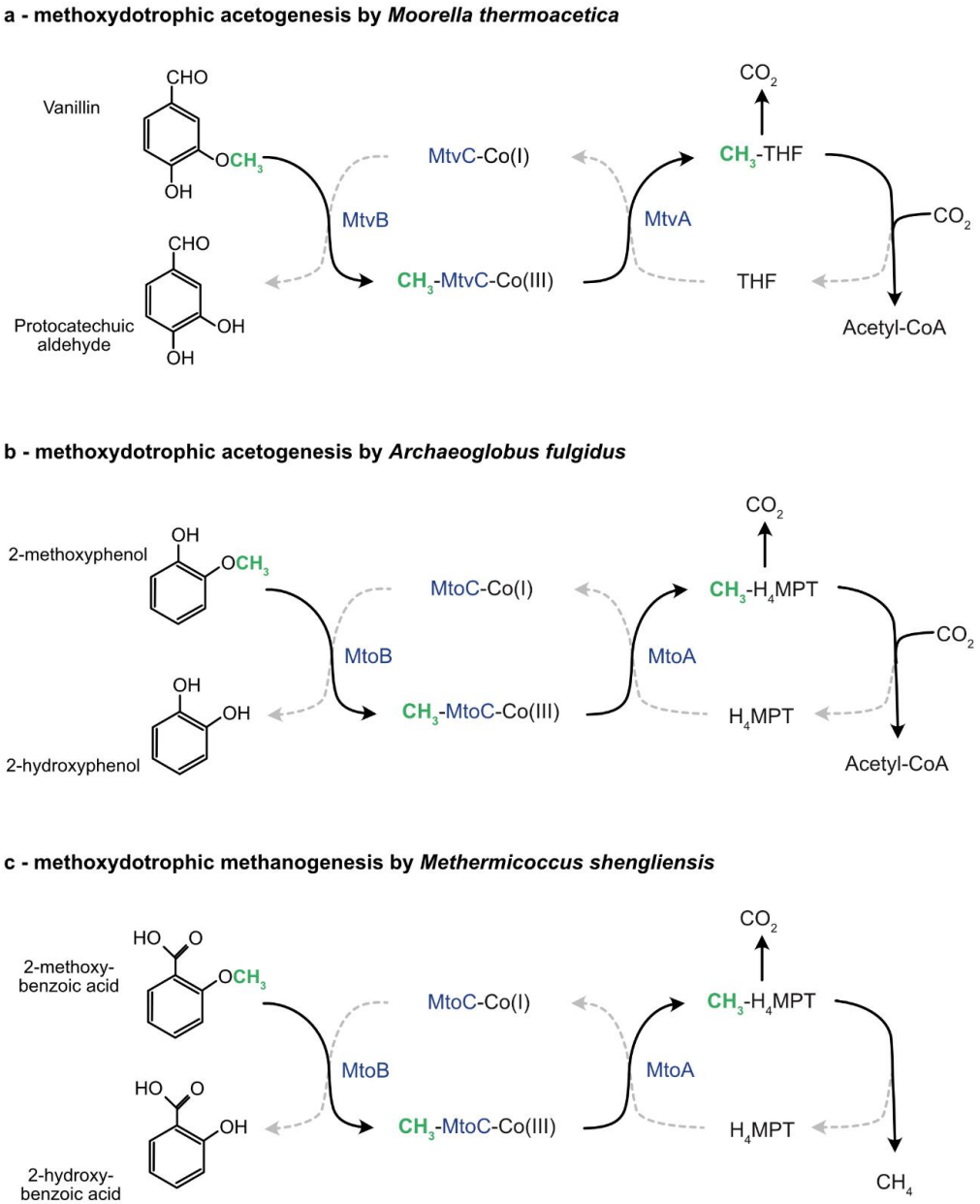

Figure 3 Demethoxylation pathways in bacteria and archaea. **(a)** Demethoxylation pathway of the acetogenic bacterium *Moorella thermoacetica* modified after Pierce et al. (2008). **(b)** Demethoxylation pathway of the acetogenic archaeon *Archaeoglobus fulgidus* modified after Welte et al. (2021). **(c)** Demethoxylation pathway of the methanogenic archaeon *Methermicoccus shengliensis* modified after Kurth et al. (2021). Transferred methyl groups are indicated in green. Genes of the *O*-demethylase systems are indicated in blue. Solid black lines indicate the transfer of the methyl group from the methoxylated aromatic compound (MAC). THF: tetrahydrofolate; H_4MPT: tetrahydromethanopterin.

Methoxydotrophic methanogenesis

Some archaea are thought to be involved in the conversion of methoxy groups to methane in the process of methoxydotrophic methanogenesis (Fig. 3c) (Welte et al. 2021). These archaea were found to use *O*-demethylase systems with high similarity to those found in acetogenic bacteria and archaea (Welte et al. 2021). Instead of the methyl group carrier coenzyme M

(CoM), which is known for methylotrophic methanogens, tetrahydromethanopterin is used as a methyl group carrier in methoxydotrophic methanogenesis (Welte et al. 2021).

1.4.1.2 Aromatic ring cleavage

After demethylation of the MAC, the remaining aromatic ring structures are further degraded anaerobically via four main intermediates (Fig. 4): benzoyl-CoA, phloroglucinol, resorcinol, or hydroxyhydroquinone (HHQ) (Schink et al. 2000, Carmona et al. 2009, Fuchs et al. 2011, Khomyakova and Slobodkin 2023).

Benzoyl-CoA pathway

Most aromatic compounds, e.g., benzoate, cresols, phenylacetate, or ethylbenzene are degraded via the benzoyl-CoA pathway (Fig. 4) (Schink et al. 2000, Fuchs et al. 2011). Benzoyl-CoA is dearomatized by benzoyl-CoA reductase (BCR) to form cyclohexa-1,5-dienoyl-1-carboxyl-CoA (Boll and Fuchs 1995, Carmona et al. 2009). In the BCR I class, this reaction is coupled to the oxidation of ferredoxin as an electron donor and the hydrolysis of ATP (Boll and Fuchs 1995, Carmona et al. 2009). The BCR II class functions ATP-independent and requires electron bifurcation (Carmona et al. 2009, Fuchs et al. 2011, Löffler et al. 2011). After de-aromatization, cyclohexa-1,5-dienoyl-1-carboxyl-CoA is degraded via a modified β-oxidation (acyl-CoA hydratase, hydroxyacyl-CoA dehydrogenase, oxoacyl-CoA hydrolase) to form a C_7-dicarboxyl-CoA compound (3-hydroxypimelyl-CoA or pimelyl-CoA) (Carmona et al. 2009), which is further transformed via glutaryl-CoA to acetyl-CoA following a β-oxidation (Carmona et al. 2009).

Phloroglucinol pathway

Phloroglucinol (1,3,5-trihydroxybenzene) contains three hydroxy groups, which allow tautomerization to destabilize the aromatic ring (Schink et al. 2000). In the phloroglucinol pathway, phloroglucinol is first reduced to dihydrophloroglucinol via an NADPH-dependent phloroglucinol reductase (PGR) (Fig. 4b) (Patel et al. 1981, Armstrong and Patel 1994, Schink et al. 2000). The ring structure is cleaved by a dihydrophloroglucinol cyclohydrolase (DPGC) to 3-hydroxy-5-oxo-hexanoate, which is further reduced by a triacetate forming dehydrogenase ((S)-3-hydroxy-5-oxohexanoate dehydrogenase) (TfD) (Zhou et al. 2023). The reduction of triacetate via triacetate:acetoacetate-lyase (TAL) leads to the formation of acetoacetate and acetoacetyl-CoA (Zhou et al. 2023). Ultimately, acetate or butyrate are formed from phloroglucinol.

Resorcinol and Hydroxyhydroquinone pathway

Resorcinol (1,3-dihydroxybenzene), similar to phloroglucinol, contains hydroxy groups, allowing tautomerization (Schink et al. 2000). In the first step, resorcinol is reduced to cyclohexanedione via a ferredoxin-dependent resorcinol reductase (Fig. 4c) (Carmona et al. 2009). The resulting intermediate is further hydrolyzed to 5-oxohexanoate and further degraded to butyrate and acetate (Tschech and Schink 1985, Schink et al. 2000). The enzymes involved in this process have not yet been characterized. Nitrate-reducing bacteria first hydroxylate resorcinol to form a destabilized HHQ using a resorcinol hydroxylase (Philipp and Schink 1998). In a subsequent reaction, HHQ is oxidized to 2-hydroxy-1,4-benzoquinone (HBQ) by an HHQ dehydrogenase (Philipp and Schink 1998). Nitrate is an electron acceptor in both oxidation reactions (Philipp and Schink 1998). As a last step, the ring is cleaved by an uncharacterized ring cleavage enzyme, and succinate and acetate are formed (Schink et al. 2000, Carmona et al. 2009).

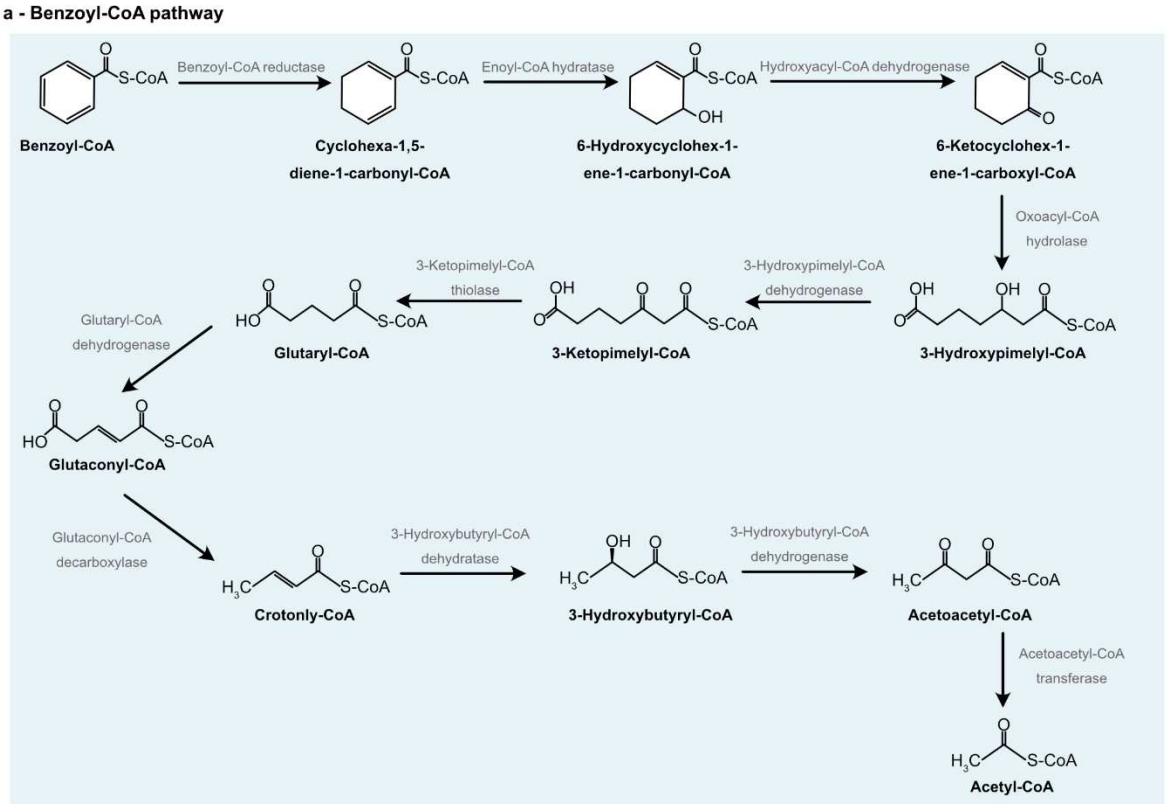

Figure 4 *Part 1.* Pathways of aromatic ring degradation. **(a)** Pathway of anaerobic benzoyl-CoA degradation modified after Carmona et al. (2009).

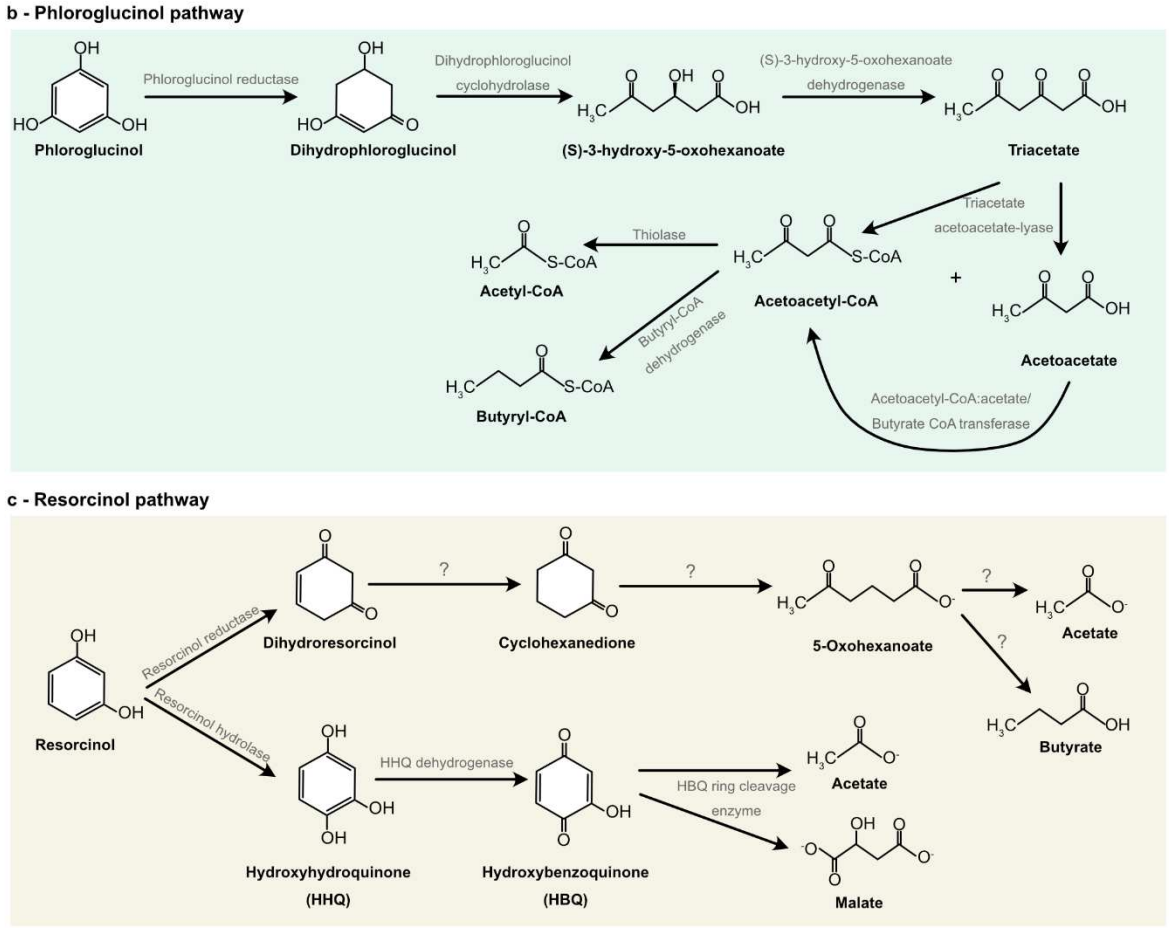

Figure 4 *Part 2*. Pathways of aromatic ring degradation. **(b)** Pathway of phloroglucinol degradation modified after Zhou et al. (2023) **(c)** Pathway of resorcinol and hydroxyhydroquinone degradation modified after Schink et al. (2000) and Carmona et al. (2009). Genes of the resorcinol pathway have not been described yet and are therefore indicated by question marks. Electron transfer processes are not displayed in this overview.

1.5 Strategies to characterize uncultivated taxa

1.5.1 Enrichments

When cultivation attempts fail, enrichment techniques can be used to characterize the properties of microorganisms and help to elucidate the role of these microorganisms in an environment (Madigan et al. 2021). Enrichment cultures are set up using an inoculum and incubation conditions, reflecting the researched environment. While these enrichment cultures might help to investigate the metabolic potential of different organisms, usually only the most abundant organisms are studied, while less abundant microorganisms are overlooked. In addition, enrichment cultures still include multiple organisms, making assigning metabolic functions to one organism more challenging. To further characterize the metabolic potential of single groups, techniques, such as metagenomics and metatranscriptomics are required. A method directly coupling the carbon utilization and microbial activity is stable isotope probing (SIP).

1.5.2 Stable isotope probing (SIP)

Stable isotope probing (SIP) can be employed to analyze active microorganisms. Stable isotopes are non-radioactive isotopes which occur in lesser amounts in the environment. Stable isotopes of carbon (^{13}C), nitrogen (^{15}N), hydrogen (^{2}H, deuterium) and oxygen (^{18}O) can be used to label substrates (Fig. 5). By incorporating these labeled substrates into nucleic acids, microbial taxa can be separated based on their isotopically labeled and unlabeled nucleic acids using density gradient centrifugation (Schildkraut et al. 1962, Wagner 1994, Felske et al. 1998). By using SIP, growing (DNA-SIP) and active taxa (DNA-SIP and RNA-SIP) are labeled (Radajewski et al. 2000, Schwartz et al. 2016). The usage of RNA-SIP thereby gives an advantage over DNA-SIP as RNA synthesis occurs faster than DNA replication and therefore isotopes can be traced much quicker in metabolically active but slow or non-growing microorganisms (Manefield et al. 2002, Dumont and Murrell 2005, Neufeld et al. 2007).

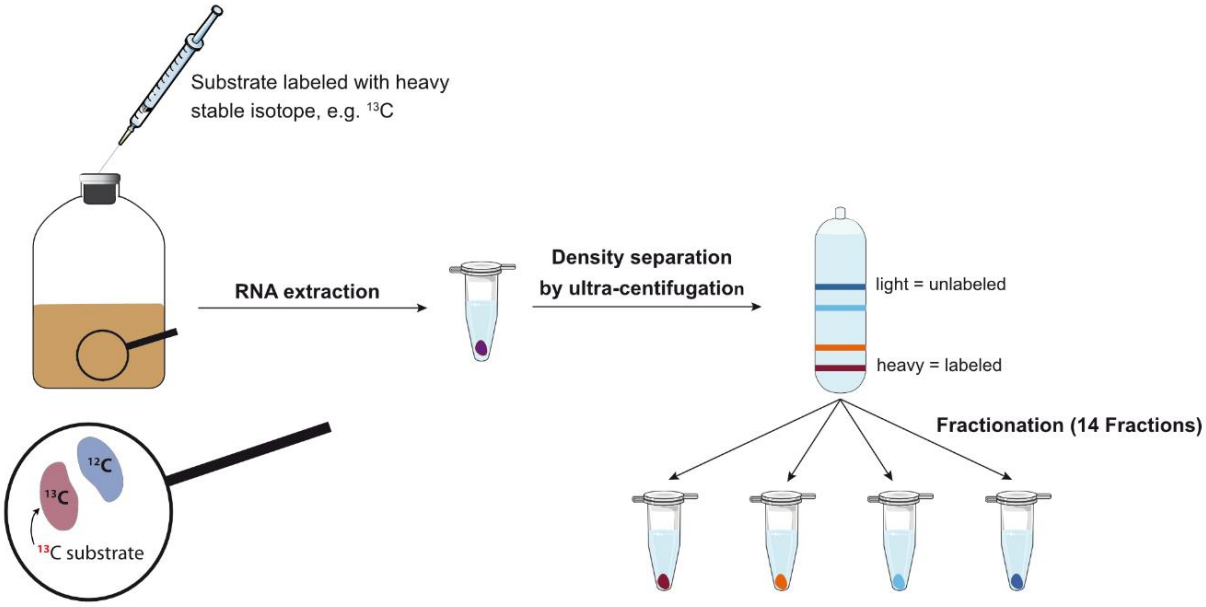

Figure 5 Schematic overview of the method of stable isotope probing (SIP), including the steps of incubation, RNA extraction, density separation by ultra-centrifugation and fractionation. The graphic of Eppendorf cups was obtained from Servier Medical Art (https://smart.servier.com, last accessed 10.12.2024) and modified to our use. The graphics are licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Organic and inorganic carbon are used as energy sources by microorganisms and are required for microbial growth. Using ^{13}C-labeled carbon sources, the carbon assimilation by microorganisms can be traced (Radajewski et al. 2000, Kim et al. 2023). In the cases of labeled growth substrates or contaminants, the stable isotope probing (SIP) gives an advantage over simple DNA/RNA sequencing, by creating a link between not only the identity of active microbial taxa but also the substrates they incorporate and degrade (Radajewski et al. 2000). Besides studying specifically labeled substrates to evaluate metabolic capabilities, $H_2{}^{18}O$ can

trace all active microbial populations in enrichments regardless of their carbon metabolism (Schwartz et al. 2016). Water is ubiquitous for all living organisms and is incorporated into phosphodiester linkages of nucleic acids through various assimilatory pathways. As such, water exchanges its oxygen atom with inorganic phosphate, pyrophosphate and ATP during enzyme activities of e.g., ATPases and phosphatases (Richards and Boyer 1966, Adair and Schwartz 2011).

1.5.3 Metagenomics

Metagenomics describes the sequencing of all genomes in a microbial community (Quince et al. 2017). With decreasing costs for sequencing and improved computational techniques and bioinformatic tools, metagenomics has become a powerful tool, allowing the analysis of the metabolic potential of whole microbial communities (Quince et al. 2017). After metagenomic sequencing, reads are quality-trimmed and *de novo* assembled to reconstruct single genomes. Most commonly, *de novo* assemblies are performed using de Bruijn graphs, in which sequencing reads are split into 'kmers'. By finding overlaps between these 'kmers', sequencing reads are merged to form contigs (Compeau et al. 2011). Different approaches then cluster these contigs, e.g., using the coverage and composition of contigs (Alneberg et al. 2014) or empirical probabilistic distances of genome abundance and tetranucleotide frequency (Kang et al. 2019, Nissen et al. 2021) into 'bins' and, by this, reconstruct single genomes. Besides these reference-free binning approaches, reference-based binning can be applied if reference genomes are available. However, this is often not the case with uncharacterized taxa; therefore, reference-free binning is used preferentially (Yue et al. 2020). Pipelines, such as metaWRAP or BASALT have enabled user-friendly approaches to refine and reassemble bins to retrieve high-quality metagenome-assembled genomes (MAGs) (Uritskiy et al. 2018, Qiu et al. 2024). Retrieved MAGs are taxonomically classified and annotated to obtain genes and reconstruct metabolic pathways (Quince et al. 2017).

1.6 Study site - Helgoland mud area

The HMA is a shallow depositional area with a water depth of < 30 m, located in the German Bight, southeast of the island Helgoland and spans an area of 500 km^2. Originally, the mud area was formed by a halotectonic depression, which was filled with muddy sediments of riverine origin (Hebbeln et al. 2003). Sediment deposition in this area is continuously influenced by longshore coastal currents, discharge from the Elbe and Weser rivers, tidal dynamics, as well as storm floods (Dominik et al. 1978, Hertweck 1983) and was found to vary between an average sedimentation of 2 mm/year and 7.7 mm/year (Reineck 1963, Dominik et al. 1978).

The total organic carbon (TOC) in the HMA varies between 0.1 – 3 wt% with the highest TOC content in the southern part closest to the Elbe River outflow (Müller et al. 2024).

Due to high sedimentation and burial rates as well as different sources of organic matter to the HMA, this makes it an excellent setting to study heterotrophic microorganisms involved in anaerobic organic matter degradation of both marine and terrestrial origin. For this study, two cores located in the western HMA were used (Fig. 6, HE483/2-2, lat 54.087627 lon 7.968191 and HE531/3-1, lat 54.103164 lon 7.96274). Recent studies showed the presence of high marine organic matter depositions deriving from algae predominantly in the sediments of the western HMA (Müller et al. 2024). In this western area, the upper sediment consists mostly of the more easily biodegradable marine organic matter, while more recalcitrant terrestrial organic matter, such as lignin, tannin and cellulose is buried in the deeper sediments of the HMA (Oni et al. 2015).

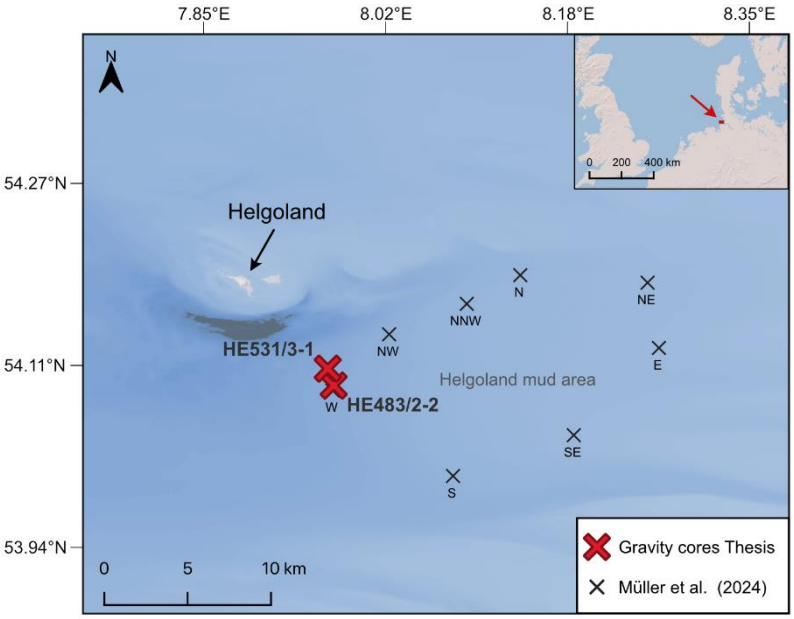

Figure 6 Overview map of the Helgoland mud area (HMA). The base map was created using QGIS v3.4 (QGIS Development Team 2024) with the ESRI World Shaded Relief base map (Esri 2009) and overlaid by the EasyGSH Bathymetry of the year 2016 (Sievers et al. 2020). Red marks indicate sediments of gravity cores used in this thesis. Black marks indicate cores used in the study by Müller et al. (2024) for investigations of depositions and organic carbon amounts in the HMA. Both cores used in this thesis are located close to those derived from the western HMA (W).

1.7 Aim of this study

Due to the lack of culturability, the vast majority of microorganisms are yet to be explored. Culture-independent approaches started to estimate the number of microorganisms expected to inhabit the earth's environments, showing us that research is still needed to characterize the unknown organisms. Advances in sequencing technologies and deeper sequencing allowed the identification of novel taxa. In recent years, novel groups, such as multiple new classes and orders within the Thermoplasmatota, Asgardarchaeota, and Crenarchaeota, were identified in anaerobic marine sediments, showing the potential to degrade not only small organic carbon compounds but also more complex carbon compounds, such as proteins and lignin-derived compounds. In this thesis, I studied the degradation of marine and terrestrial-derived organic matter in the Helgoland mud area (HMA). I aimed to understand organic matter degradation processes through primary fermentation by uncultured bacteria and archaea and detect syntrophs involved in secondary fermentation and respiratory organisms involved in the subsequent degradation of fermentation products in anoxic marine sediments.

For the study, the Helgoland mud area sediments were used, as these sediments are characterized by the input of both marine and terrestrial organic matter. This thesis used model compounds to study primary fermentation processes. The model compounds were proteins, which are significant components of marine organic matter, and MACs, deriving from lignin and, therefore, a significant component of terrestrial organic matter (Burdige 2005, 2007).

My thesis followed four main hypotheses:

1. The dominant microorganisms detected in organic carbon-depleted sediments are autotrophic and heterotrophic organisms acting on DIC and fermentation products.

2. Autotrophic taxa become predominantly active in organic carbon-depleted sediments.

3. Due to the lack of knowledge on primary fermenting microorganisms, using different carbon sources in anoxic marine sediments will reveal uncharacterized primary fermenting organisms besides already known taxa.

4. Besides main players in anoxic marine sediments, low abundant microbial taxa are important for the degradation of organic matter.

1.8 Chapter overview

Chapter II: Linking sulfur and carbon cycling with $H_2{}^{18}O$ RNA-SIP: active taxa and metabolic interactions in anoxic marine sediments. In this chapter, RNA stable isotope probing (SIP) with ^{18}O water was used to label active microorganisms involved in element cycling in *in situ* sediments of the Helgoland mud area. For this, a new strategy was developed to perform small-scale anoxic SIP-enrichments. We could detect key players Desulfocapsaceae, Desulfurivibrionaceae, Desulfobulbaceae and Sva1033 as the most active bacteria performing most likely sulfur disproportionation, sulfate reduction, and dissimilatory iron reduction. A high taxonomic novelty was detected among members of these families. Further fermentative bacteria, such as Fusobacteriia, Bacteroidia and Clostridia, were identified, most likely supplying autotrophic and heterotrophic bacteria with electron donors.

Chapter III: Extensive data mining uncovers novel diversity among members of the rare biosphere within the Thermoplasmatota. In this chapter, enrichments were combined with metagenomic sequencing to characterize an undescribed class within the phylum of Thermoplasmatota. Due to its underrepresentation in public databases, a data-mining strategy was developed, including mining genome assemblies and metagenomic short read data. This increased the representation of the EX4484-6 class extensively. The class EX4484-6 was classified as a member of the rare biosphere. As such, it performed anaerobic heterotrophic protein and amino acid degradation. An unexpectedly high number of unknown genes was detected within the new class, indicating taxonomic distance to other Thermoplasmatota classes.

Chapter IV: Diverse methoxydotrophic acetogens facilitate terrestrial carbon in anoxic marine sediments. In this chapter, RNA-SIP and enrichment experiments were conducted to detect organisms growing on methoxy groups of methoxylated aromatic compounds (MACs) and identify their substrate spectrum. Clostridia were detected as the sole methoxydotrophs, utilizing up to nine different compounds. The degradation of MACs was supported by a decrease of MACs, measured by LC-MS/MS. Using real-time quantitative PCR, the growth of Clostridia was tracked throughout the incubation time, and high abundances were observed. Clostridia of the orders Acetivibrionales, Eubacteriales, Oscillospirales and Tissierellales were involved in methoxydotrophy via a reductive acetyl-CoA pathway, as shown by metagenomic analyses. A comparison of these Clostridia with other Clostridia representatives indicated a high abundance of methyltransferases and, therefore, the involvement of Clostridia in C_1-carbon compound metabolism.

1.9 References

Adair, K. and Schwartz, E. (2011). Chapter seven - Stable isotope probing with ^{18}O-water to investigate growth and mortality of ammonia oxidizing bacteria and archaea in soil. *In:* M. G. Klotz, editor: Methods in Enzymology. Academic Press. p. 155-169. doi: 10.1016/B978-0-12-381294-0.00007-9.

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F. and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. Nat. Methods **11**:1144-1146. doi: 10.1038/nmeth.3103.

Andreotti, G., Cubellis, M. V., Nitti, G., Sannia, G., Mai, X., Adams, M. W. and Marino, G. (1995). An extremely thermostable aromatic aminotransferase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Biochim. Biophys. Acta **1247**:90-96. doi: 10.1016/0167-4838(94)00211-x.

Andreotti, G., Cubellis, M. V., Nitti, G., Sannia, G., Mai, X., Marino, G. and Adams, M. W. (1994). Characterization of aromatic aminotransferases from the hyperthermophilic archaeon *Thermococcus litoralis*. Eur. J. Biochem. **220**:543-549. doi: 10.1111/j.1432-1033.1994.tb18654.x.

Armstrong, R. A., Lee, C., Hedges, J. I., Honjo, S. and Wakeham, S. G. (2001). A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals. Deep-Sea Res. II: Top. Stud. Oceanogr. **49**:219-236. doi: 10.1016/S0967-0645(01)00101-1.

Armstrong, S. M. and Patel, T. R. (1994). Microbial degradation of phloroglucinol and other polyphenolic compounds. J. Basic Microbiol. **34**:123-135. doi: 10.1002/jobm.3620340208.

Arndt, S., Jørgensen, B. B., LaRowe, D. E., Middelburg, J. J., Pancost, R. D. and Regnier, P. (2013). Quantifying the degradation of organic matter in marine sediments: a review and synthesis. Earth-Sci. Rev. **123**:53-86. doi: 10.1016/j.earscirev.2013.02.008.

Arnosti, C. (2004). Speed bumps and barricades in the carbon cycle: substrate structural effects on carbon cycling. Mar. Chem. **92**:263-273. doi: 10.1016/j.marchem.2004.06.030.

Atwood, T. B., Witt, A., Mayorga, J., Hammill, E. and Sala, E. (2020). Global patterns in marine sediment carbon stocks. Front. Mar. Sci. **7**. doi: 10.3389/fmars.2020.00165.

Berner, R. A. (1982). Burial of organic carbon and pyrite sulfur in the modern ocean: its geochemical and environmental significance. Am. J. Sci. **282**:451-473. doi: 10.1016/0016-7037(83)90151-5.

Bianchi, T. S., Cui, X., Blair, N. E., Burdige, D. J., Eglinton, T. I. and Galy, V. (2018). Centers of organic carbon burial and oxidation at the land-ocean interface. Org. Geochem. **115**:138-155. doi: 10.1016/j.orggeochem.2017.09.008.

Biddle, J. F., Lipp, J. S., Lever, M. A., Lloyd, K. G., Sørensen, K. B., Anderson, R., Fredricks, H. F., Elvert, M., Kelly, T. J., Schrag, D. P., et al. (2006). Heterotrophic archaea dominate sedimentary subsurface ecosystems off Peru. Proc. Natl. Acad. Sci. U. S. A. **103**:3846-3851. doi: 10.1073/pnas.0600035103.

Blair, N. E. and Aller, R. C. (2012). The fate of terrestrial organic carbon in the marine environment. Ann. Rev. Mar. Sci. **4**:401-423. doi: 10.1146/annurev-marine-120709-142717.

Blamey, J. M. and Adams, M. W. (1993). Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Biochim. Biophys. Acta **1161**:19-27. doi: 10.1016/0167-4838(93)90190-3.

Boerjan, W., Ralph, J. and Baucher, M. (2003). Lignin biosynthesis. Annu. Rev. Plant Biol. **54**:519-546. doi: 10.1146/annurev.arplant.54.031902.134938.

Boll, M. and Fuchs, G. (1995). Benzoyl-coenzyme A reductase (dearomatizing), a key enzyme of anaerobic aromatic metabolism. ATP dependence of the reaction, purification and some properties of the enzyme from *Thauera aromatica* strain K172. Eur. J. Biochem. **234**:921-933. doi: 10.1111/j.1432-1033.1995.921_a.x.

Braun, S., Mhatre, S. S., Jaussi, M., Røy, H., Kjeldsen, K. U., Pearce, C., Seidenkrantz, M. S., Jørgensen, B. B. and Lomstein, B. A. (2017). Microbial turnover times in the deep seabed studied by amino acid racemization modelling. Sci. Rep. **7**:5680. doi: 10.1038/s41598-017-05972-z.

Brown, M. E. and Chang, M. C. Y. (2014). Exploring bacterial lignin degradation. Curr. Opin. Chem. Biol. **19**:1-7. doi: 10.1016/j.cbpa.2013.11.015.

Bugg, T. D. H., Ahmad, M., Hardiman, E. M. and Rahmanpour, R. (2011a). Pathways for degradation of lignin in bacteria and fungi. Nat. Prod. Rep. **28**:1883-1896. doi: 10.1039/C1NP00042J.

Bugg, T. D. H., Ahmad, M., Hardiman, E. M. and Singh, R. (2011b). The emerging role for bacteria in lignin degradation and bio-product formation. Curr. Opin. Biotechnol. **22**:394-400. doi: 10.1016/j.copbio.2010.10.009.

Burdige, D. J. (1993). The biogeochemistry of manganese and iron reduction in marine sediments. Earth-Sci. Rev. **35**:249-284. doi: 10.1016/0012-8252(93)90040-E.

Burdige, D. J. (2005). Burial of terrestrial organic matter in marine sediments: a re-assessment. Global Biogeochem. Cycles **19**. doi: 10.1029/2004GB002368.

Burdige, D. J. (2007). Preservation of organic matter in marine sediments: controls, mechanisms, and an imbalance in sediment organic carbon budgets? Chem. Rev. **107**:467-485. doi: 10.1021/cr050347q.

Canfield, D. E. (1994). Factors influencing organic carbon preservation in marine sediments. Chem. Geol. **114**:315-329. doi: 10.1016/0009-2541(94)90061-2.

Canfield, D. E. and Thamdrup, B. (2009). Towards a consistent classification scheme for geochemical environments, or, why we wish the term 'suboxic' would go away. Geobiology **7**:385-392. doi: 10.1111/j.1472-4669.2009.00214.x.

Carmona, M., Zamarro, M. T., Blázquez, B., Durante-Rodríguez, G., Juárez, J. F., Valderrama, J. A., Barragán, M. J., García, J. L. and Díaz, E. (2009). Anaerobic catabolism of aromatic compounds: a genetic and genomic view. Microbiol. Mol. Biol. Rev. **73**:71-133. doi: 10.1128/mmbr.00021-08.

Compeau, P. E. C., Pevzner, P. A. and Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. p. 987-991 Nat. Biotechnol. doi: 10.1038/nbt.2023.

Consalvi, V., Chiaraluce, R., Politi, L., Vaccaro, R., De Rosa, M. and Scandurra, R. (1991). Extremely thermostable glutamate dehydrogenase from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. Eur. J. Biochem. **202**:1189-1196. doi: 10.1111/j.1432-1033.1991.tb16489.x.

Crawford, D. L. and Crawford, R. L. (1980). Microbial degradation of lignin. Enzyme Microb. Technol. **2**:11-22. doi: 10.1016/0141-0229(80)90003-4.

Dauwe, B., Middelburg, J. J., Herman, P. M. J. and Heip, C. H. R. (1999). Linking diagenetic alteration of amino acids and bulk organic matter reactivity. Limnol. Oceanogr. **44**:1809-1814. doi: 10.4319/lo.1999.44.7.1809.

Devol, A. H. (2015). Denitrification, anammox, and N_2 production in marine sediments. Ann. Rev. Mar. Sci. **7**:403-423. doi: 10.1146/annurev-marine-010213-135040.

Doelle, H. W. (1975). Bacterial metabolism. 2nd edition. Academic Press, New York, London.

Dominik, J., Förstner, U., Mangini, A. and Reineck, H.-E. (1978). Pb and ^{137}Cs chronology of heavy metal pollution in a sediment core from the German Bight (North Sea). Senckenb. Marit. **10**:213-227. doi:

Drake, H. L. and Daniel, S. L. (2004). Physiology of the thermophilic acetogen *Moorella thermoacetica*. Res. Microbiol. **155**:869-883. doi: 10.1016/j.resmic.2004.10.002.

Dumont, M. G. and Murrell, J. C. (2005). Stable isotope probing - linking microbial identity to function. Nat. Rev. Microbiol. **3**:499-504. doi: 10.1038/nrmicro1162.

Emerson, S. (1985). Organic carbon preservation in marine sediments. *In:* E. T. Sundquist and W. S. Broecker, editors: The Carbon Cycle and Atmospheric CO_2: Natural Variations Archean to Present. American Geophysical Union, Washington, DC. p. 78-87. doi: 10.1029/GM032p0078.

Esri. (2009). World Shaded Relief. https://www.arcgis.com/home/item.html?id=9c5370d0b54f4de1b48a3792d7377ff2. (accessed December 20, 2024).

Felske, A., Wolterink, A., Van Lis, R. and Akkermans, A. D. (1998). Phylogeny of the main bacterial 16S rRNA sequences in Drentse A grassland soils (The Netherlands). Appl. Environ. Microbiol. **64**:871-879. doi: 10.1128/AEM.64.3.871-879.1998.

Fisher, A. B. and Fong, S. S. (2014). Lignin biodegradation and industrial implications. AIMS Bioeng. **1**:92-112. doi: 10.3934/bioeng.2014.2.92.

Follett, C. L., Repeta, D. J., Rothman, D. H., Xu, L. and Santinelli, C. (2014). Hidden cycle of dissolved organic carbon in the deep ocean. Proc. Natl. Acad. Sci. U. S. A. **111**:16706-16711. doi: 10.1073/pnas.1407445111.

Fonknechten, N., Chaussonnerie, S., Tricot, S., Lajus, A., Andreesen, J. R., Perchat, N., Pelletier, E., Gouyvenoux, M., Barbe, V., Salanoubat, M., et al. (2010). *Clostridium sticklandii*, a specialist in amino acid degradation: revisiting its metabolism through its genome sequence. BMC Genomics **11**:555. doi: 10.1186/1471-2164-11-555.

Francois, R., Honjo, S., Krishfield, R. and Manganini, S. (2002). Factors controlling the flux of organic carbon to the bathypelagic zone of the ocean. Global Biogeochem. Cycles **16**:34-31-34-20. doi: 10.1029/2001GB001722.

Fuchs, G., Boll, M. and Heider, J. (2011). Microbial degradation of aromatic compounds - from one strategy to four. Nat. Rev. Microbiol. **9**:803-816. doi: 10.1038/nrmicro2652.

Gargulak, J. D., Lebo, S. E. and McNally, T. J. (2015). Lignin. *In*: Kirk-Othmer Encyclopedia of Chemical Technology. John Wiley & Sons, Inc. p. 1-26. doi: 10.1002/0471238961.12090714120914.a01.pub3.

Glasemacher, J., Bock, A. K., Schmid, R. and Schønheit, P. (1997). Purification and properties of acetyl-CoA synthetase (ADP-forming), an archaeal enzyme of acetate formation and ATP synthesis, from the hyperthermophile *Pyrococcus furiosus*. Eur. J. Biochem. **244**:561-567. doi: 10.1111/j.1432-1033.1997.00561.x.

Glud, R. N. (2008). Oxygen dynamics of marine sediments. Mar. Biol. Res. **4**:243-289. doi: 10.1080/17451000801888726.

Grbić-Galić, D. (1983). Anaerobic degradation of coniferyl alcohol by methanogenic consortia. Appl. Environ. Microbiol. **46**:1442-1446. doi: 10.1128/aem.46.6.1442-1446.1983.

Grigoryan A. A., Cornish S. L., Buziak, B., Lin, S., Cavallaro, A., Arensdorf J. J. and Voordouw, G. (2008). Competitive oxidation of volatile fatty acids by sulfate- and nitrate-reducing bacteria from an oil field in Argentina. Appl. Environ. Microbiol. **74**:4324-4335. doi: 10.1128/AEM.00419-08.

Han, Y., Zhang, C., Zhao, Z., Peng, Y., Liao, J., Jiang, Q., Liu, Q., Shao, Z. and Dong, X. (2023). A comprehensive genomic catalog from global cold seeps. Sci. Data **10**:596. doi: 10.1038/s41597-023-02521-4.

Hansell, D. A. (2013). Recalcitrant dissolved organic carbon fractions. Ann. Rev. Mar. Sci. **5**:421-445. doi: 10.1146/annurev-marine-120710-100757.

Hansell, D. A., Carlson, C. A., Repeta, D. J. and Schlitzer, R. (2009). Dissolved organic matter in the ocean a controversy stimulates new insights Oceanography **22**:202-211. doi: 10.5670/oceanog.2009.109.

Hebbeln, D., Scheurle, C. and Lamy, F. (2003). Depositional history of the Helgoland mud area, German Bight, North Sea. Geo-Mar. Lett. **23**:81-90. doi: 10.1007/s00367-003-0127-0.

Hedges, J. I. and Keil, R. G. (1995). Sedimentary organic matter preservation: an assessment and speculative synthesis. Mar. Chem. **49**:81-115. doi: 10.1016/0304-4203(95)00008-F.

Heider, J., Mai, X. and Adams, M. W. (1996). Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A-dependent enzyme involved in peptide fermentation by hyperthermophilic archaea. J. Bacteriol. **178**:780-787. doi: 10.1128/jb.178.3.780-787.1996.

Hertweck, G. (1983). Das Schlickgebiet in der inneren Deutschen Bucht: Aufnahme mit dem Sedimentechographen. Senckenb. Marit. **15**:219-249.

Hoppe, H.-G., Arnosti, C. and Herndl, G. (2002). Ecological significance of bacterial enzymes in the marine environment *in* R. G. Burns and R. P. Dick, editors: Enzymes in the Environment. Marcel Dekker, New York, Basel. p. 73-107. doi: 10.1201/9780203904039.ch3.

Hoshino, T. and Inagaki, F. (2018). Abundance and distribution of archaea in the subseafloor sedimentary biosphere. ISME J. **13**:227-231. doi: 10.1038/s41396-018-0253-3.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., Ise, K., et al. (2016). A new view of the tree of life. Nat. Microbiol. **1**:16048. doi: 10.1038/nmicrobiol.2016.48.

Hugenholtz, P., Goebel, B. M. and Pace, N. R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. **180**:4765-4774. doi: 10.1128/jb.180.18.4765-4774.1998.

Iversen, M. H., Nowald, N., Ploug, H., Jackson, G. A. and Fischer, G. (2010). High resolution profiles of vertical particulate organic matter export off Cape Blanc, Mauritania: degradation processes and ballasting effects. Deep-Sea Res. I: Oceanogr. Res. Pap. **57**:771-784. doi: 10.1016/j.dsr.2010.03.007.

Jørgensen, B. B., Findlay, A. J. and Pellerin, A. (2019). The biogeochemical sulfur cycle of marine sediments. Front. Microbiol. **10**. doi: 10.3389/fmicb.2019.00849.

Jørgensen, B. B., Wenzhöfer, F., Egger, M. and Glud, R. N. (2022). Sediment oxygen consumption: role in the global marine carbon cycle. Earth-Sci. Rev. **228**:103987. doi: 10.1016/j.earscirev.2022.103987.

Kamimura, N., Takahashi, K., Mori, K., Araki, T., Fujita, M., Higuchi, Y. and Masai, E. (2017). Bacterial catabolism of lignin-derived aromatics: new findings in a recent decade: update on bacterial lignin catabolism. Environ. Microbiol. Rep. **9**:679-705. doi: 10.1111/1758-2229.12597.

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H. and Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. Peer J. **7**:e7359. doi: 10.7717/peerj.7359.

Kato, S., Chino, K., Kamimura, N., Masai, E., Yumoto, I. and Kamagata, Y. (2015). Methanogenic degradation of lignin-derived monoaromatic compounds by microbial enrichments from rice paddy field soil. Sci. Rep. **5**:14295. doi: 10.1038/srep14295.

Kaufmann, F., Wohlfarth, G. and Diekert, G. (1997). Isolation of *O*-demethylase, an ether-cleaving enzyme system of the homoacetogenic strain MC. Arch. Microbiol. **168**:136-142. doi: 10.1007/s002030050479.

Kaufmann, F., Wohlfarth, G. and Diekert, G. (1998). *O*-demethylase from *Acetobacterium dehalogenans* - substrate specificity and function of the participating proteins. Eur. J. Biochem. **253**:706-711. doi: 10.1046/j.1432-1327.1998.2530706.x.

Kawasaki, N., Sohrin, R., Ogawa, H., Nagata, T. and Benner, R. (2011). Bacterial carbon content and the living and detrital bacterial contributions to suspended particulate organic carbon in the North Pacific Ocean. Aquat. Microb. Ecol. **62**:165-176. doi: 10.3354/ame01462.

Keil, R. G. and Kirchman, D. L. (1994). Abiotic transformation of labile protein to refractory protein in sea water. Mar. Chem. **45**:187-196. doi: 10.1016/0304-4203(94)90002-7.

Khomyakova, M. A. and Slobodkin, A. I. (2023). Transformation of methoxylated aromatic compounds by anaerobic microorganisms. Microbiol. **92**:97-118. doi: 10.1134/S0026261722603128.

Kim, J., Hwangbo, M., Shih, C.-H. and Chu, K.-H. (2023). Advances and perspectives of using stable isotope probing (SIP)-based technologies in contaminant biodegradation. Water Res. X. **20**:100187. doi: 10.1016/j.wroa.2023.100187.

Kirk, T. K. and Farrell, R. L. (1987). Enzymatic “combustion": the microbial degradation of lignin. Annu. Rev. Microbiol. **41**:465-501. doi: 10.1146/annurev.mi.41.100187.002341.

Klaas, C. and Archer, D. E. (2002). Association of sinking organic matter with various types of mineral ballast in the deep sea: implications for the rain ratio. Global Biogeochem. Cycles **16**:63-61-63-14. doi: 10.1029/2001GB001765.

Kletzin, A. and Adams, M. W. (1996). Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus* and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima*. J. Bacteriol. **178**:248-257. doi: 10.1128/jb.178.1.248-257.1996.

Kremp, F., Poehlein, A., Daniel, R. and Müller, V. (2018). Methanol metabolism in the acetogenic bacterium *Acetobacterium woodii*. Environ. Microbiol. **20**:4369-4384. doi: 10.1111/1462-2920.14356.

Kurth, J. M., Nobu, M. K., Tamaki, H., de Jonge, N., Berger, S., Jetten, M. S. M., Yamamoto, K., Mayumi, D., Sakata, S., Bai, L., et al. (2021). Methanogenic archaea use a bacteria-like methyltransferase system to demethoxylate aromatic compounds. ISME J. **15**:3549-3565. doi: 10.1038/s41396-021-01025-6.

Kurth, J. M., Op den Camp, H. J. M. and Welte, C. U. (2020). Several ways one goal - methanogenesis from unconventional substrates. Appl. Microbiol. Biotechnol. **104**:6839-6854. doi: 10.1007/s00253-020-10724-7.

Laanbroek, H. J. and Pfennig, N. (1981). Oxidation of short-chain fatty acids by sulfate-reducing bacteria in freshwater and in marine sediments. Arch. Microbiol. **128**:330-335. doi: 10.1007/BF00422540.

Lee, H., Feng, X., Mastalerz, M. and Feakins, S. J. (2019). Characterizing lignin: combining lignin phenol, methoxy quantification, and dual stable carbon and hydrogen isotopic techniques. Org. Geochem. **136**:103894. doi: 10.1016/j.orggeochem.2019.07.003.

Lengeler, J. W., Drews, G. and Schlegel, H. G., editors. (1999). Biology of the Prokaryotes. Georg Thieme Verlag, Germany.

Li, J.-L., Duan, L., Wu, Y., Ahmad, M., Yin, L.-Z., Luo, X.-Q., Wang, X., Fang, B.-Z., Li, S.-H., Huang, L.-N., et al. (2022). Unraveling microbe-mediated degradation of lignin and lignin-derived aromatic fragments in the Pearl River Estuary sediments. Chemosphere **296**:133995. doi: 10.1016/j.chemosphere.2022.133995.

Lipp, J. S., Morono, Y., Inagaki, F. and Hinrichs, K.-U. (2008). Significant contribution of archaea to extant biomass in marine subsurface sediments. Nature **454**:991-994. doi: 10.1038/nature07174.

Liu, Y., Ji, M., Yu, T., Zaugg, J., Anesio, A. M., Zhang, Z., Hu, S., Hugenholtz, P., Liu, K., Liu, P., et al. (2022). A genome and gene catalog of glacier microbiomes. Nat. Biotechnol. **40**:1341-1348. doi: 10.1038/s41587-022-01367-2.

Lloyd, K. G., Schreiber, L., Petersen, D. G., Kjeldsen, K. U., Lever, M. A., Steen, A. D., Stepanauskas, R., Richter, M., Kleindienst, S., Lenk, S., et al. (2013). Predominant archaea in marine sediments degrade detrital proteins. Nature **496**:215-218. doi: 10.1038/nature12033.

Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J. and Crosby, L. (2018). Phylogenetically novel uncultured microbial cells dominate earth microbiomes. mSystems **3**. doi: 10.1128/mSystems.00055-18.

Löffler, C., Kuntze, K., Vazquez, J. R., Rugor, A., Kung, J. W., Böttcher, A. and Boll, M. (2011). Occurrence, genes and expression of the W/Se-containing class II benzoyl-coenzyme A reductases in anaerobic bacteria. Environ. Microbiol. **13**:696-709. doi: 10.1111/j.1462-2920.2010.02374.x.

Lomstein, B. A., Langerhuus, A. T., D'Hondt, S., Jørgensen, B. B. and Spivack, A. J. (2012). Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. Nature **484**:101-104. doi: 10.1038/nature10905.

Lutz, M., Dunbar, R. and Caldeira, K. (2002). Regional variability in the vertical flux of particulate organic carbon in the ocean interior. Global Biogeochem. Cycles **16**:11-11-11-18. doi: 10.1029/2000GB001383.

Ma, K., Hutchins, A., Sung, S. J. and Adams, M. W. (1997). Pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Pyrococcus furiosus*, functions as a CoA-dependent pyruvate decarboxylase. Proc. Natl. Acad. Sci. U. S. A. **94**:9608-9613. doi: 10.1073/pnas.94.18.9608.

Madigan, M., Sattley, W., Aiyer, J., Stahl, D. and Buckley, D., editors. (2021). Brock Biology of Microorganisms, Global Edition. Pearson, Germany.

Mai, X. and Adams, M. W. (1994). Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. A new enzyme involved in peptide fermentation. J. Biol. Chem. **269**:16726-16732. doi: 10.1016/S0021-9258(19)89451-6.

Mai, X. and Adams, M. W. (1996a). Characterization of a fourth type of 2-keto acid-oxidizing enzyme from a hyperthermophilic archaeon: 2-ketoglutarate ferredoxin oxidoreductase from *Thermococcus litoralis*. J. Bacteriol. **178**:5890-5896. doi: 10.1128/jb.178.20.5890-5896.1996.

Mai, X. and Adams, M. W. (1996b). Purification and characterization of two reversible and ADP-dependent acetyl coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. **178**:5897-5903. doi: 10.1128/jb.178.20.5897-5903.1996.

Manefield, M., Whiteley, A. S., Griffiths, R. I. and Bailey, M. J. (2002). RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. Appl. Environ. Microbiol. **68**:5367-5373. doi: 10.1128/AEM.68.11.5367-5373.2002.

Matsui, I., Matsui, E., Sakai, Y., Kikuchi, H., Kawarabayasi, Y., Ura, H., Kawaguchi, S., Kuramitsu, S. and Harata, K. (2000). The molecular structure of hyperthermostable aromatic aminotransferase with novel substrate specificity from *Pyrococcus horikoshii*. J. Biol. Chem. **275**:4871-4879. doi: 10.1074/jbc.275.7.4871.

Middelburg, J. J. (2018). Reviews and syntheses: to the bottom of carbon processing at the seafloor. Biogeosciences **15**:413-427. doi: 10.5194/bg-15-413-2018.

Middelburg, J. J., Vlug, T., Jaco, F. and van der Nat, W. A. (1993). Organic matter mineralization in marine systems. Global Planet. Change **8**:47-58. doi: 10.1016/0921-8181(93)90062-S.

Müller, D., Liu, B., Geibert, W., Holtappels, M., Sander, L., Miramontes, E., Taubner, H., Henkel, S., Hinrichs, K. U., Bethke, D., et al. (2024). Depositional controls and budget of organic carbon burial in fine-grained sediments of the North Sea - the Helgoland mud area as a test field. EGUsphere **2024**:1-37. doi: 10.5194/egusphere-2024-1632.

Nagpal, S., Chuichulcherm, S., Livingston, A. and Peeva, L. (2000). Ethanol utilization by sulfate-reducing bacteria: an experimental and modeling study. Biotechnol. Bioeng. **70**:533-543. doi: 10.1002/1097-0290(20001205)70:53.3.CO;2-3.

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I. M., Huntemann, M., et al. (2021). A genomic catalog of Earth's microbiomes. Nat. Biotechnol. **39**:499-509. doi: 10.1038/s41587-020-0718-6.

Neidhardt, F. C., Ingraham, J. L. and Schaechter, M., editors. (1990). Physiology of the bacterial cell. A molecular approach. Sinauer Associates Inc., U.S., Sunderland, MA.

Neufeld, J. D., Dumont, M. G., Vohra, J. and Murrell, J. C. (2007). Methodological considerations for the use of stable isotope probing in microbial ecology. Microb. Ecol. **53**:435-442. doi: 10.1007/s00248-006-9125-x.

Nisman, B. (1954). The Stickland reaction. Bacteriol. Rev. **18**:16-42. doi: 10.1128/br.18.1.16-42.1954.

Nissen, J. N., Johansen, J., Allesøe, R. L., Sønderby, C. K., Armenteros, J. J. A., Grønbech, C. H., Jensen, L. J., Nielsen, H. B., Petersen, T. N., Winther, O., et al. (2021). Improved metagenome binning and assembly using deep variational autoencoders. Nat. Biotechnol. **39**:555-560. doi: 10.1038/s41587-020-00777-4.

Nunn, B. L., Norbeck, A. and Keil, R. G. (2003). Hydrolysis patterns and the production of peptide intermediates during protein degradation in marine systems. Mar. Chem. **83**:59-73. doi: 10.1016/S0304-4203(03)00096-3.

Oni, O. E., Schmidt, F., Miyatake, T., Kasten, S., Witt, M., Hinrichs, K.-U. and Friedrich, M. W. (2015). Microbial communities and organic matter composition in surface and subsurface sediments of the Helgoland mud area, North Sea. Front. Microbiol. **6**. doi: 10.3389/fmicb.2015.01290.

Orsi, W. D., Richards, T. A. and Francis, W. R. (2018). Predicted microbial secretomes and their target substrates in marine sediment. Nat. Microbiol. **3**:32-37. doi: 10.1038/s41564-017-0047-9.

Orsi, W. D., Schink, B., Buckel, W. and Martin, W. F. (2020). Physiological limits to life in anoxic subseafloor sediment. FEMS Microbiol. Rev. **44**:219-231. doi: 10.1093/femsre/fuaa004.

Ozawa, Y., Nakamura, T., Kamata, N., Yasujima, D., Urushiyama, A., Yamakura, F., Ohmori, D. and Imai, T. (2005). *Thermococcus profundus* 2-ketoisovalerate ferredoxin oxidoreductase, a key enzyme in the archaeal energy-producing amino acid metabolic pathway. J. Biochem. **137**:101-107. doi: 10.1093/jb/mvi012.

Ozawa, Y., Siddiqui, M. A., Takahashi, Y., Urushiyama, A., Ohmori, D., Yamakura, F., Arisaka, F. and Imai, T. (2012). Indolepyruvate ferredoxin oxidoreductase: An oxygen-sensitive iron-sulfur enzyme from the hyperthermophilic archaeon *Thermococcus profundus*. J. Biosci. Bioeng. **114**:23-27. doi: 10.1016/j.jbiosc.2012.02.014.

Pace, N. R., Stahl, D. A., Lane, D. J. and Olsen, G. J. (1986). The analysis of natural microbial populations by ribosomal RNA sequences. *In:* K. C. Marshall, editor. Advances in Microbial Ecology. Springer US, Boston, MA. p. 1-55. doi: 10.1007/978-1-4757-0611-6_1.

Paoli, L., Ruscheweyh, H.-J., Forneris, C. C., Hubrich, F., Kautsar, S., Bhushan, A., Lotti, A., Clayssen, Q., Salazar, G., Milanese, A., et al. (2022). Biosynthetic potential of the global ocean microbiome. Nature **607**:111-118. doi: 10.1038/s41586-022-04862-3.

Parkes, R. J., Cragg, B., Roussel, E., Webster, G., Weightman, A. and Sass, H. (2014). A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere-geosphere interactions. Mar. Geol. **352**:409-425. doi: 10.1016/j.margeo.2014.02.009.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A. and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. **36**:996-1004. doi: 10.1038/nbt.4229.

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P. and Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. **2**:1533-1542. doi: 10.1038/s41564-017-0012-7.

Patel, T. R., Jure, K. G. and Jones, G. A. (1981). Catabolism of phloroglucinol by the rumen anaerobe *Coprococcus*. Appl. Environ. Microbiol. **42**:1010-1017. doi: 10.1128/aem.42.6.1010-1017.1981.

Pelikan, C., Wasmund, K., Glombitza, C., Hausmann, B., Herbold, C. W., Flieder, M. and Loy, A. (2021). Anaerobic bacterial degradation of protein and lipid macromolecules in subarctic marine sediment. ISME J. **15**:833-847. doi: 10.1038/s41396-020-00817-6.

Peng, Q., Lin, L., Tu, Q., Wang, X., Zhou, Y., Chen, J., Jiao, N. and Zhou, J. (2023). Unraveling the roles of coastal bacterial consortia in degradation of various lignocellulosic substrates. mSystems **8**:e01283-01222. doi: 10.1128/msystems.01283-22.

Philipp, B. and Schink, B. (1998). Evidence of two oxidative reaction steps initiating anaerobic degradation of resorcinol (1,3-dihydroxybenzene) by the denitrifying bacterium *Azoarcus anaerobius*. J. Bacteriol. **180**:3644-3649. doi: 10.1128/jb.180.14.3644-3649.1998.

Pierce, E., Xie, G., Barabote, R. D., Saunders, E., Han, C. S., Detter, J. C., Richardson, P., Brettin, T. S., Das, A. and Ljungdahl, L. G. (2008). The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). Environ. Microbiol. **10**:2550-2573. doi: 10.1111/j.1462-2920.2008.01679.x.

QGIS Development Team. (2024). QGIS Geographic Information System. https://www.qgis.org.

Qiu, Z., Yuan, L., Lian, C.-A., Lin, B., Chen, J., Mu, R., Qiao, X., Zhang, L., Xu, Z., Fan, L., et al. (2024). BASALT refines binning from metagenomic data and increases resolution of genome-resolved metagenomic analysis. Nat. Commun. **15**:2179. doi: 10.1038/s41467-024-46539-7.

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. Nat. Biotechnol. **35**:833-844. doi: 10.1038/nbt.3935.

Radajewski, S., Ineson, P., Parekh, N. R. and Murrell, J. C. (2000). Stable isotope probing as a tool in microbial ecology. Nature **403**:646-649. doi: 10.1038/35001054.

Rahman, R. N., Fujiwara, S., Takagi, M. and Imanaka, T. (1998). Sequence analysis of glutamate dehydrogenase (GDH) from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1 and comparison of the enzymatic characteristics of native and recombinant GDHs. Mol. Gen. Genet. **257**:338-347. doi: 10.1007/s004380050655.

Rashid, G. M., Taylor, C. R., Liu, Y., Zhang, X., Rea, D., Fülöp, V. and Bugg, T. D. (2015). Identification of manganese superoxide dismutase from *Sphingobacterium* sp. T2 as a novel bacterial enzyme for lignin oxidation. ACS Chem. Biol. **10**:2286-2294. doi: 10.1021/acschembio.5b00298.

Reineck, H. E. (1963). Sedimentgefüge im Bereich der südlichen Nordsee. Abhandlungen der Senckenbergischen Naturforschenden Gesellschaft, Band 505.

Richards, O. C. and Boyer, P. D. (1966). ^{18}O Labeling of deoxyribonucleic acid during synthesis and stability of the label during replication. J. Mol. Biol. **19**:109-119. doi: 10.1016/S0022-2836(66)80053-0.

Rodríguez del Río, Á., Giner-Lamia, J., Cantalapiedra, C. P., Botas, J., Deng, Z., Hernández-Plaza, A., Munar-Palmer, M., Santamaría-Hernando, S., Rodríguez-Herva, J. J., Ruscheweyh, H.-J., et al. (2024). Functional and evolutionary significance of unknown genes from uncultivated taxa. Nature **626**:377-384. doi: 10.1038/s41586-023-06955-z.

Sakuraba, H., Kawakami, R., Takahashi, H. and Ohshima, T. (2004). Novel archaeal alanine:glyoxylate aminotransferase from *Thermococcus litoralis*. J. Bacteriol. **186**:5513-5518. doi: 10.1128/jb.186.16.5513-5518.2004.

Sánchez, C. (2009). Lignocellulosic residues: biodegradation and bioconversion by fungi. Biotechnol. Adv. **27**:185-194. doi: 10.1016/j.biotechadv.2008.11.001.

Schäfer, T., Selig, M. and Schönheit, P. (1993). Acetyl-CoA synthetase (ADP forming) in archaea, a novel enzyme involved in acetate formation and ATP synthesis. Arch. Microbiol. **159**:72-83. doi: 10.1007/BF00244267

Schildkraut, C. L., Marmur, J. and Doty, P. (1962). Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. J. Mol. Biol. **4**:430-443. doi: 10.1016/S0022-2836(62)80100-4.

Schink, B. (2006). The genus *Propionigenium.in* M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt, editors. The Prokaryotes: Volume 7: Proteobacteria: Delta, Epsilon Subclass. Springer New York, New York, NY. p. 955-959. doi: 10.1007/0-387-30747-8_41.

Schink, B., Philipp, B. and Müller, J. A. (2000). Anaerobic degradation of phenolic compounds. Naturwissenschaften. doi: 10.1007/s001140050002.

Schlesinger, W. H. and Melack, J. M. (1981). Transport of organic carbon in the world's rivers. Tellus **33**:172-187. doi: 10.3402/tellusa.v33i2.10706.

Schut, G. J., Menon, A. L. and Adams, M. W. W. (2001). 2-keto acid oxidoreductases from *Pyrococcus furiosus* and *Thennococcus litoralis*. Methods in Enzymology **331**:144-158. doi: 10.1016/S0076-6879(01)31053-4.

Schwartz, E., Hayer, M., Hungate, B. A., Koch, B. J., McHugh, T. A., Mercurio, W., Morrissey, E. M. and Soldanova, K. (2016). Stable isotope probing with ^{18}O-water to investigate microbial growth and death in environmental samples. Curr. Opin. Biotechnol. **41**:14-18. doi: 10.1016/j.copbio.2016.03.003.

Shikata, K., Fukui, T., Atomi, H. and Imanaka, T. (2007). A novel ADP-forming succinyl-CoA synthetase in *Thermococcus kodakaraensis* structurally related to the archaeal nucleoside diphosphate-forming acetyl-CoA synthetases. J. Biol. Chem. **282**:26963-26970. doi: 10.1074/jbc.M702694200.

Siddiqui, M. A., Fujiwara, S. and Imanaka, T. (1997). Indolepyruvate ferredoxin oxidoreductase from *Pyrococcus* sp. KOD1 possesses a mosaic structure showing features of various oxidoreductases. Mol. Gen. Genet. **254**:433-439. doi: 10.1007/pl00008607.

Sievers, J., Rubel, M. and Milbradt, P. (2020). EasyGSH-DB: Themengebiet - Geomorphologie [Data set]. Bundesanstalt für Wasserbau.

Smith, E. T., Blamey, J. M. and Adams, M. W. (1994). Pyruvate ferredoxin oxidoreductases of the hyperthermophilic archaeon, *Pyrococcus furiosus*, and the hyperthermophilic bacterium, *Thermotoga maritima*, have different catalytic mechanisms. Biochem. **33**:1008-1016. doi: 10.1021/bi00170a020.

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L. and Ettema, T. J. G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature **521**:173-179. doi: 10.1038/nature14447.

Stumm, W. and Morgan, J. J. (2013). Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters. John Wiley & Sons, Inc.

Thamdrup, B., Fossing, H. and Jørgensen, B. B. (1994). Manganese, iron and sulfur cycling in a coastal marine sediment, Aarhus bay, Denmark. Geochim. Cosmochim. Acta **58**:5115-5129. doi: 10.1016/0016-7037(94)90298-4.

Tian, J.-H., Pourcher, A.-M., Bouchez, T., Gelhaye, E. and Peu, P. (2014). Occurrence of lignin degradation genotypes and phenotypes among prokaryotes. Appl. Microbiol. Biotechnol. **98**:9527-9544. doi: 10.1007/s00253-014-6142-4.

Tschech, A. and Schink, B. (1985). Fermentative degradation of resorcinol and resorcylic acids. Arch. Microbiol. **143**:52-59. doi: 10.1007/BF00414768.

Uritskiy, G. V., DiRuggiero, J. and Taylor, J. (2018). MetaWRAP - a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome **6**:158. doi: 10.1186/s40168-018-0541-1.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J. and Boerjan, W. (2010). Lignin biosynthesis and structure. Plant Physiol. **153**:895-905. doi: 10.1104/pp.110.155119.

Wagner, R. (1994). The regulation of ribosomal RNA synthesis and bacterial cell growth. Arch. Microbiol. **161**:100-109. doi: 10.1007/bf00276469.

Wakeham, S. G., Hedges, J. I., Lee, C., Peterson, M. L. and Hernes, P. J. (1997a). Compositions and transport of lipid biomarkers through the water column and surficial sediments of the equatorial Pacific Ocean. Deep-Sea Res. II: Top. Stud. Oceanogr. **44**:2131-2162. doi: 10.1016/S0967-0645(97)00035-0.

Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J. and Peterson, M. J. (1997b). Molecular indicators of diagenetic status in marine organic matter. Geochim. Cosmochim. Acta **61**:5363-5369. doi: 10.1016/S0016-7037(97)00312-8.

Wang, X., Lin, L. and Zhou, J. (2021). Links among extracellular enzymes, lignin degradation and cell growth establish the models to identify marine lignin-utilizing bacteria. Environ. Microbiol. **23**:160-173. doi: 10.1111/1462-2920.15289.

Ward, D. E., de Vos, W. M. and van der Oost, J. (2002). Molecular analysis of the role of two aromatic aminotransferases and a broad-specificity aspartate aminotransferase in the aromatic amino acid metabolism of *Pyrococcus furiosus*. Archaea **1**:133-141. doi: 10.1155/2002/959031.

Ward, D. E., Kengen, S. W., van Der Oost, J. and de Vos, W. M. (2000). Purification and characterization of the alanine aminotransferase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its role in alanine production. J. Bacteriol. **182**:2559-2566. doi: 10.1128/jb.182.9.2559-2566.2000.

Ward, D. M., Weller, R. and Bateson, M. M. (1990). 16S rRNA sequences reveal uncultured inhabitants of a well-studied thermal community. FEMS Microbiol. Rev. **6**:105-115. doi: 10.1111/j.1574-6968.1990.tb04088.x.

Welte, C. U., de Graaf, R., Dalcin Martins, P., Jansen, R. S., Jetten, M. S. M. and Kurth, J. M. (2021). A novel methoxydotrophic metabolism discovered in the hyperthermophilic archaeon *Archaeoglobus fulgidus*. Environ. Microbiol. **23**:4017-4033. doi: 10.1111/1462-2920.15546.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., Whitman, W. B., Euzéby, J., Amann, R. and Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat. Rev. Microbiol. **12**:635-645. doi: 10.1038/nrmicro3330.

Yin, X., Zhou, G., Cai, M., Zhu, Q.-Z., Richter-Heitmann, T., Aromokeye, D. A., Liu, Y., Nimzyk, R., Zheng, Q., Tang, X., et al. (2022). Catabolic protein degradation in marine sediments confined to distinct archaea. ISME J. **16**:1617-1626. doi: 10.1038/s41396-022-01210-1.

Yokooji, Y., Sato, T., Fujiwara, S., Imanaka, T. and Atomi, H. (2013). Genetic examination of initial amino acid oxidation and glutamate catabolism in the hyperthermophilic archaeon *Thermococcus kodakarensis*. J. Bacteriol. **195**:1940-1948. doi: 10.1128/jb.01979-12.

Yue, Y., Huang, H., Qi, Z., Dou, H.-M., Liu, X.-Y., Han, T.-F., Chen, Y., Song, X.-J., Zhang, Y.-H. and Tu, J. (2020). Evaluating metagenomics tools for genome binning with real metagenomic datasets and CAMI datasets. BMC Bioinformatics **21**:334. doi: 10.1186/s12859-020-03667-3.

Zhou, Y., Wei, Y., Jiang, L., Jiao, X. and Zhang, Y. (2023). Anaerobic phloroglucinol degradation by *Clostridium scatologenes*. mBio **14**:e01099-01023. doi: 10.1128/mbio.01099-23.

Zoghlami, A. and Paës, G. (2019). Lignocellulosic biomass: Understanding recalcitrance and predicting hydrolysis. Front. Chem. **7**. doi: 10.3389/fchem.2019.00874.

Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. Microbiol. Mol. Biol. Rev. **61**:533-616. doi: 10.1128/mmbr.61.4.533-616.1997.

Chapter II

Linking sulfur and carbon cycling with $H_2{}^{18}O$ RNA-SIP: active taxa and metabolic interactions in anoxic marine sediments

Mara Maeke, Xiuran Yin, Michael W. Friedrich

Manuscript in preparation

Running title:
In situ activity in the HMA

Contribution to the manuscript:

Experimental concept and design	50%
Acquisition of experimental data	60%
Data analysis and interpretation	70%
Preparation of figures and tables	100%
Drafting of manuscript	90%

Linking sulfur and carbon cycling with $H_2{}^{18}O$ RNA-SIP: active taxa and metabolic interactions in anoxic marine sediments

Mara Maeke[1], Xiuran Yin[1,2,3*], Michael W. Friedrich[1,3*]

[1]Microbial Ecophysiology Group, Faculty of Biology/Chemistry, University of Bremen, Bremen, Germany

[2]State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou, China

[3]MARUM – Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany

Correspondence *
Prof. Dr. Michael W. Friedrich
University of Bremen, Microbial Ecophysiology Group
James-Watt-Straße 1 28359 Bremen, Germany
Phone: +49 (0) 421 218-63060
Email: michael.friedrich@uni-bremen.de

Dr. Xiuran Yin
Hainan University, State Key Laboratory of Marine Resource Utilization in South China Sea
Renmin Ave. No.58 570228 Haikou, China
Email: 996383@hainanu.edu.cn

2.1.1 Abstract

The sulfur and carbon cycle are closely interconnected in anoxic marine sediments, with sulfate-reducing organisms playing an essential role in remineralizing up to 50% of organic matter. Yet, due to the cryptic sulfur cycling, processes of sulfur compound cycling are complex to track. Such complexity is partially attributed to sulfur compound disproportionation which drives formation of multiple forms of sulfur compounds. However, the importance of sulfur disproportionation in marine sediments remains largely unexplored, particularly concerning the active player, biochemical pathways and functional genes involved. Therefore, we developed a small-scale $H_2{}^{18}O$ RNA-based stable isotope probing (RNA-SIP) to identify active microorganisms and utilized their taxonomic affiliation to elaborate on their metabolic roles in sulfur cycling. The ^{18}O labeling revealed that 70% of the active microorganisms are sulfur compound cycling organisms of the families Desulfocapsaceae, Desulfobulbaceae, and Desulfurivibrionaceae. Key active taxa included the sulfur-disproportionating bacteria of the genus *Desulfocapsa*. Unclassified members of the Desulfocapsaceae, Desulfurivibrionaceae and Desulfobulbaceae showed potential for sulfur compound reduction. Besides active sulfur compound cycling organisms, the active heterotrophic iron-cycling organism Sva1033 was detected, linking the sulfur and iron cycle. Moreover, active secondary fermenters were detected, most likely supplying the sulfur compound and iron-cycling organisms with electron donors as fermentation products.

2.1.2 Main

The sulfur and carbon cycle are directly linked in anaerobic marine sediments (Wasmund et al. 2017, Jørgensen 2021). Sulfur compound cycling microorganisms use fermentation products derived from the breakdown of organic matter in a terminal oxidation process (Wasmund et al. 2017, Jørgensen 2021). Specifically, sulfate-reducing microorganisms are estimated to account for up to 50% of organic matter remineralization in marine sediments and are regarded as drivers of the biogeochemical cycling of sulfur compounds (Jørgensen 1982). Yet, due to the rapid cycling of sulfur compounds and low pool sizes involved, processes within the sulfur cycle are hard to track geochemically and are therefore described as a cryptic sulfur cycle (Wasmund et al. 2017, Jørgensen 2021). About 80-90% of the sulfide produced during sulfate reduction is reoxidized to sulfate or other intermediate sulfur compounds, e.g. sulfite, elemental sulfur, polysulfides, or thiosulfate (Wasmund et al. 2017). The remaining sulfide is buried after reaction with metal ions, e.g., as pyrite (FeS_2), interconnecting the sulfur and iron cycle (Wasmund et al. 2017, Jørgensen 2021). Besides the reduction of sulfate or sulfur cycle

intermediates and sulfur compound oxidation, the process of sulfur compound disproportionation plays an important role in marine sediments (Wasmund et al. 2017, Jørgensen 2021). Sulfur compound disproportionation by bacteria was first detected by Bak and Cypionka (1987) and until now remains an understudied process. During disproportionation elemental sulfur, thiosulfate and sulfite serve as both, electron donor and electron acceptor (Bak and Cypionka 1987, Thamdrup et al. 1993). A few studies have enumerated disproportionating bacteria by most probable number cultivation, e.g., over 10^5 thiosulfate-disproportionating bacteria per cm^{-3} in upper layers of Aarhus Bay sediment (Jørgensen and Bak 1991). By using elegant radiotracer experiments, the process was shown to be significant in marine sediments, but experiments are expensive (e.g., custom synthesis of S-35 labeled thiosulfate), and can be affected by S-isotope exchange, especially with elemental sulfur (Jørgensen 2021). Microbial elemental sulfur disproportionation is accompanied by strong isotope fractionation of S-34 in both, sulfide (depleted) and sulfate (enriched) formed, but only triple sulfur isotope analysis (S-32, S-33, S-34) can discriminate co-occurring sulfate reduction and disproportionation in natural systems (Jørgensen 2021). Active sulfur disproportionating microorganisms are difficult to identify in the environment, since biochemical pathways of disproportionation are still elusive and thus, distinct functional marker genes cannot be used so far, and the identification of disproportionating species has so far been limited to cultivated strains (Slobodkin and Slobodkina 2019).

Here, we applied small-scale $H_2{}^{18}O$ RNA-based stable isotope probing (RNA-SIP) to identify microorganisms involved in sulfur compound cycling in unamended anaerobic marine sediments. SIP with $H_2{}^{18}O$ as isotopic tracer (Schwartz et al. 2016), is an alternative to ^{13}C-based SIP approaches that does not require knowledge of the carbon source to identify active microorganisms under *in situ* conditions. Active microbial taxa, i.e. those having an energy source, will incorporate $H_2{}^{18}O$ thereby generating ^{18}O-labeled, heavy RNA, which can be density separated by gradient centrifugation from unlabeled nucleic acids, and subsequently, active microbes, even at very low abundance, can be identified by Illumina amplicon sequencing (Aoyagi et al. 2015). Our SIP incubations showed that ^{18}O amendment labeled taxa known for sulfur cycling capabilities to a high extent. Using their taxonomic affiliation, we elaborated on their possible metabolic interactions in Helgoland mud area sediments.

To gain insights into the activity of microbial communities in anaerobic marine sediment samples of the Helgoland mud area (0-25 cm depth), we extracted RNA after 7 days of incubation from four biological replicates amended with either artificial sea water (ASW, unlabeled) or labeled ^{18}O-ASW. For retrieval of sufficient RNA, replicates were pooled and,

thus, reflect an average of the communities within single treatments. Density profiles of RNA-SIP incubations showed labeling within the ^{18}O-ASW treatment (Figure S1, Supplementary results). In bacterial samples, the sequence analysis of density-separated RNA-SIP fractions of ^{18}O-ASW treated samples showed a clear distinction between "heavy" (active) and "light" (non-active) fractions, compared to the unlabeled ASW control (Fig. 1a, Fig. S1). Archaea sequenced in this study were not considered active as they could not be detected in heavy fractions due to low community coverage (Fig. S2, Supplementary results). Amplicon sequence variants (ASVs) found in the ultra-heavy and heavy fractions, and hereafter considered active, were predominantly from the phylum Desulfobacterota (Desulfobulbia and Desulfuromonadia; 87.7% relative abundance in the ultra-heavy fraction) (Fig. 1b-d).

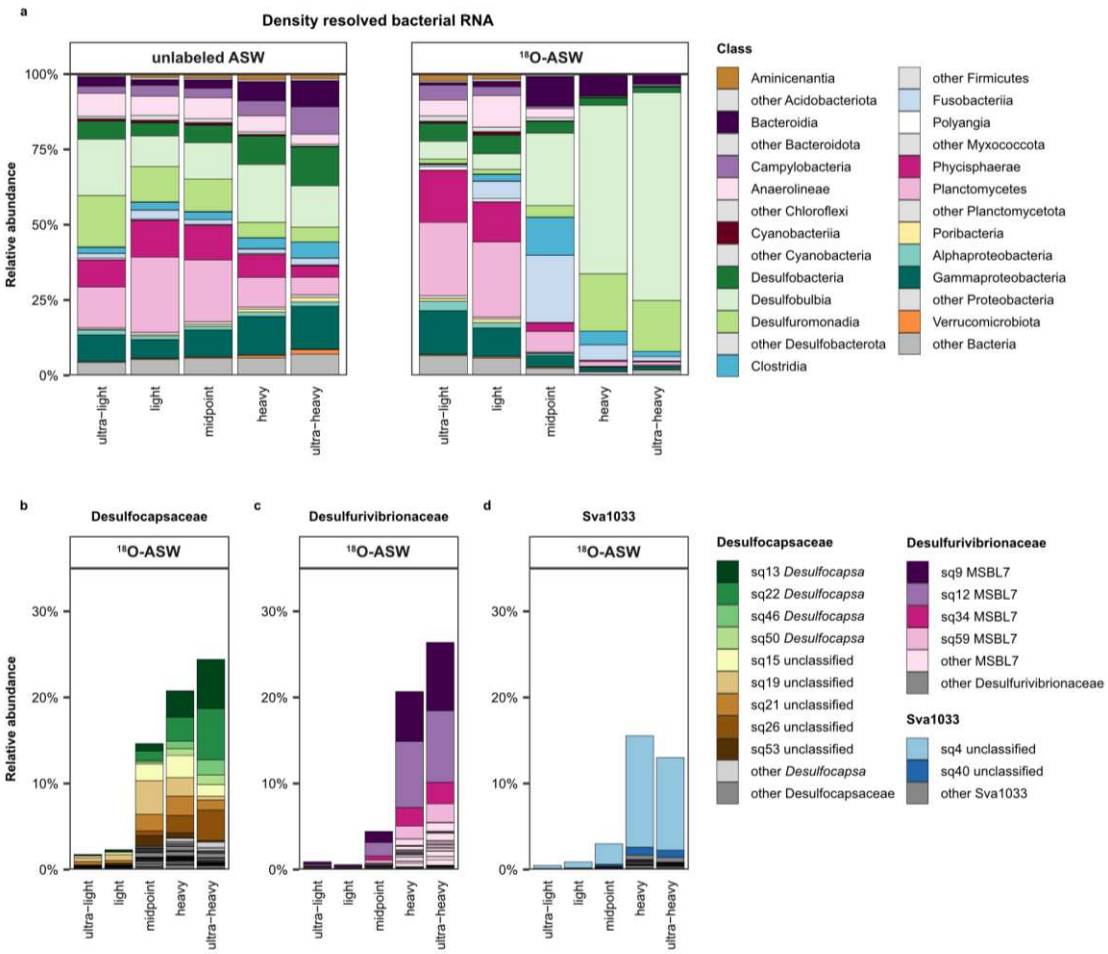

Figure 1 (**a**) Relative abundances of the bacterial RNA of the unlabeled ASW and of the ^{18}O-ASW amended incubations on class level, separated by density fractions (ultra-light, light, midpoint, heavy and ultraheavy) after 7 days of incubation. Relative abundances of amplicon sequence variants (ASVs) in ^{18}O-ASW amended sample after 7 days of incubation by fractions (ultra-light, light, midpoint, heavy and ultra-heavy). Highest relative abundances were observed in ^{18}O-ASW for Desulfobulbia (ultra-heavy (UH): 69 %, heavy (H): 56% vs unlabeled water UH: 13.7 %, H: 19.2 %) and Desulfuromonadia (UH: 16.9 %, H: 19 % vs. unlabeled water UH: 4.9 %, H: 5%). Relative abundances of (**b**) the family Desulfocapsaceae (UH: 24.4 %, H: 20.8 %) with high relative abundances of the genus *Desulfocapsa* (UH: 15.72 %, H: 8.2 %), (**c**) the family Desulfurivibrionaceae (UH: 26.4 %, H: 20.7 %) and (**d.**) the family Sva1033 (UH: 13 %, H: 15.5 %).

Within the Desulfobulbia (69%) two main abundant families were observed: Desulfocapsaceae (24.4%) and Desulfurivibrionaceae (26.4%) (Fig. S3-S5). Further, unclassified Desulfobulbaceae (13.5%) and unclassified Desulfobulbales (4.4%) were found predominantly in heavy fractions (Fig. S6).

Among the Desulfocapsaceae (Fig. 1b), the most active microorganisms (ASVs: sq13, sq22, sq26, sq46, sq50, sq53; 15.7%) were closest related to *Desulfocapsa sulfexigens* (96.41-98.41% 16S rRNA identity) (Fig. S4, Table S1). *Desulfocapsa* species are known to grow by sulfur disproportionation using thiosulfate, sulfite and elemental sulfur (Finster 2008). Sulfur disproportionation of elemental sulfur requires the presence of sulfide-scavenging agents such as dissolved iron or manganese (Thamdrup et al. 1993). Besides, further *Desulfocapsaceae* ASVs unclassified on genus level with highest similarities to *Desulfomarina profundi* (sq15, 98.01%) and *Desulfocastanea catecholica* (sq19 and sq21, 97.21-98.41 %) were identified (Table S1). The shift of these taxa to the heavy (11.6%) and midpoint (10.9%) fraction might reflect an activity, lower than that of *Desulfocapsa*. *Desulfomarina profundi* and *Desulfocastanea catecholica* reduce sulfate, thiosulfate and sulfite with fermentation products or autotrophically; for neither of these microorganisms, sulfur compound disproportionation was reported (Szewzyk and Pfennig 1987, Hashimoto et al. 2021). The sulfate-reducing capability of members of the family Desulfocapsaceae*,* excluding the genus *Desulfocapsa,* was recently demonstrated in incubations with upper sediment layers of the Helgoland mud area (Yin et al. 2024). To unravel the physiology of these unclassified Desulfocapsaceae additional physiological experiments are required.

Active ASVs within the family Desulfurivibrionaceae (Fig. 1c, Fig. S5) were assigned to the genus MSBL7 (26%). The Mediterranean Sea Brine Lake group 7 (MSBL7) was first detected in Mediterranean Sea brines (Borin et al. 2009) and has no cultured representative yet. The closest type strains were *Thiovibrio frasassiensis* (94.44-95.24% 16S rRNA gene identity, Table S2) and *Desulfogranum mediterraneum* (94.02-94.42% 16S rRNA gene, Table S2). Though *Thiovibrio frasassiensis* was described as sulfur disproportionating bacterium by Aronson et al. (2023), the MSBL7 ASVs found were only assigned to the same family and thus not close enough to infer disproportionation. *Desulfogranum mediterraneum* has not been reported as sulfur compound disproportionating bacterium but uses sulfate and thiosulfate as terminal electron acceptors (Sass et al. 2002). Therefore, it remains elusive here whether MSBL7 ASVs were involved in sulfur compound disproportionation. To further unravel the phylogeny of the ASVs found for this group and gain further insights into their possible metabolic potential we constructed a 16S rRNA gene phylogenetic tree using available 16S

rRNA genes of good quality genomes (completeness > 80%, contamination < 5%) present in the GTDB database r214 and compared the sequence identity of the ASVs with all 16S rRNA genes retrieved (Fig. S7, Table S3-S6). The MSBL7 ASVs sq9 and sq59 were highly identical with a 16S rRNA gene found in the genome GCA_002868945.1 (98.805-99.203% 16s rRNA gene identity, Table S5) that was assigned to the genus BM506, following the GTDB taxonomy. In a recent study, the genus BM506 showed the genetic potential to reduce elemental sulfur (Barnum et al. 2018). The ASVs sq12 and sq34 were more closely related to GCA_024277655.1 (97.61-98.41% 16S rRNA gene identity, Table S5) that was assigned to the genus JAJRUT01. The genome sequence GCA_024277655.1 was affiliated with a study by Zhong et al. (2022), in which Desulfurivibrionaceae in general were described as sulfur oxidizers yet also showed potential for sulfate and thiosulfate reduction. Based on this information the ASVs sq12 and sq34 assigned to the MSBL7 are most likely involved in reducing sulfur compounds, with the potential to oxidize sulfur, yet no clear metabolism or electron donor could be determined from this previous study for members of the JAJRUT01 specifically. The presence of genes for sulfur compound reduction and sulfur oxidation could hint at sulfur disproportionation, though not being proven for this group.

ASVs of unclassified Desulfobulbaceae (sq25, sq31, Fig. S6) were also found to be related on family level to the type strains of *Thiovibrio frasassiensis* (93.6 % similarity, Table S2) and *Desulfogranum mediterraneum* (94.02% similarity, Table S2). Comparing the ASV sequences to those 16S rRNA genes obtained from genomes present in the GTDB database, these taxa are closest related to the genera JAHEDT01 (98.4% 16S rRNA gene identity) and JAKITW01 (96.4% 16S rRNA gene identity), both of which have not been described yet. Generally, Desulfobulbaceae use sulfate, sulfite or thiosulfate as electron acceptor, however also sulfur compound disproportionation was described for members of this family (Slobodkin and Slobodkina 2019, Galushko and Kuever 2020). Thus, found Desulfobulbaceae ASVs are involved in sulfur compound cycling. Yet, without further insights into the genomic potential of these taxa, no clear metabolism can be assigned. Since the biochemical pathways of sulfur compound disproportionation remain elusive (Slobodkin and Slobodkina 2019), additional physiological experiments are required to finally describe these unclassified community members.

During the processes of sulfur compound disproportionation and sulfur compound reduction, sulfide is formed, which can form FeS in the presence of ferrous iron (Fe^{2+}) and further transform into pyrite (FeS_2) (Thamdrup et al. 1993, Finster 2008). Dissimilatory iron reducers mediate the availability of such ferrous iron. We detected active ASVs within the class

Desulfuromonadia, which were mostly reflected by the family Sva1033 (13%) (Fig. 1d, Fig. S8-S9). The family Sva1033 of the class Desulfuromonadia was initially detected in the permanently cold sediments from Svalbard (Ravenschlag et al. 1999) and later identified by RNA-SIP as dissimilatory iron reducers in marine sediment using acetate as an electron donor (Wunder et al. 2021). Due to the close relationship with other sulfur cycling microorganisms (e.g., *Desulfuromonas acetoxidans*, 95.62-96.02% 16S rRNA gene identity, Table S2), sulfur cycling cannot be dismissed currently for Sva1033. In previous studies, the presence of acetate primarily stimulated Gammaproteobacteria and Arcobacteraceae (Yin et al. 2024), however these taxa were not identified as active in this present study, raising the question if Sva1033 is actually using acetate in these sediments or instead relying on carbon flux from other active community members.

Further identified heterotrophic taxa were affiliated with Clostridia (*Fusibacter*), Fusobacteriia (*Propionigenium*) and Bacteroidia (*Marinifilum*) and were found only in the midpoint gradient fractions, pointing towards lower activities (Fig. S10-S12, Supplementary results). *Fusibacter, Marinifilum* and *Propionigenium* are known as fermentative bacteria capable to ferment carbohydrates (Schink and Pfennig 1982, Janssen and Liesack 1995, Fadhlaoui et al. 2015, Fu et al. 2018, Brioukhanov et al. 2023). Additionally, *Propionigenium* was shown to ferment succinate, organic and amino acids (Schink and Pfennig 1982, Janssen and Liesack 1995). These fermentative organisms might supply the sediments with fermentation products, which are then coupled to the processes of sulfate, thiosulfate and sulfite reduction by sulfur compound cycling bacteria and dissimilatory iron reduction by Sva1033.

Considering the microorganisms found in the heavy and ultra-heavy fractions as most active, about 70% of these microorganisms are related to taxa that are known to be involved in sulfur compound cycling by either performing sulfur disproportionation or cycling intermediate sulfur species. As the sediment used in this study had not received a supply of fresh organic matter or undergone any infaunal activity since sampling in 2019, the sediment was most likely depleted of possible electron donors. Regardless, we detected sulfur compound cycling as the most dominant process in sediments of the Helgoland mud area. Despite the family Sva1033, potentially involved in heterotrophic dissimilatory iron reduction, other heterotrophic organisms were depleted in heavy fractions and thus might play only minor roles in the analyzed sediment, suggesting that involved organisms are mostly autotrophic. Though only *Desulfocapsa* could be assigned to the process of sulfur disproportionation based on their high identity with other known sulfur disproportionating bacteria, other families, such as the Desulfurivibrionaceae or Desulfobulbaceae could be involved in sulfur compound

disproportionation. Many anaerobic bacteria involved in the respiration of sulfur compounds were also shown to disproportionate sulfur compounds; among these, members of the Desulfobulbaceae and Desulfocapsaceae were detected (Wasmund et al. 2017, Slobodkin and Slobodkina 2019). A reconstruction of possible metabolic pathways within the sediment is shown in Figure 2.

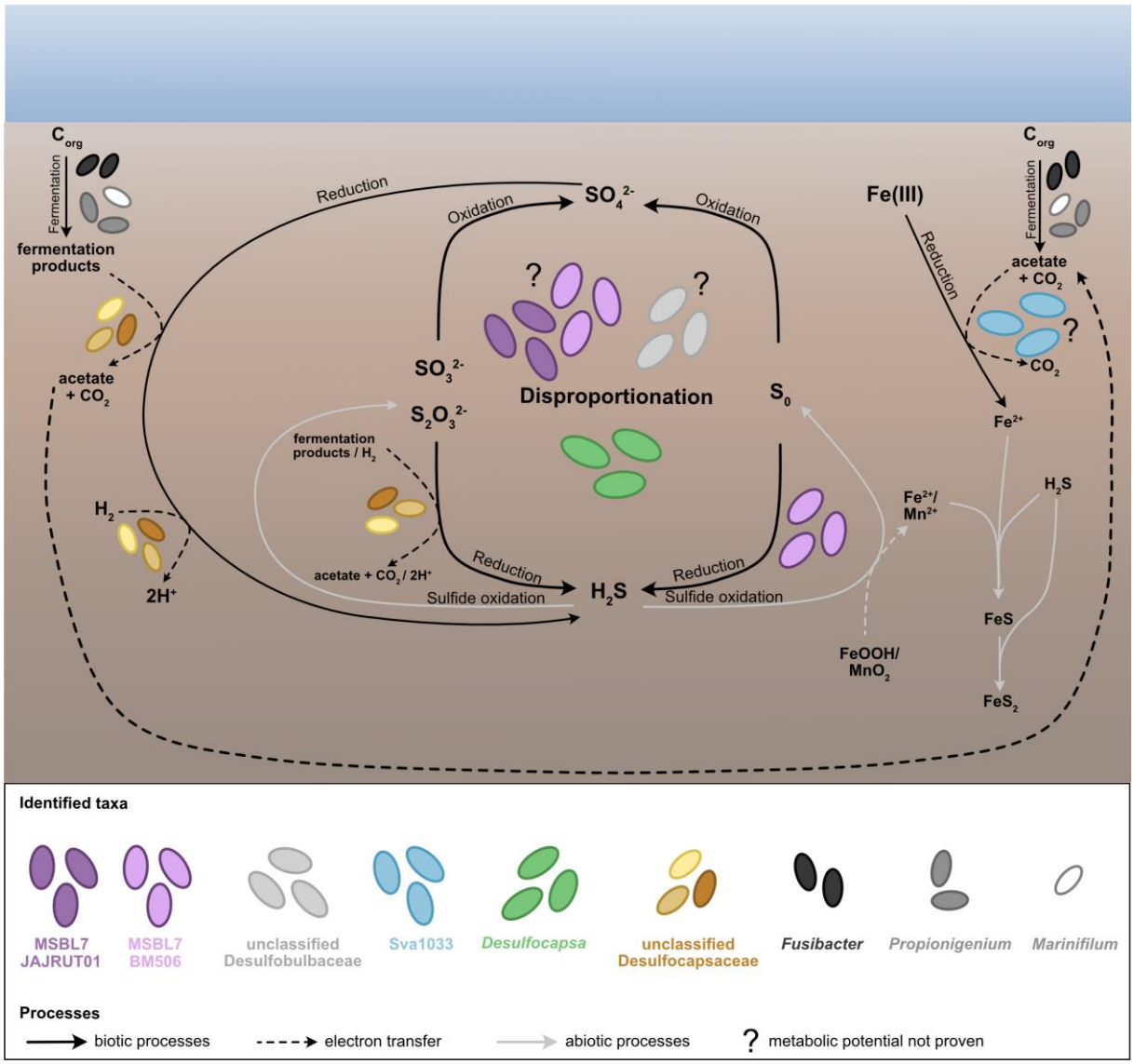

Figure 2 Reconstruction of microbial activity in the Helgoland mud area ^{18}O-labeled taxa and their metabolic capabilities in Helgoland mud area sediments as based on pure culture and isotope probing studies (Schink and Pfennig 1982, Szewzyk and Pfennig 1987, Thamdrup et al. 1993, Janssen and Liesack 1995, Janssen et al. 1996, Finster et al. 1998, Finster et al. 2013, Fadhlaoui et al. 2015, Barnum et al. 2018, Fu et al. 2018, Hashimoto et al. 2021, Wunder et al. 2021, Zhong et al. 2022, Yin et al. 2024). *Desulfocapsa* sp. disproportionate sulfur compounds (ca. 16% of taxa in ultraheavy gradient fractions). Unclassified Desulfocapsaceae (12% in heavy gradient fractions) reduce sulfate, thiosulfate and sulfite autotrophically or heterotrophically. For both genera of the MSBL7 and unclassified Desulfobulbaceae, sulfur compound disproportionation is a possible, yet unconfirmed, metabolic potential. Sediment endogenous organic carbon compounds (C_{org}) can be fermented by different microorganisms such as *Propionigenium*, *Marinifilum* and *Fusibacter* (ca. 36% in medium gradient fractions. Fermentation products formed may be oxidized by dissimilatory iron reduction by Sva1033 (13% in ultraheavy gradient fractions).

With the results obtained during this study, we could detect active taxa in the *in situ* environment and elaborate their metabolic capabilities based on their taxonomic affiliation. Although direct carbon utilization cannot be established based on the $H_2{}^{18}O$ RNA-SIP, we gained insights into the activities within sediments from the Helgoland mud area by identifying (**i.)** dominant sulfur compound cycling bacteria, thereby linking the sulfur and carbon fixation: Desulfocapsaceae, MSBL7 and Desulfobulbaceae. (**ii.)** active heterotrophic taxa, involved in iron cycling: Sva1033. (**iii.)** less active primary fermenters, e.g., *Marinifilum*, *Fusibacter* and *Propionigenium,* supplying sulfur and iron cycling taxa with electron donors in the form of fermentation products. Using the method of $H_2{}^{18}O$ RNA-SIP on a small scale on anaerobic sediments provided a great starting point for investigating the importance of *in situ* processes in untreated sediments. In combination with additional geochemical measurements of the sediment matrix and metagenomic analyses of identified active taxa, these analyses offer a powerful tool to link taxa to microbial processes, such as unresolved sulfur disproportionation.

2.1.3 Data availability

All code used to perform analyses in this study is available on https://github.com/mmaeke/18O_RNA-SIP. Amplicon raw reads for this study will be deposited in the European Nucleotide Archive (ENA) at EMBL-EBI upon manuscript submission and are currently available upon request.

2.1.4 Funding statement

This study was supported by DFG under Germany's excellence Strategy, no. EXC-2077-390741603.

2.1.5 References

Aoyagi, T., Hanada, S., Itoh, H., Sato, Y., Ogata, A., Friedrich, M. W., Kikuchi, Y. and Hori, T. (2015). Ultra-high-sensitivity stable isotope probing of rRNA by high-throughput sequencing of isopycnic centrifugation gradients. Environ. Microbiol. Rep. **7**:282-287. doi: 10.1111/1758-2229.12243.

Aronson, H. S., Thomas, C., Bhattacharyya, M. K., Eckstein, S. R., Jensen, S. R., Barco, R. A., Macalady, J. L. and Amend, J. P. (2023). *Thiovibrio frasassiensis* gen. nov., sp. nov., an autotrophic, elemental sulphur disproportionating bacterium isolated from sulphidic karst sediment, and proposal of *Thiovibrionaceae* fam. nov. Int. J. Syst. Evol. Micr. **73**. doi: 10.1099/ijsem.0.006003.

Bak, F. and Cypionka, H. (1987). A novel type of energy metabolism involving fermentation of inorganic sulphur compounds. Nature **326**:891-892. doi: 10.1038/326891a0.

Barnum, T. P., Figueroa, I. A., Carlström, C. I., Lucas, L. N., Engelbrektson, A. L. and Coates, J. D. (2018). Genome-resolved metagenomics identifies genetic mobility, metabolic interactions, and unexpected diversity in perchlorate-reducing communities. ISME J. **12**:1568-1581. doi: 10.1038/s41396-018-0081-5.

Borin, S., Brusetti, L., Mapelli, F., D'Auria, G., Brusa, T., Marzorati, M., Rizzi, A., Yakimov, M., Marty, D., De Lange, G. J., et al. (2009). Sulfur cycling and methanogenesis primarily drive microbial colonization of the highly sulfidic Urania deep hypersaline basin. Proc. Natl. Acad. Sci. U. S. A. **106**:9151-9156. doi: 10.1073/pnas.0811984106.

Brioukhanov, A. L., Kadnikov, V. V., Beletsky, A. V. and Savvichev, A. S. (2023). Aerotolerant thiosulfate-reducing bacterium *Fusibacter* sp. strain WBS isolated from littoral bottom sediments of the White Sea - Biochemical and genome analysis. Microorganisms **11**:1642. doi: 10.3390/microorganisms11071642.

Fadhlaoui, K., Ben Hania, W., Postec, A., Fauque, G., Hamdi, M., Ollivier, B. and Fardeau, M.-L. (2015). *Fusibacter fontis* sp. nov., a sulfur-reducing, anaerobic bacterium isolated from a mesothermic Tunisian spring. Int. J. Syst. Evol. Micr. **65**:3501-3506. doi: 10.1099/ijsem.0.000445.

Finster, K. (2008). Microbiological disproportionation of inorganic sulfur compounds. J. Sulfur Chem. **29**:281-292. doi: 10.1080/17415990802105770.

Finster, K., Liesack, W. and Thamdrup, B. (1998). Elemental sulfur and thiosulfate disproportionation by *Desulfocapsa sulfoexigens* sp. nov., a new anaerobic bacterium isolated from marine surface sediment. Appl. Environ. Microbiol. **64**:119-125. doi: 10.1128/AEM.64.1.119-125.1998.

Finster, K. W., Kjeldsen, K. U., Kube, M., Reinhardt, R., Mussmann, M., Amann, R. and Schreiber, L. (2013). Complete genome sequence of *Desulfocapsa sulfexigens*, a marine deltaproteobacterium specialized in disproportionating inorganic sulfur compounds. Stand. Genomic Sci. **8**:58-68. doi: 10.4056/sigs.3777412.

Fu, T., Jia, C., Fu, L., Zhou, S., Yao, P., Du, R., Sun, H., Yang, Z., Shi, X. and Zhang, X.-H. (2018). *Marinifilum breve* sp. nov., a marine bacterium isolated from the Yongle Blue Hole in the South China Sea and emended description of the genus *Marinifilum*. Int. J. Syst. Evol. Micr. **68**:3540-3545. doi: 10.1099/ijsem.0.003027.

Galushko, A. and Kuever, J. (2020). Desulfobulbaceae. *In:* M. E. Trujillo, S. Dedysh, P. DeVos, B. Hedlund, P. Kämpfer, F. A. Rainey, and W. B. Whitman, editors: Bergey's Manual of Systematics of Archaea and Bacteria. John Wiley & Sons, Inc., in association with Bergey's Manual Trust. p. 1-4. doi: 10.1002/9781118960608.fbm00194.pub2.

Hashimoto, Y., Tame, A., Sawayama, S., Miyazaki, J., Takai, K. and Nakagawa, S. (2021). *Desulfomarina profundi* gen. nov., sp. nov., a novel mesophilic, hydrogen-oxidizing, sulphate-reducing chemolithoautotroph isolated from a deep-sea hydrothermal vent chimney. Int. J. Syst. Evol. Micr. **71**. doi: 10.1099/ijsem.0.005083.

Janssen, P. H. and Liesack, W. (1995). Succinate decarboxylation by *Propionigenium maris* sp. nov., a new anaerobic bacterium from an estuarine sediment. Arch. Microbiol. **164**:29-35. doi: 10.1007/bf02568731.

Janssen, P. H., Schuhmann, A., Bak, F. and Liesack, W. (1996). Disproportionation of inorganic sulfur compounds by the sulfate-reducing bacterium *Desulfocapsa thiozymogenes* gen. nov., sp. nov. Arch. Microbiol. **166**:184-192. doi: 10.1007/s002030050374.

Jørgensen, B. B. (1982). Mineralization of organic matter in the sea bed - the role of sulphate reduction. Nature **296**:643-645. doi: 10.1038/296643a0.

Jørgensen, B. B. (2021). Sulfur biogeochemical cycle of marine sediments. Geochem. Perspect. **10**:145-307. doi: 10.7185/geochempersp.10.2.

Jørgensen, B. B. and Bak, F. (1991). Pathways and microbiology of thiosulfate transformations and sulfate eeduction in a marine sediment (Kattegat, Denmark). Appl. Environ. Microbiol. **57**:847-856. doi: 10.1128/aem.57.3.847-856.1991.

Ravenschlag, K., Sahm, K., Pernthaler, J. and Amann, R. (1999). High bacterial diversity in permanently cold marine sediments. Appl. Environ. Microbiol. **65**:3982-3989. doi: 10.1128/aem.65.9.3982-3989.1999.

Sass, A., Rütters, H., Cypionka, H. and Sass, H. (2002). *Desulfobulbus mediterraneus* sp. nov., a sulfate-reducing bacterium growing on mono- and disaccharides. Arch. Microbiol. **177**:468-474. doi: 10.1007/s00203-002-0415-5.

Schink, B. and Pfennig, N. (1982). *Propionigenium modestum* gen. nov. sp. nov. a new strictly anaerobic, nonsporing bacterium growing on succinate. Arch. Microbiol. **133**:209-216. doi: 10.1007/BF00415003.

Schwartz, E., Hayer, M., Hungate, B. A., Koch, B. J., McHugh, T. A., Mercurio, W., Morrissey, E. M. and Soldanova, K. (2016). Stable isotope probing with ^{18}O-water to investigate microbial growth and death in environmental samples. Curr. Opin. Biotechnol. **41**:14-18. doi: 10.1016/j.copbio.2016.03.003.

Slobodkin, A. I. and Slobodkina, G. B. (2019). Diversity of sulfur-disproportionating microorganisms. Microbiology **88**:509-522. doi: 10.1134/S0026261719050138.

Szewzyk, R. and Pfennig, N. (1987). Complete oxidation of catechol by the strictly anaerobic sulfate-reducing *Desulfobacterium catecholicum* sp. nov. Arch. Microbiol. **147**:163-168. doi: 10.1007/BF00415278.

Thamdrup, B., Finster, K., Hansen Jens, W. and Bak, F. (1993). Bacterial disproportionation of elemental sulfur coupled to chemical reduction of iron or manganese. Appl. Environ. Microbiol. **59**:101-108. doi: 10.1128/aem.59.1.101-108.1993.

Wasmund, K., Mußmann, M. and Loy, A. (2017). The life sulfuric: microbial ecology of sulfur cycling in marine sediments. Environ. Microbiol. Rep. **9**:323-344. doi: 10.1111/1758-2229.12538.

Wunder, L. C., Aromokeye, D. A., Yin, X., Richter-Heitmann, T., Willis-Poratti, G., Schnakenberg, A., Otersen, C., Dohrmann, I., Römer, M., Bohrmann, G., et al. (2021). Iron and sulfate reduction structure microbial communities in (sub-)Antarctic sediments. ISME J. **15**:3587-3604. doi: 10.1038/s41396-021-01014-9.

Yin, X., Zhou, G., Wang, H., Han, D., Maeke, M., Richter-Heitmann, T., Wunder, L. C., Aromokeye, D. A., Zhu, Q.-Z., Nimzyk, R., et al. (2024). Unexpected carbon utilization activity of sulfate-reducing microorganisms in temperate and permanently cold marine sediments. ISME J. **18**. doi: 10.1093/ismejo/wrad014.

Zhong, Y. W., Zhou, P., Cheng, H., Zhou, Y. D., Pan, J., Xu, L., Li, M., Tao, C. H., Wu, Y. H. and Xu, X. W. (2022). Metagenomic features characterized with microbial iron oxidoreduction and mineral interaction in southwest Indian Ridge. Microbiol. Spectr. **10**:e0061422. doi: 10.1128/spectrum.00614-22.

2.2 Supplementary

2.2.1 Supplementary Methods

2.2.1.1 Small-scale $H_2{}^{18}O$ RNA Stable Isotope Probing

Sediment was collected from the Helgoland mud area (54°06'11.3904''N, 7°57'45.864''E) by gravity coring in 2019 during the RV HEINCKE cruise HE531. Sediments from a depth of 0-25 cm were selected from gravity core HE531/3-1 for Stable Isotope Probing (SIP) incubations (Fig. S13). Incubations were set up using sediment and sterilized artificial sea water (ASW; composition 26.4 g NaCl, 11.2 g $MgCl_2 \cdot 6\ H_2O$, 1.5 g $CaCl_2 \cdot 2\ H_2O$, 0.7 g KCl per liter) at a ratio of 1:4 (w/v). On a small scale, artificial seawater was prepared using 660 mg NaCl, 280 mg $MgCl_2$, 37.5 mg $CaCl_2$ x H_2O and 17.5 mg KCl in 25 ml unlabeled or ^{18}O-labeled ASW ($H_2{}^{18}O$). To obtain sufficient amounts of RNA for RNA-SIP, samples were set up in quadruplets for each treatment, either containing $H_2{}^{18}O$- or unlabeled ASW. For single incubations, 1 g of wet sediment was weighed in screw cap vials and centrifuged at 20,817 *g* at 4°C for 5 minutes to remove most of the residue water in the sediment. Samples for $H_2{}^{18}O$ incubations were flushed once with 350 µl filter sterilized $H_2{}^{18}O$-ASW to dilute residues of unlabeled water. Sediment within screw cap vials was carefully mixed with the labeled water using a pipet tip and subsequently briefly vortexed. Samples were centrifuged again at 20,817 *g* at 4°C for 5 minutes to remove residue water. Samples for $H_2{}^{18}O$- and unlabeled ASW incubations were transferred into an anaerobic chamber. Samples were either amended with 700 µl of $H_2{}^{18}O$- or unlabeled ASW. Within the anaerobic chamber, samples were mixed as mentioned above. To establish an anaerobic environment within the sample tubes, tubes were incubated with open lids within the anaerobic chamber for several minutes. Samples were then moved into an anaerobic jar and removed from the anaerobic chamber. The anaerobic jar was flushed with N_2 to remove the remaining oxygen in the jar and incubated at 10°C for 7 days.

2.2.1.2 Density separation and gradient fractionation of RNA

After 7 days, samples were removed from the anaerobic jar, centrifuged for 5 minutes and supernatant discarded. The sediment within each screw cap vial was separated into 2 vials for each replicate. Nucleic acid extraction, removal of DNA, quantification, isopycnic centrifugation and gradient fractionation were performed according to Yin et al. (2019). DNA was removed according to the RQ1 DNase kit (Promega, Madison, Wisconsin, USA) at 37°C for 45 min. Following DNase digestion, RNA was purified by using phenol-chloroform-isoamyl alcohol and chloroform-isoamyl alcohol, precipitation was performed using polyethylene glycol 6000 (~30%). RNA extracts of all replicates were pooled to retrieve a

sufficient RNA concentration for the following SIP. RNA was quantified with Quanti-iT RiboGreen. For each incubation setup (unlabeled water day 7 and ^{18}O-water day 7), 4µl of the pooled RNA was used directly for cDNA synthesis. The remaining 205-250 ng RNA was used for density separation by ultracentrifugation (Fig. S1). From each treatment, 14 fractions were separated, with fraction 1 having the highest density and fraction 14 having the lowest. From each of the treatments, fractions 3 (ultra-heavy, 1.830-1.835 g/ml), 5 (heavy, 1.815 – 1.822 g/ml), 7 (midpoint, 1.801-1.808 g/ml), 9 (light, 1.788 - 1.795 g/ml), and 11 (ultra-light, 1.774 – 1.781 g/ml) were used for cDNA synthesis. Fractions used for 16S rRNA gene amplicon sequencing are listed in Table S7.

2.2.1.3 16S rRNA gene amplicon sequencing of SIP enrichments

Illumina amplicon sequencing libraries were prepared for the V4 region of bacterial and archaeal 16S rRNA. Primers targeting the V4 region of the bacterial 16S rRNA were Bac515F (5′-GTGYCAGCMGCCGCGGTAA-3′) (Parada et al. 2016) and Bac805R (5′-GACTACHVGGGTATCTAATCC-3′) (Herlemann et al. 2011), primers used for the V4 region of archaeal 16S rRNA were Arc519F (5'-CAGCMGCCGCGGTAA-3') (Ovreås et al. 1997) and Arc806R (5'-GGACTACVSGGGTATCTAAT-3') (Takai and Horikoshi 2000). The library preparation was performed according to Wunder et al. (2024). Thermal cycling conditions used for the sequencing PCR were as follows: initial denaturation at 95°C for 3 min, denaturation at 95°C for 20 s, annealing at 60°C for 15 s, elongation at 72°C for 15 s, final elongation at 72°C for 1 min. For bacterial samples 30 PCR cycles of denaturation, annealing and elongation were performed: archaeal samples underwent 35 PCR cycles. Sequencing of multiplexed libraries was performed at Novogene (Cambridge, UK) on the NovaSeq 6000 platform (2 x 250 bp, Illumina) in mixed orientation by ligation, therefore resulting in forward and reverse amplicon orientation in both forward (R1) and reverse reads (R2) for both, archaea and bacteria separately. Reads were demultiplexed and primer clipped using cutadapt v2.1 (Martin 2011) and further processed using the package dada2 v1.24.0 (Callahan et al. 2016) in R v4.2.1 (R Core Team 2020). Forward and reverse reads were trimmed at 140 bp and 150 bp for both, bacterial and archaeal reads with a maximum error rate of 2. Subsequently, error rates were learned and samples were dereplicated and denoised independently for each library orientation by pooling the data from all samples, using a modified loess function adapted for libraries with binned quality scores (Salazar 2020). Error-corrected R1 and R2 reads were merged into amplicon sequence variants (ASVs) and sequence tables for forward and reverse orientations were combined by reorientation of the reverse-forward ASVs. Chimeras, ASVs of

unexpected lengths (< 250 bp and > 253 bp for bacteria, < 252 bp and > 255 bp for archaea) and singletons were removed. A bootstrap cutoff of 70 was used to perform taxonomic classification with the assignTaxonomy function of dada2 with the SILVA nr99 v138.1 train set reference database (Quast et al. 2013, Yilmaz et al. 2013). For further processing of ASVs, chloroplasts, mitochondria and archaea were removed from the bacterial data set, all non-archaeal taxa were removed from the archaeal data set. Rarefaction curves were computed for bacterial and archaeal data sets to validate sufficient sequencing community composition using the package iNEXT v3.0.0 (Fig. S2) (Hsieh et al. 2016).

2.2.1.4 16S rRNA gene phylogenetic tree

For the phylogenetic placement of the retrieved ASVs, a 16S rRNA gene phylogenetic tree was constructed. All Desulfobulbia and Desulfuromonadia affiliated species present on GTDB were filtered to a checkM completeness of > 80% and contamination of < 5%. The quality-filtered species were downloaded from RefSeq yin

or GenBank (Sayers et al. 2019). 16S rRNA genes were extracted from all genomes using barrnap v0.9 (Seemann 2018). Additional 16S rRNA genes of type strains within the Desulfobulbia and Desulfuromonadia were acquired through the SILVA database (Quast et al. 2013) and NCBI (Sayers et al. 2022). All sequences, including the ASV sequences, were aligned using SINA v1.7.2 (Pruesse et al. 2012) against the SILVA nr99 v138.1 ARB database (Quast et al. 2013). Sequences were trimmed to fixed start and end positions of the full-length 16S rRNA gene (E.coli position 1,043 – 6,884) using seqtk v1.3-r106 (https://github.com/lh3/seqtk) and all gaps caused by one sequence removed with clipkit v2.2.4 (Steenwyk et al. 2020). For tree construction, sequences with at least 1,000 bp were chosen. Redundancy in sequences was removed by seqkit v2.3.1 (Shen et al. 2016). For all non-redundant sequences above 1,000 bp in length, a GTR+I+G4 model for the 16S rRNA gene phylogenetic tree was determined using modeltest-ng v0.1.7 (Darriba et al. 2019), which was further used for calculation of the tree with raxml-ng v1.1.0 (Kozlov et al. 2019). In total, 70 starting trees were inferred; bootstrap convergence at a cutoff of 0.02 was reached after 1,300 trees. Shorter 16S rRNA gene sequences found in Desulfobulbia and Desulfuromonadia genomes, along with ASVs from the SIP enrichment were added after tree calculation. Short sequences were placed into the existing tree using epa-ng v0.3.8 (Barbera et al. 2018) and gappa v0.7.1 (Czech et al. 2020). The tree was manually rooted and grouped in iTOL v6 (Letunic and Bork 2024). The 16S rRNA gene sequences of the obtained Desulfobulbia and Desulfuromonadia ASVs (section 3) were searched in the blastn suite of the online NCBI

BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome, accessed 12.08.2024) against the rRNA/ITS database with the megablast option to find highly similar sequences. The top hit for each ASV is listed in Tables S1-S2.

Further, the pairwise identity of all ASVs against all 16S rRNA genes within the phylogenetic tree was computed using the blastn function of BLAST+2.16.0 with an e-value threshold of 1×10^{-10} (Altschul et al. 1990). Results for the top 5 hits are summed in Tables S3-S6.

2.2.2 Supplementary Results

2.2.2.1 Successful labeling of microorganisms in ^{18}O-ASW treatments

For RNA-SIP, the highest amounts of RNA were found at a density of 1.78 g/mL for the unlabeled ASW (Fig. S1a) and the ^{18}O-ASW (Fig. S1b) density curves. Within the profile obtained from the^{18}O-ASW RNA-SIP run, a second peak between a density of 1.80 and 1.82 g/mL was observed (Figure S1b), indicating successful labeling of RNA by ^{18}O-ASW after 7 days.

2.2.2.2 Activity of archaea was not detected after seven days

Total archaeal RNA of both treatments (unlabeled and ^{18}O) showed high relative abundances of Bathyarchaeia (83.8 and 61.3%). Lokiarchaeia (6.2 and 13.7%), ANME-1 (1.6 and 5.2%), Methanosarcinia (3.4 and 8.8%) and Thermoplasmata (3.4 and 8%) were of lower relative abundance (Fig. S14). Heavy and ultraheavy fractions of the ^{18}O-RNA-SIP did not yield sufficient community coverage and were therefore excluded from the data set (Fig. S2). Assuming highly active taxa would be found in these heavy fractions, as they incorporated the heavier ^{18}O-isotope, archaea are regarded as non-active or hardly active.

2.2.2.3 High relative abundances of fermentative organisms in the midpoint fraction

Three groups were primarily detected in the midpoint – ultra-heavy fractions with the highest relative abundances at the midpoint. Fusobacteriia had a relative abundance of 22.3% in the midpoint (Fig. S11). Within this group, two main ASVs were detected affiliated with the genus *Propionigenium* (22.2%). Clostridia showed a relative abundance of 12.6%, with highest abundant ASVs being affiliated with the genus *Fusibacter* (11.1%) (Fig. S10). The class Bacteroidia showed the lowest relative abundance among these three groups (9.8 %) (Fig. S12). The highest abundant ASVs of the Bacteroidia were detected for the genus *Marinifilum* in the midpoint of the ^{18}O-ASW treated sample (3.1%). Besides, low abundant ASVs of Bacteroidales and Bacteroidetes VC2.1 Bac 22 were detected in heavy and midpoint fractions.

2.2.3 Supplementary Tables

Table S1 Top hit of the megablast search of ASVs against the NCBI online search tool rRNA/ITS database for ASVs of the family Desulfocapsaceae.

Query ASV	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
sq13	*Desulfocapsa sulfexigens* DSM 10523 16S ribosomal RNA, partial sequence	*Desulfocapsa sulfexigens* DSM 10523	431	431	100%	1.00E-120	97.61%	1551	NR_102510.1
sq15	*Desulfomarina profundi* strain KT2 16S ribosomal RNA, partial sequence	*Desulfomarina profundi*	436	436	100%	3.00E-122	98.01%	1550	NR_179352.1
sq19	*Desulfocastanea catecholica* strain NZva20 16S ribosomal RNA, partial sequence	*Desulfocastanea catecholica*	425	425	100%	6.00E-119	97.21%	1501	NR_028895.1
sq21	*Desulfocastanea catecholica* strain NZva20 16S ribosomal RNA, partial sequence	*Desulfocastanea catecholica*	442	442	100%	6.00E-124	98.41%	1501	NR_028895.1
sq22	*Desulfocapsa sulfexigens* DSM 10523 16S ribosomal RNA, partial sequence	*Desulfocapsa sulfexigens* DSM 10523	431	431	100%	1.00E-120	97.61%	1551	NR_102510.1
sq26	*Desulfocapsa sulfexigens* DSM 10523 16S ribosomal RNA, partial sequence	*Desulfocapsa sulfexigens* DSM 10523	436	436	100%	3.00E-122	98.01%	1551	NR_102510.1
sq46	*Desulfocapsa sulfexigens* DSM 10523 16S ribosomal RNA, partial sequence	*Desulfocapsa sulfexigens* DSM 10523	442	442	100%	6.00E-124	98.41%	1551	NR_102510.1
sq50	*Desulfocapsa sulfexigens* DSM 10523 16S ribosomal RNA, partial sequence	*Desulfocapsa sulfexigens* DSM 10523	420	420	100%	3.00E-117	96.81%	1551	NR_102510.1
sq53	*Desulfocapsa sulfexigens* DSM 10523 16S ribosomal RNA, partial sequence	*Desulfocapsa sulfexigens* DSM 10523	414	414	100%	1.00E-115	96.41%	1551	NR_102510.1

Table S2 Top hit of the megablast search of ASVs against the NCBI online search tool rRNA/ITS database for ASVs of the families Desulfurivvibrionaceae (genus MSBL7), Sva1033 and Desulfobulbaceae.

Query ASV	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
MSBL7									
sq9	*Desulfogranum mediterraneum* strain 86FS1 16S ribosomal RNA, partial sequence	*Desulfogranum mediterraneum*	387	387	100%	3.00E-107	94.42%	1457	NR_025150.1
sq12	*Thiovibrio frasassiensis* strain RS19-109 16S ribosomal RNA, partial sequence	*Thiovibrio frasassiensis*	398	398	100%	1.00E-110	95.24%	1415	NR_189258.1
sq34	T*hiovibrio frasassiensis* strain RS19-109 16S ribosomal RNA, partial sequence	*Thiovibrio frasassiensis*	387	387	100%	3.00E-107	94.44%	1415	NR_189258.1
sq59	*Desulfogranum mediterraneum* strain 86FS1 16S ribosomal RNA, partial sequence	*Desulfogranum mediterraneum*	381	381	100%	1.00E-105	94.02%	1457	NR_025150.1
Sva1033									
sq4	*Desulfuromonas acetoxidans* strain DSM 684 16S ribosomal RNA, partial sequence	*Desulfuromonas acetoxidans*	409	409	100%	6.00E-114	96.02%	1558	NR_121678.1
sq40	*Desulfuromonas acetoxidans* strain DSM 684 16S ribosomal RNA, partial sequence	*Desulfuromonas acetoxidans*	403	403	100%	3.00E-112	95.62%	1558	NR_121678.1
Desulfobulbaceae									
sq25	*Desulfogranum mediterraneum* strain 86FS1 16S ribosomal RNA, partial sequence	*Desulfogranum mediterraneum*	381	381	100%	1.00E-105	94.02%	1457	NR_025150.1
sq31	*Thiovibrio frasassiensis* strain RS19-109 16S ribosomal RNA, partial sequence	*Thiovibrio frasassiensis*	375	375	100%	6.00E-104	93.65%	1415	NR_189258.1

Table S3 Top five hits of the blastn search of ASVs against all input sequences of the 16S rRNA gene phylogenetic tree for ASVs of the family Desulfocapsaceae.

	ASV query	Closest hits	Taxonomy	Identity [%]	Length	Mismatch	E-value	Bit-score	Query cover
Desulfocapsaceae	sq13	sq50_Desulfocapsaceae		99.203	251	2	1.59E-126	444	100
		sq22_Desulfocapsaceae		98.406	251	4	8.22E-124	435	100
		CP003985.2137188.2138627	*Desulfocapsa sulfexigens* DSM 10523	97.61	251	6	4.26E-121	426	100
		GCF_000341395.1	*Desulfocapsa sulfexigens*	97.61	251	6	4.26E-121	426	100
		sq26_Desulfocapsaceae		97.211	251	7	1.81E-119	422	100
	sq15	sq26_Desulfocapsaceae		98.406	251	4	8.22E-124	435	100
		NR_179352.1	*Desulfomarina profundi*	98.008	251	5	3.5E-122	431	100
		GCF_019703855.1	*Desulfomarina profundi*	98.008	251	5	3.5E-122	431	100
		GCA_016765495.1	*Desulfocapsa* sp016765495	98.008	251	5	3.5E-122	431	100
		sq53_Desulfocapsaceae		97.61	251	6	4.26E-121	426	100
	sq19	sq21_Desulfocapsaceae		97.61	251	6	4.26E-121	426	100
		NR_028895.1	*Desulfocastanea catecholica*	97.211	251	7	1.81E-119	422	100
		GCF_005116645.1	*Desulforhopalus* sp005116645	97.211	251	7	1.81E-119	422	100
		AF118453.1.1498	*Desulforhopalus singaporensis*	96.813	251	8	2.21E-118	417	100
		GCF_900104445.1	*Desulforhopalus singaporensis*	96.813	251	8	2.21E-118	417	100
	sq20	GCF_010646885.2	*Desulfopila* sp010646885	99.203	251	2	1.59E-126	444	100
		sq22_Desulfocapsaceae		97.211	251	7	1.81E-119	422	100
		GCF_900143695.1	*Desulfopila aestuarii*	97.211	251	7	1.81E-119	422	100
		AB110542.1.1471	*Desulfopila aestuarii*	97.189	249	7	2.21E-118	418	99
		NR_173613.1	*Desulfosediminicola flagellatus*	96.414	251	9	9.38E-117	413	100
	sq21	GCF_005116645.1	*Desulforhopalus* sp005116645	99.602	251	1	1.3E-127	449	100
		NR_028895.1	*Desulfocastanea catecholica*	98.406	251	4	8.22E-124	435	100
		GCF_900143695.1	*Desulfopila aestuarii*	98.406	251	4	8.22E-124	435	100
		AB110542.1.1471	*Desulfopila aestuarii*	98.394	249	4	1E-122	432	99
		NR_173613.1	*Desulfosediminicola flagellatus*	97.61	251	6	4.26E-121	426	100

	ASV query	Closest hits	Taxonomy	Identity [%]	Length	Mismatch	E-value	Bit-score	Query cover
Desulfocapsaceae	sq22	sq50		98.406	251	4	8.22E-124	435	100
		sq13		98.406	251	4	8.22E-124	435	100
		sq26		98.008	251	5	3.5E-122	431	100
		CP003985.2137188.2138627	*Desulfocapsa sulfexigens* DSM 10523	97.61	251	6	4.26E-121	426	100
		GCF_000341395.1	*Desulfocapsa sulfexigens*	97.61	251	6	4.26E-121	426	100
	sq26	sq15		98.406	251	4	8.22E-124	435	100
		CP003985.2137188.2138627	*Desulfocapsa sulfexigens* DSM 10523	98.008	251	5	3.5E-122	431	100
		sq46		98.008	251	5	3.5E-122	431	100
		sq22		98.008	251	5	3.5E-122	431	100
		GCF_000341395.1	*Desulfocapsa sulfexigens*	98.008	251	5	3.5E-122	431	100
	sq46	GCA_016765495.1	*Desulfocapsa* sp016765495	99.203	251	2	1.59E-126	444	100
		CP003985.2137188.2138627	*Desulfocapsa sulfexigens* DSM 10523	98.406	251	4	8.22E-124	435	100
		GCF_000341395.1	*Desulfocapsa sulfexigens*	98.406	251	4	8.22E-124	435	100
		sq26		98.008	251	5	3.5E-122	431	100
		sq15		97.211	251	7	1.81E-119	422	100
	sq50	sq13		99.203	251	2	1.59E-126	444	100
		sq22		98.406	251	4	8.22E-124	435	100
		CP003985.2137188.2138627	*Desulfocapsa sulfexigens* DSM 10523	96.813	251	8	2.21E-118	417	100
		GCF_000341395.1	*Desulfocapsa sulfexigens*	96.813	251	8	2.21E-118	417	100
		sq26		96.414	251	9	9.38E-117	413	100
	sq53	sq26		97.61	251	6	4.26E-121	426	100
		sq15		97.61	251	6	4.26E-121	426	100
		CP003985.2137188.2138627	*Desulfocapsa sulfexigens* DSM 10523	96.414	251	9	9.38E-117	413	100
		sq46		96.414	251	9	9.38E-117	413	100
		GCF_000341395.1	*Desulfocapsa sulfexigens*	96.414	251	9	9.38E-117	413	100

Table S4 Top five hits of the blastn search of ASVs against all input sequences of the 16S rRNA gene phylogenetic tree for ASVs of the genus MSBL7 within the family Desulfurivibrionaceae.

	ASV query	Closest hits	Taxonomy	Identity [%]	Length	Mismatch	E-value	Bit-score	Query cover
MSBL7	sq9	sq59		99.602	251	1	1.30E-127	449	100
		GCA_002868945.1	BM506 sp002868945	99.203	251	2	1.59E-126	444	100
		GCA_024641085.1	JAHEDT01 sp024641085	95.618	251	11	4.86E-114	404	100
		sq12		95.219	251	12	5.92E-113	399	100
		sq25		94.821	251	13	2.52E-111	395	100
	sq12	sq34		99.203	251	2	1.59E-126	444	100
		GCA_024277655.1	JAJRUT01 sp024277655	98.406	251	4	8.22E-124	435	100
		GCA_002868945.1	BM506 sp002868945	96.016	251	10	1.14E-115	408	100
		GCA_013791995.1	*Thiovibrio* sp013791995	95.635	252	10	1.70E-113	402	100
		sq9		95.219	251	12	5.92E-113	399	100
	sq34	sq12		99.203	251	2	1.59E-126	444	100
		GCA_024277655.1	JAJRUT01 sp024277655	97.61	251	6	4.26E-121	426	100
		GCA_002868945.1	BM506 sp002868945	95.219	251	12	5.92E-113	399	100
		GCA_013791995.1	*Thiovibrio* sp013791995	94.841	252	12	8.79E-111	393	100
		sq9		94.422	251	14	3.07E-110	390	100
	sq59	sq9		99.602	251	1	1.3E-127	449	100
		GCA_002868945.1	BM506 sp002868945	98.805	251	3	6.75E-125	440	100
		GCA_024641085.1	JAHEDT01 sp024641085	95.219	251	12	5.92E-113	399	100
		sq12		94.821	251	13	2.52E-111	395	100
		sq25		94.422	251	14	3.07E-110	390	100

Table S5 Top five hits of the blastn search of ASVs against all input sequences of the 16S rRNA gene phylogenetic tree for ASVs of the family Desulfobulbaceae.

	ASV query	Closest hits	Taxonomy	Identity [%]	Length	Mismatch	E-value	Bit-score	Query Cover
Desulfobulbaceae	sq25	GCA_024641085.1	JAHEDT01 sp024641085	98.406	251	4	8.22E-124	435	100
		sq9		94.821	251	13	2.52E-111	395	100
		GCA_002868945.1	BM506 sp002868945	94.821	251	13	2.52E-111	395	100
		sq59		94.422	251	14	3.07E-110	390	100
		GCF_000429965.1	*Desulfogranum mediterraneum*	94.024	251	15	1.3E-108	386	100
	sq27	sq31		99.203	251	2	1.59E-126	444	100
		GCA_021647905.1	JAKITW01 sp021647905	96.813	251	8	2.21E-118	417	100
		GCA_015231515.1	JADGBM01 sp015231515	96.047	253	6	2.07E-112	398	100
		GCA_002868945.1	BM506 sp002868945	94.821	251	13	2.52E-111	395	100
		GCA_015488275.1	S012-135 sp015488275	94.422	251	14	3.07E-110	390	100
	sq31	sq27		99.203	251	2	1.59E-126	444	100
		GCA_021647905.1	JAKITW01 sp021647905	96.414	251	9	9.38E-117	413	100
		GCA_015231515.1	JADGBM01 sp015231515	94.422	251	14	3.07E-110	390	100
		GCA_002868945.1	BM506 sp002868945	94.422	251	14	3.07E-110	390	100
		GCA_903883485.1	*Thiovibrio* sp903869265	94.444	252	13	1.07E-109	389	100
	sq42	GCA_023544645.1	UBA10518 sp023544645	96.414	251	9	9.38E-117	413	100
		GCA_021647275.1	DRLX01 sp024280455	96.414	251	9	9.38E-117	413	100
		GCA_022766165.1	*Electrothrix gigas*	96.016	251	10	1.14E-115	408	100
		GCA_022765945.1	*Electrothrix gigas*	96.016	251	10	1.14E-115	408	100
		GCA_004028505.1	*Electrothrix arhusiensis*	96.016	251	10	1.14E-115	408	100

Table S6 Top five hits of the blastn search of ASVs against all input sequences of the 16S rRNA gene phylogenetic tree for ASVs of the family Sva1033 within the order Desulfuromonadales.

	ASV query	Closest hits	Taxonomy	Identity [%]	Length	Mismatch	E-value	Bit-score	Query cover
Sva1033	sq4	AJ240983.1	uncultured delta proteobacterium Sva1033	100	251	0	3.07E-129	453	100
		sq40		96.414	251	9	9.38E-117	413	100
		GCA_024641085.1	JAHEDT01 sp024641085	96.414	251	9	9.38E-117	413	100
		GCF_015234255.1	*Desulfuromonas acetoxidans*	96.016	251	10	1.14E-115	408	100
		GCF_013371615.1	*Desulfuromonas acetoxidans*	96.016	251	10	1.14E-115	408	100
	sq40	GCA_024641085.1	JAHEDT01 sp024641085	96.813	251	8	2.21E-118	417	100
		AJ240983.1	uncultured delta proteobacterium Sva1033	96.414	251	9	9.38E-117	413	100
		sq4		96.414	251	9	9.38E-117	413	100
		GCF_015234255.1	*Desulfuromonas acetoxidans*	95.618	251	11	4.86E-114	404	100
		GCF_013371615.1	*Desulfuromonas acetoxidans*	95.618	251	11	4.86E-114	404	100

Table S7 Sequenced fractions of the unlabeled and labeled (^{18}O-ASW) ASW treatments. All fractions used for amplicon sequencing are listed with their density after density separation and fractionation and the amount of RNA retrieved.

Treatment	Fraction	Density (g/ml)	RNA amount per Fraction (ng)
unlabeled ASW	3	1.829	0.151
	5	1.815	1.286
	7	1.801	5.887
	9	1.788	12.057
	11	1.774	19.222
^{18}O-ASW	3	1.835	15.482
	5	1.822	26.263
	7	1.808	26.011
	9	1.795	52.042
	11	1.781	76.199

2.2.4 Supplementary Figures

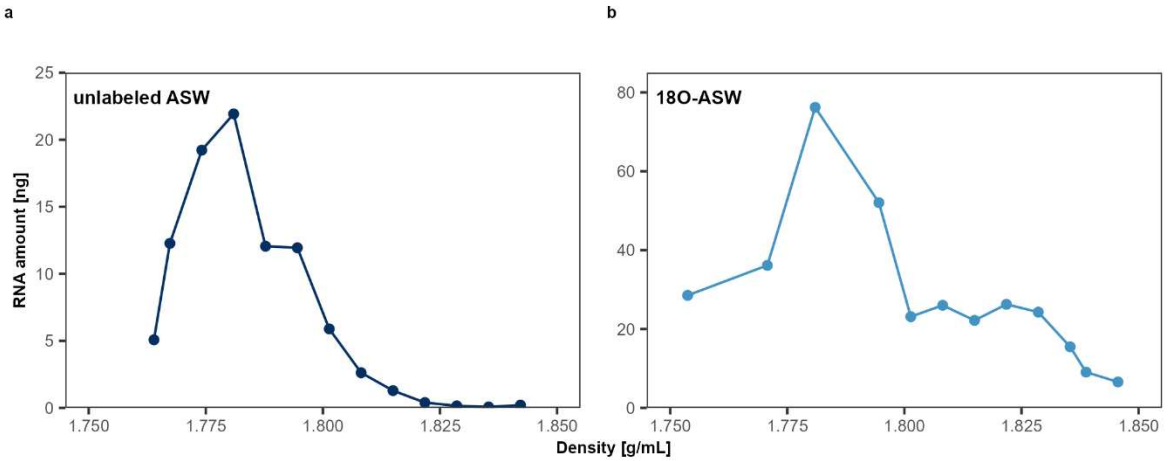

Figure S1 Density profiles of SIP incubations amended with **(a)** unlabeled artificial sea water (ASW) and **(b)** labeled ^{18}O-ASW (n=4, pooled for separation) after 7 days of incubation.

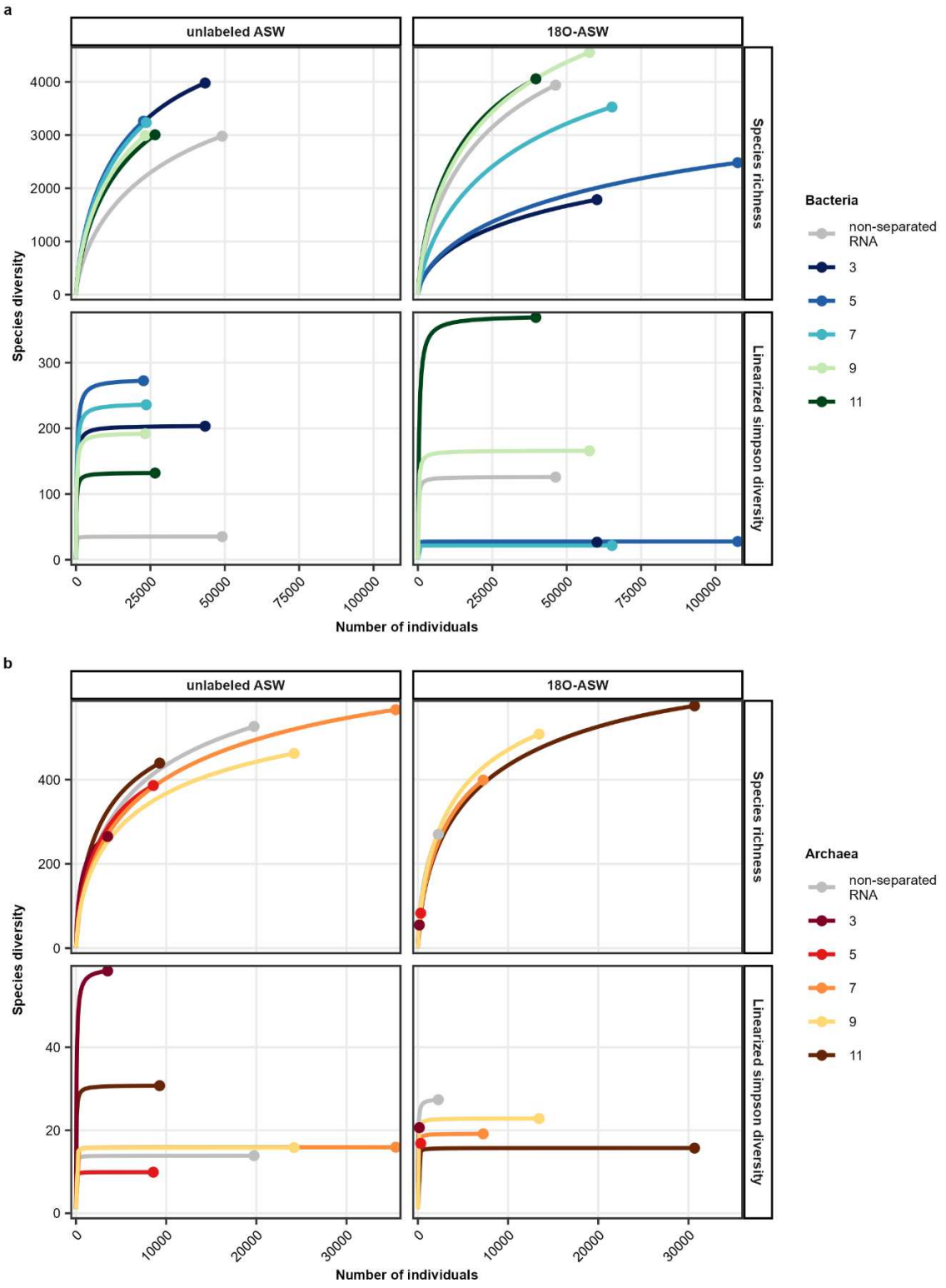

Figure S2 Rarefaction curves for the treatments of unlabeled ASW and ^{18}O-labeled ASW of **(a)** bacterial amplicon sequences for total RNA retrieved before density separation (non-separated RNA) and sequenced fractions (3 = ultra-heavy, 5 = heavy, 7 = midpoint, 9 = light, 11 = ultra-light) after 7 days of incubation. **(b)** archaeal amplicon sequences for total RNA retrieved before density separation (non-separated RNA) and sequenced fractions (3 = ultra-heavy, 5 = heavy, 7 = midpoint, 9 = light, 11 = ultra-light) after 7 days of incubation.

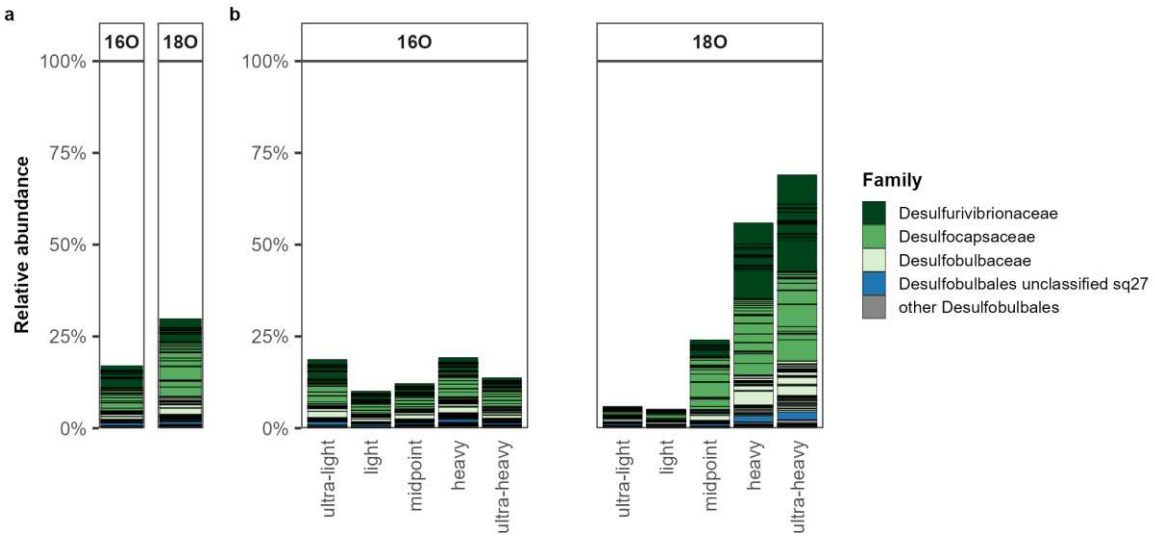

Figure S3 Relative abundance of bacterial RNA. Relative abundances of ASVs of the class Desulfobulbia in (**a)** total bacterial RNA of the unlabeled Artificial Sea Water (ASW, ^{16}O in plot) or labeled ^{18}O-ASW treated samples. **(b)** RNA-SIP fractions, retrieved after density separation and ultracentrifugation of the unlabeled ASW or labeled ^{18}O-ASW treated samples. Fractions are separated by density into ultra-light, light, midpoint, heavy and ultra-heavy.

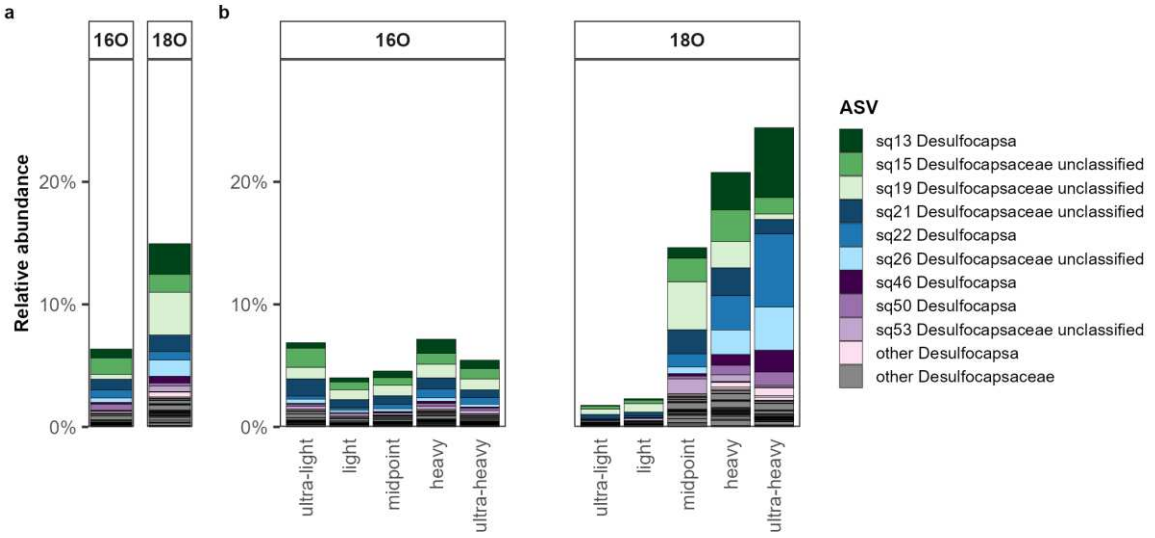

Figure S4 Relative abundance of bacterial RNA. Relative abundances of ASVs of the family Desulfocapsaceae in (**a)** total bacterial RNA of the unlabeled Artificial Sea Water (ASW, ^{16}O in plot) or labeled ^{18}O-ASW treated samples. **(b)** RNA-SIP fractions, retrieved after density separation and ultracentrifugation of the unlabeled ASW or labeled ^{18}O-ASW treated samples. Fractions are separated by density into ultra-light, light, midpoint, heavy and ultra-heavy.

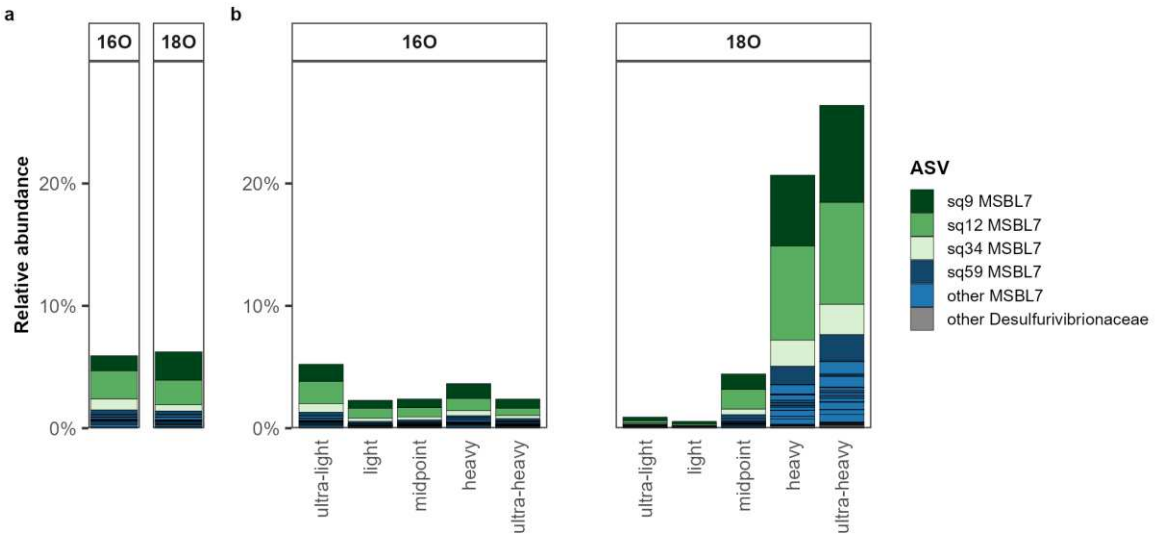

Figure S5 Relative abundance of bacterial RNA. Relative abundances of ASVs of the family Desulfurivibrionaceae in (**a)** total bacterial RNA of the unlabeled Artificial Sea Water (ASW, ^{16}O in plot) or labeled ^{18}O-ASW treated samples. **(b)** RNA-SIP fractions, retrieved after density separation and ultracentrifugation of the unlabeled ASW or labeled ^{18}O-ASW treated samples. Fractions are separated by density into ultra-light, light, midpoint, heavy and ultra-heavy.

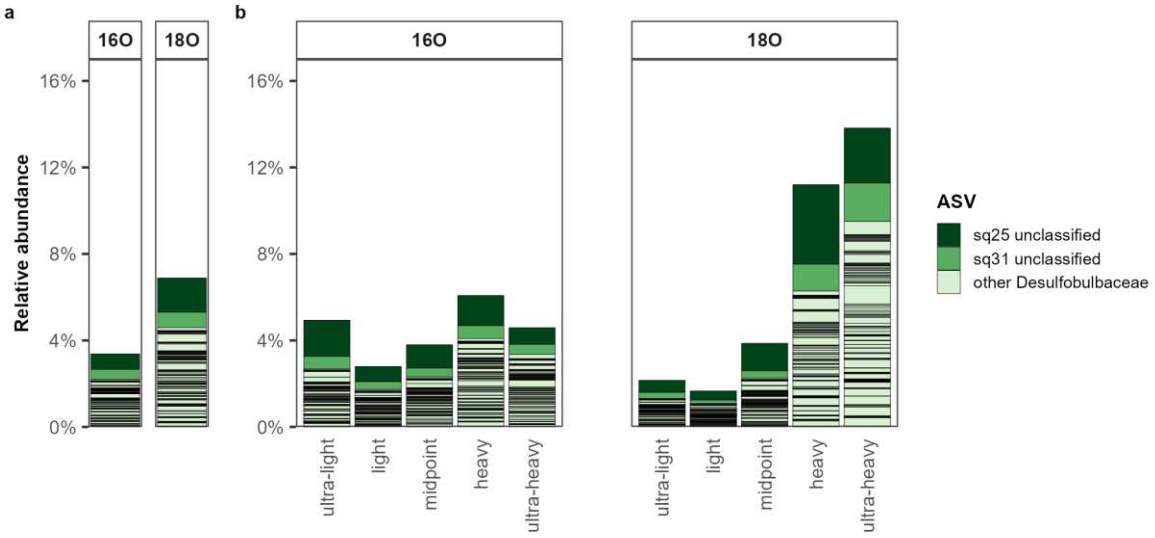

Figure S6 Relative abundance of bacterial RNA. Relative abundances of ASVs of the family Desulfobulbaceae in (**a)** total bacterial RNA of the unlabeled Artificial Sea Water (ASW, ^{16}O in plot) or labeled ^{18}O-ASW treated samples. **(b)** RNA-SIP fractions, retrieved after density separation and ultracentrifugation of the unlabeled ASW or labeled ^{18}O-ASW treated samples. Fractions are separated by density into ultra-light, light, midpoint, heavy and ultra-heavy.

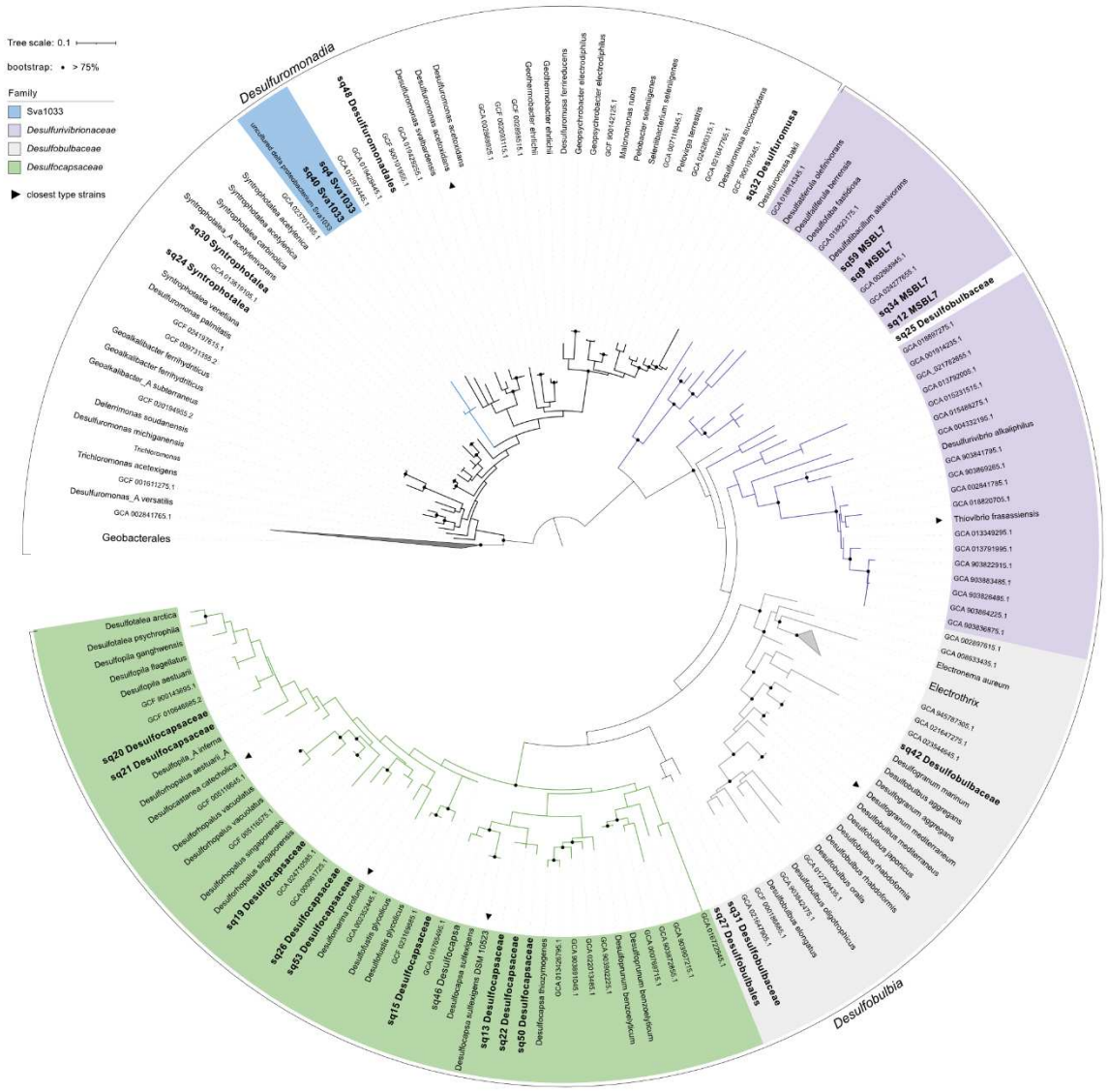

Figure S7 Maximum-likelihood tree (raxml, 1300 bootstraps) of 228 sequences with a length > 1,000 bp. Shorter 16S rRNA gene sequences of ASVs were added to the existing tree after initial calculation. Families of Sva1033, Desulfurivibrionaceae, Desulfobulbaceae and Desulfocapsaceae are indicated by color. Triangles mark the closest type strains (Table S1-S2).

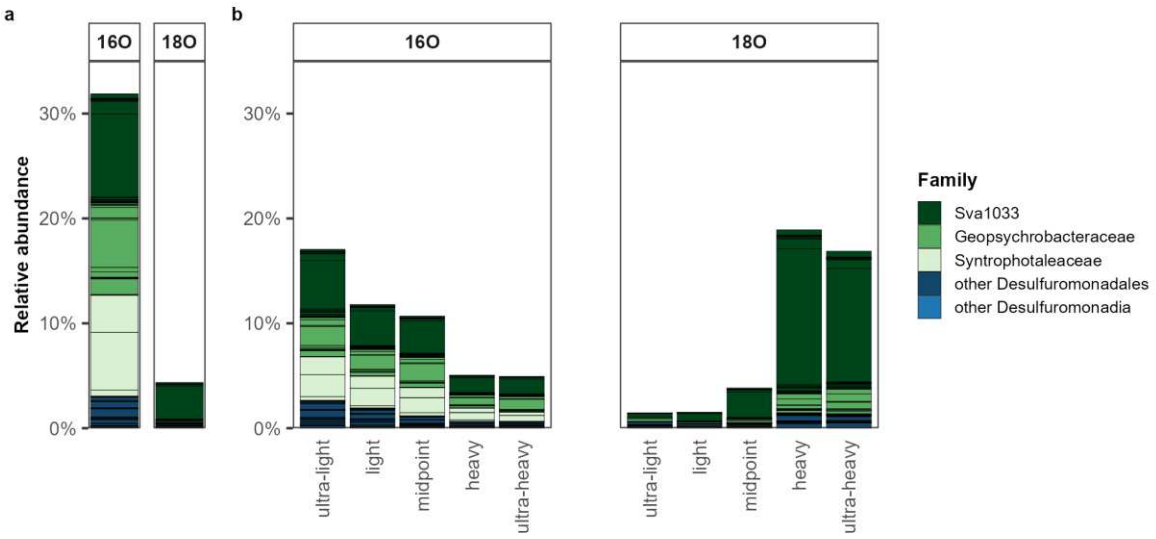

Figure S8 Relative abundance of bacterial RNA. Relative abundances of ASVs of the class Desulfuromonadia in (**a**) total bacterial RNA of the unlabeled Artificial Sea Water (ASW, ^{16}O in plot) or labeled ^{18}O-ASW treated samples. **(b)** RNA-SIP fractions, retrieved after density separation and ultracentrifugation of the unlabeled ASW or labeled ^{18}O-ASW treated samples. Fractions are separated by density into ultra-light, light, midpoint, heavy and ultra-heavy.

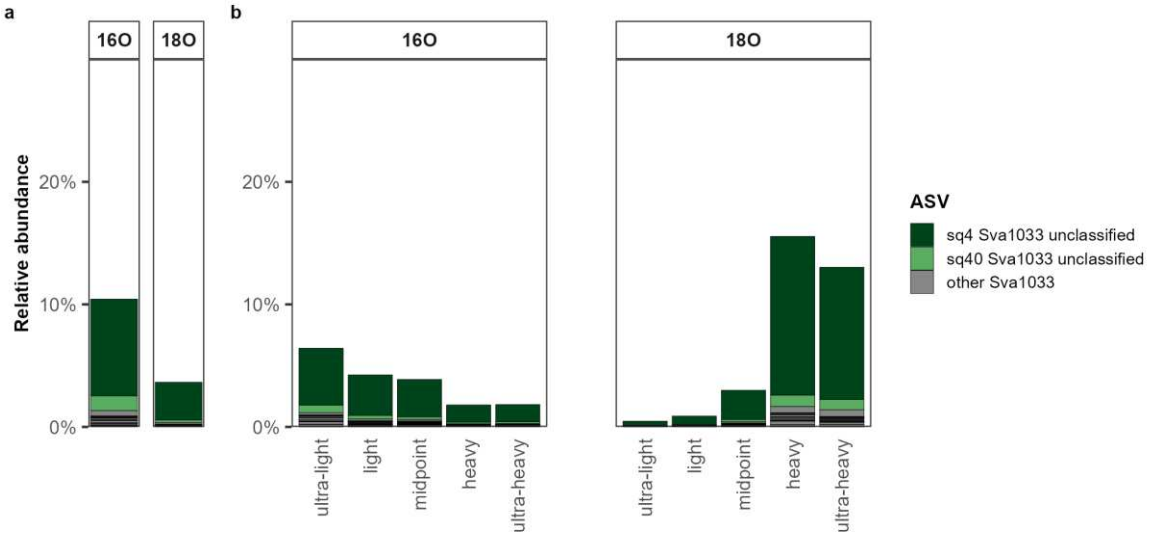

Figure S9 Relative abundance of bacterial RNA. Relative abundances of ASVs of the family Sva1033 in (**a**) total bacterial RNA of the unlabeled Artificial Sea Water (ASW, ^{16}O in plot) or labeled ^{18}O-ASW treated samples. **(b)** RNA-SIP fractions, retrieved after density separation and ultracentrifugation of the unlabeled ASW or labeled ^{18}O-ASW treated samples. Fractions are separated by density into ultra-light, light, midpoint, heavy and ultra-heavy.

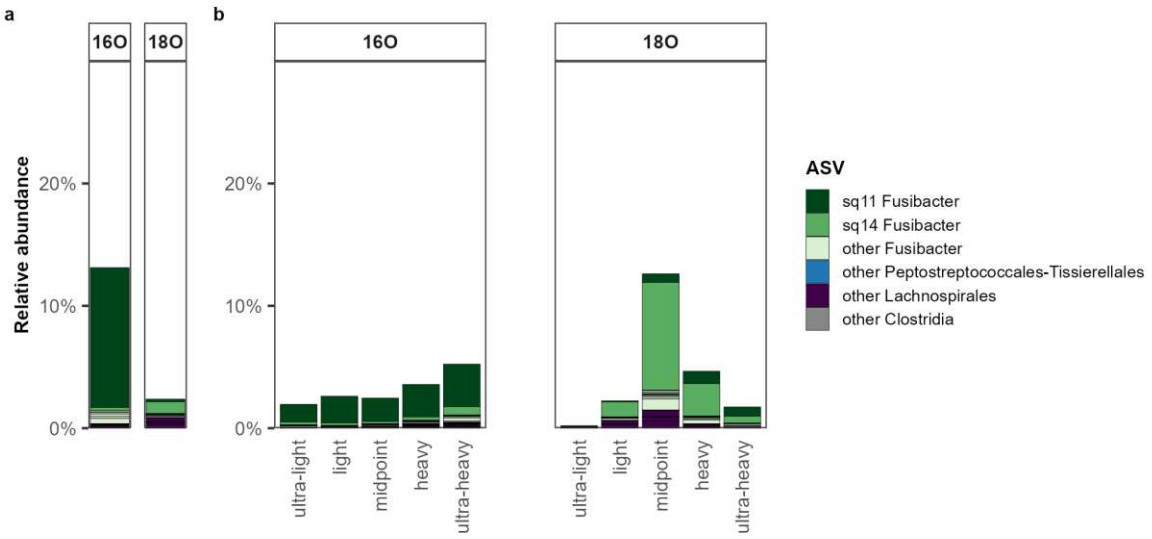

Figure S10 Relative abundance of bacterial RNA. Relative abundances of ASVs of the class Clostridia in (**a**) total bacterial RNA of the unlabeled Artificial Sea Water (ASW, ^{16}O in plot) or labeled ^{18}O-ASW treated samples. (**b**) RNA-SIP fractions, retrieved after density separation and ultracentrifugation of the unlabeled ASW or labeled ^{18}O-ASW treated samples. Fractions are separated by density into ultra-light, light, midpoint, heavy and ultra-heavy.

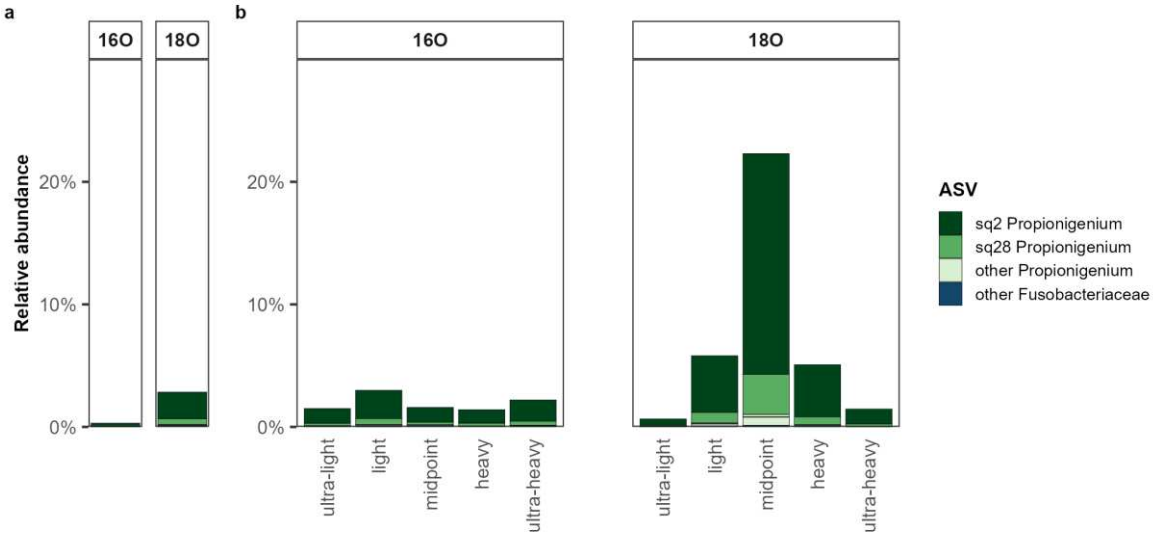

Figure S11 Relative abundance of bacterial RNA. Relative abundances of ASVs of the class Fusobacteriia in (**a**) total bacterial RNA of the unlabeled Artificial Sea Water (ASW, ^{16}O in plot) or labeled ^{18}O-ASW treated samples. (**b**) RNA-SIP fractions, retrieved after density separation and ultracentrifugation of the unlabeled ASW or labeled ^{18}O-ASW treated samples. Fractions are separated by density into ultra-light, light, midpoint, heavy and ultra-heavy.

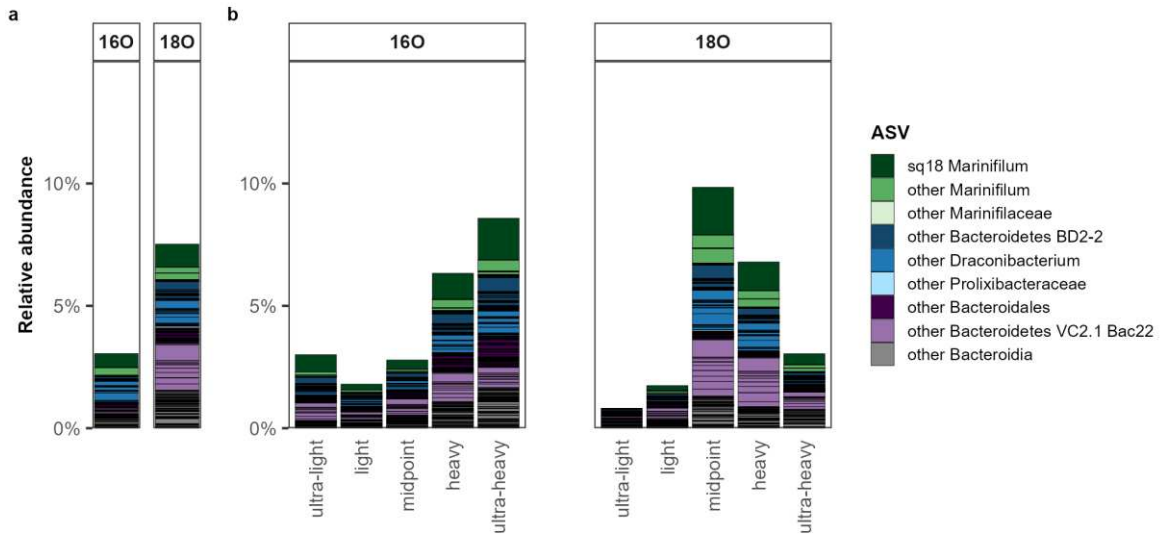

Figure S12 Relative abundance of bacterial RNA. Relative abundances of ASVs of the class Bacteroidia in (**a**) total bacterial RNA of the unlabeled Artificial Sea Water (ASW, ^{16}O in plot) or labeled ^{18}O-ASW treated samples. (**b**) RNA-SIP fractions, retrieved after density separation and ultracentrifugation of the unlabeled ASW or labeled ^{18}O-ASW treated samples. Fractions are separated by density into ultra-light, light, midpoint, heavy and ultra-heavy.

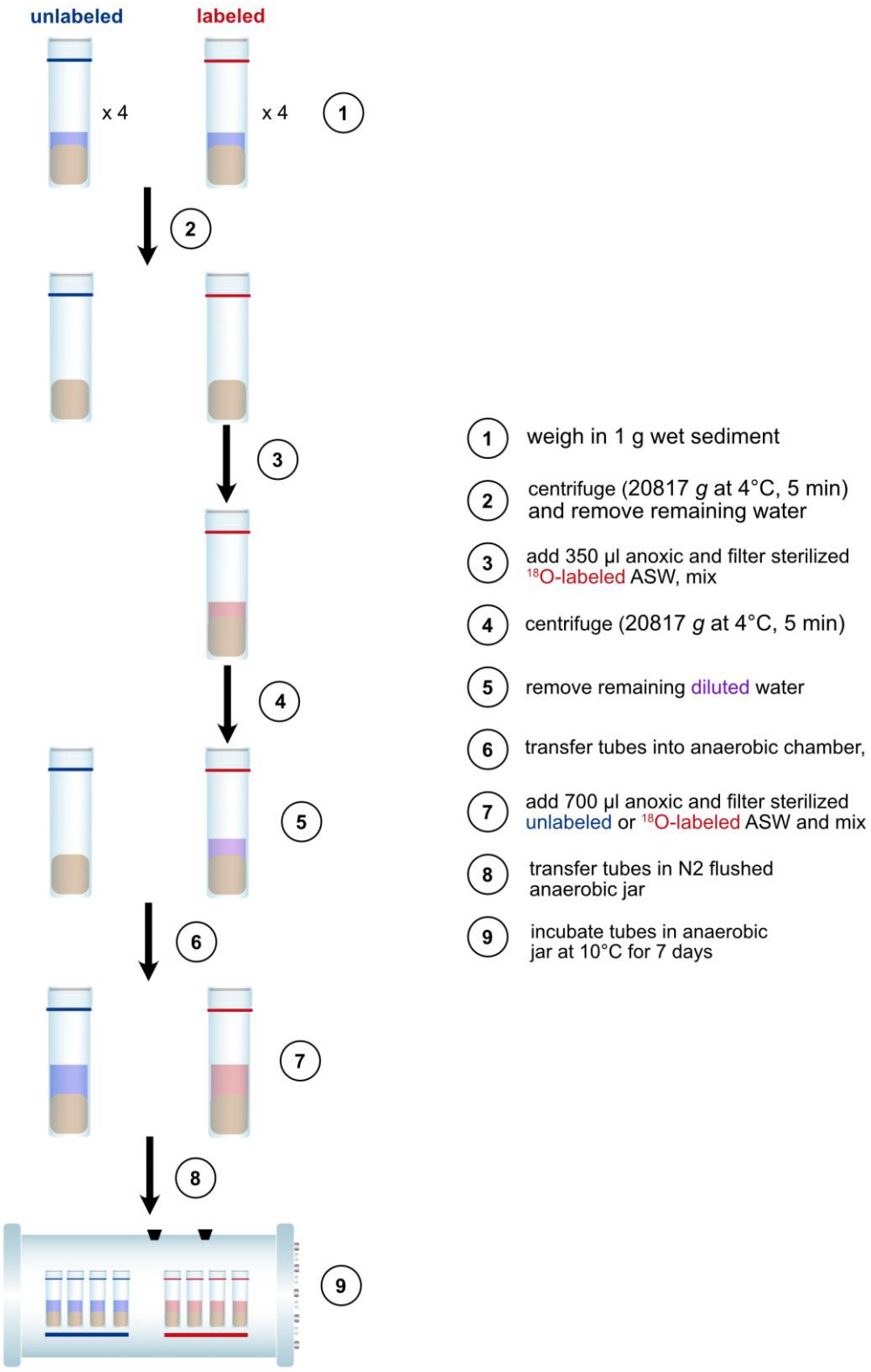

Figure S13 Schematic overview of the newly developed anoxic small-scale ^{18}O-RNA-SIP method.

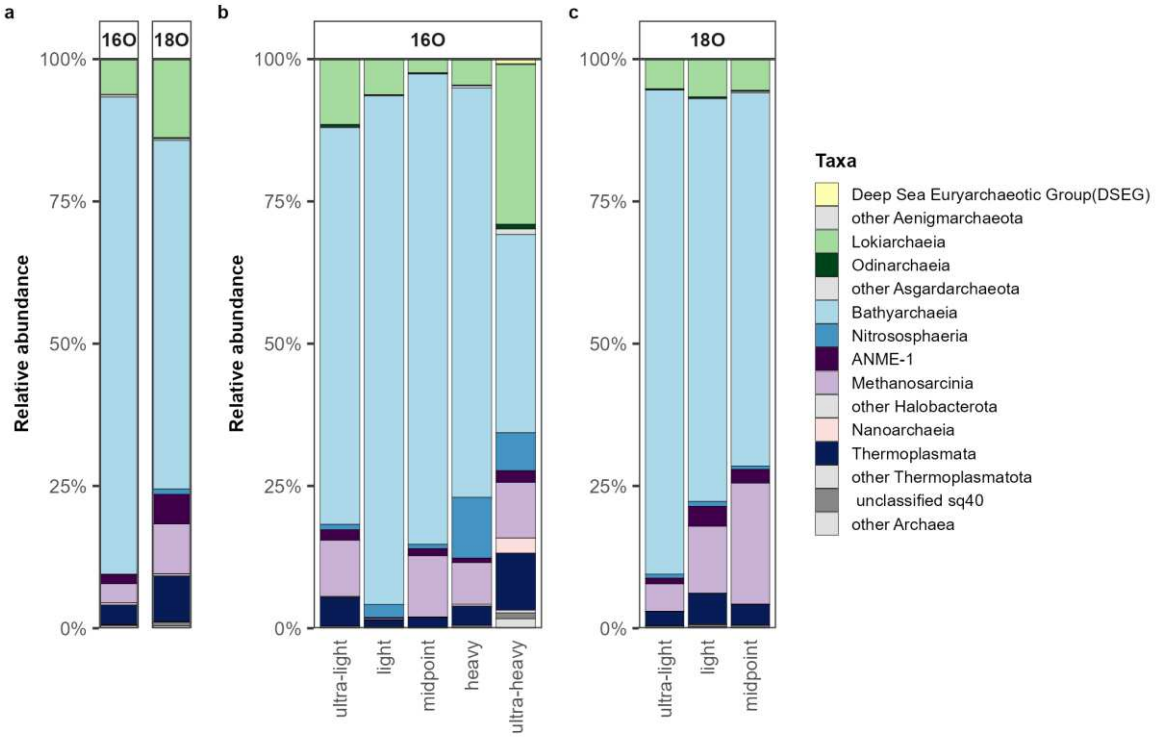

Figure S14 Relative abundance of archaeal classes in (**a)** total archaeal RNA of the unlabeled Artificial Sea Water (ASW, ^{16}O in plot) or labeled ^{18}O-ASW treated samples. (**b)** RNA-SIP fractions, retrieved after density separation and ultracentrifugation of the unlabeled ASW or labeled ^{18}O-ASW treated samples. Fractions are separated by density into ultra-light, light, midpoint, heavy and ultra-heavy.

2.2.5 Supplementary References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. **215**:403-410. doi: 10.1016/s0022-2836(05)80360-2.

Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T. and Stamatakis, A. (2018). EPA-ng: massively parallel evolutionary placement of genetic sequences. Syst. Biol. **68**:365-369. doi: 10.1093/sysbio/syy054.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods **13**:581-583. doi: 10.1038/nmeth.3869.

Czech, L., Barbera, P. and Stamatakis, A. (2020). Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. Bioinformatics **36**:3263-3265. doi: 10.1093/bioinformatics/btaa070.

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B. and Flouri, T. (2019). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol. Biol. Evol. **37**:291-294. doi: 10.1093/molbev/msz189.

Herlemann, D. P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J. and Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. **5**:1571-1579. doi: 10.1038/ismej.2011.41.

Hsieh, T. C., Ma, K. H. and Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol. Evol. **7**:1451-1456. doi: 10.1111/2041-210X.12613.

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics **35**:4453-4455. doi: 10.1093/bioinformatics/btz305.

Letunic, I. and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. **52**:gkae268. doi: 10.1093/nar/gkae268.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. **17**:3. doi: 10.14806/ej.17.1.200.

Ovreås, L., Forney, L., Daae, F. L. and Torsvik, V. (1997). Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. Appl. Environ. Microbiol. **63**:3367-3373. doi: 10.1128/aem.63.9.3367-3373.1997.

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

Pruesse, E., Peplies, J. and Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics **28**:1823-1829. doi: 10.1093/bioinformatics/bts252.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. **41**:D590-596. doi: 10.1093/nar/gks1219.

R Core Team. (2020). R: a Language and Environment for Statistical Computing. https://www.R-project.org/.

Salazar, G. (2020). https://github.com/benjjneb/dada2/issues/938#issuecomment-589657164. (accessed 15. January 2024).

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., et al. (2022). Database resources of the national center for biotechnology information. Nucleic Acids Res. **50**:D20-d26. doi: 10.1093/nar/gkab1112.

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D. and Karsch-Mizrachi, I. (2019). GenBank. Nucleic Acids Res. **48**:D84-D86. doi: 10.1093/nar/gkz956.

Seemann, T. (2018). barrnap 0.9 : rapid ribosomal RNA prediction. https://github.com/tseemann/barrnap.

Shen, W., Le, S., Li, Y. and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLOS One **11**:e0163962. doi: 10.1371/journal.pone.0163962.

Steenwyk, J. L., Buida, T. J., 3rd, Li, Y., Shen, X. X. and Rokas, A. (2020). ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. **18**:e3001007. doi: 10.1371/journal.pbio.3001007.

Takai, K. and Horikoshi, K. (2000). Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. Appl. Environ. Microbiol. **66**:5066-5072. doi: 10.1128/aem.66.11.5066-5072.2000.

Wunder, L. C., Breuer, I., Willis-Poratti, G., Aromokeye, D. A., Henkel, S., Richter-Heitmann, T., Yin, X. and Friedrich, M. W. (2024). Manganese reduction and associated microbial communities in Antarctic surface sediments. Front. Microbiol. **15**. doi: 10.3389/fmicb.2024.1398021.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. and Glöckner, F. O. (2013). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. **42**:D643-D648. doi: 10.1093/nar/gkt1209.

Yin, X., Kulkarni, A. C. and Friedrich, M. W. (2019). DNA and RNA stable isotope probing of methylotrophic methanogenic archaea. *In:* M. Dumont and M. Hernández García, editors: Stable Isotope Probing. Methods in Molecular Biology. Humana, New York, NY. p. 189-206. doi: 10.1007/978-1-4939-9721-3_15.

Chapter III

Extensive data mining uncovers novel diversity among members of the rare biosphere within the Thermoplasmatota

Mara D. Maeke, Xiuran Yin, Lea C. Wunder, Chiara Vanni, Tim Richter-Heitmann, Samuel Miravet-Verde, Hans-Joachim Ruscheweyh, Shinichi Sunagawa, Jenny Fabian, Judith Piontek, Michael W. Friedrich & Christiane Hassenrück

Manuscript in review at Microbiome
Preprint at Research Square https://doi.org/10.21203/rs.3.rs-5240808/v1

Running title:
EX4484-6 data-mining

Contribution to the manuscript:

Experimental concept and design	60%
Acquisition of experimental data	70%
Data analysis and interpretation	70%
Preparation of figures and tables	100%
Drafting of manuscript	100%

Extensive data mining uncovers novel diversity among members of the rare biosphere within the Thermoplasmatota

Mara D. Maeke[1], Xiuran Yin[1,2*], Lea C. Wunder[1], Chiara Vanni[3], Tim Richter-Heitmann[1], Samuel Miravet-Verde[4], Hans-Joachim Ruscheweyh[4], Shinichi Sunagawa[4], Jenny Fabian[5], Judith Piontek[5], Michael W. Friedrich[1,3*] & Christiane Hassenrück[5*]

[1]Microbial Ecophysiology Group, Faculty of Biology/Chemistry, University of Bremen, Bremen, Germany

[2]State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou, China

[3]MARUM – Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany

[4]Department of Biology, Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Zurich, Zurich, Switzerland

[5]Biological Oceanography, Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Rostock, Germany

*Correspondence:
Prof. Dr. Michael W. Friedrich
University of Bremen, Microbial Ecophysiology Group
James-Watt-Straße 1 28359 Bremen, Germany
Phone: +49 (0) 421 218-63060
Email: michael.friedrich@uni-bremen.de

Dr. Christiane Hassenrück
Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Biological Oceanography
Seestrasse 15 18119 Rostock-Warnemünde, Germany
Phone: +49 (0)381 5197 227
Email: christiane.hassenrueck@io-warnemuende.de

Dr. Xiuran Yin
Hainan University, State Key Laboratory of Marine Resource Utilization in South China Sea
Renmin Ave. No.58 570228 Haikou, China
Email: 996383@hainanu.edu.cn

3.1.1 Abstract

Background:

Rare species, especially of the marine sedimentary biosphere, have long been overlooked owing to the complexity of sediment microbial communities, their sporadic temporal and patchy spatial abundance and challenges in cultivating environmental microorganisms. In this study, we combined enrichments, targeted metagenomic sequencing and extensive data-mining to uncover uncultivated members of the archaeal rare biosphere in marine sediments.

Results:

In protein-amended enrichments, we detected the ecologically and metabolically uncharacterized class EX4484-6 within the phylum Thermoplasmatota. By extensively screening more than 8,000 metagenomic runs and 11,479 published genome assemblies, we expanded the phylogeny of class EX4484-6 by three novel orders. All six identified families of this class show low abundance in environmental samples characteristic of rare biosphere members. Members of the EX4484-6 class were predicted to be involved in organic matter degradation in anoxic, carbon rich habitats. All EX4484-6 families contain high numbers of taxon-specific orthologous genes, highlighting their environmental adaptations and habitat specificity. Besides, members of this group exhibit the highest proportion of unknown genes within the entire phylum Thermoplasmatota, suggesting a high degree of functional novelty in this class.

Conclusions:

In this study, we emphasize the necessity of targeted, data-integrative approaches to deepen our understanding of the rare biosphere and uncover the functions and metabolic potential hidden within these understudied taxa.

3.1.2 Introduction

In the environment, the vast majority of microbial species is represented by low abundant microorganisms, known as the ‘rare biosphere’ [1]. While many studies define rare taxa as those being less than 0.01-0.1% abundant in a sample at a specific time point [2, 3], rarity is not only confined to population sizes but can also be measured by geographic range and habitat specificity [4, 5]. Rare taxa are hypothesized to play an important role in ecosystems by carrying a gene pool, which can be accessed under changing environmental conditions, functioning as seed bank for other microbial taxa, or by supporting the community with key functions [6, 7]. These functions can be nutrient cycling [8-10], degradation of pollutants [11] or the promotion of community resilience [12, 13]. Due to high intra- or interspecific

competition [6], as well as the assumed limited environmental distribution most rare taxa occupy, these taxa mostly show only temporally and spatially constrained abundance [7], making the study of the rare biosphere challenging.

Most studies so far conducted on the rare marine biosphere have focused on diversity assessments of bacterial and archaeal plankton using high-throughput 16S rRNA gene surveys [2, 3, 14-20]. More recently, first metagenomic approaches have been applied to investigate the rare biosphere in marine bacterioplankton [21]. However, the rare biosphere in marine sediments remains largely unexplored, owing to the complexity of sediment communities [22, 23].

To reduce the complexity of sediment communities and thereby making the rare biosphere more accessible for the analysis of metabolic functions, enrichment experiments with substrate amendments were designed to selectively promote the growth of rare taxa [24-27]. While isolation would be preferable to further characterize the metabolic capabilities of rare taxa, most microorganisms remain uncultured despite recent advances in cultivation techniques [28, 29]. Therefore, enrichment techniques coupled to metagenome analyses simplify the description of the full metabolic potential of rare uncultivated microorganisms in the absence of isolates [2, 29-31]. Yet, even the combination of these methods only provides a snapshot of the enriched taxa at a specific time point in a specific setting and cannot offer further information regarding their global diversity or habitat selection.

Recent advances in high-throughput sequencing and computational techniques have enabled deep sequencing of microbial communities, including the rare biosphere [2, 32-34] and led to rapid data accumulation on public databases. Data from next generation sequencing runs deposited on the Sequence Read Archive (SRA) under the umbrella of the National Center for Biotechnology Information (NCBI) reached a data volume of 57.9 petabytes in 2024 [35], while the number of assembled genomes exceeds 2.3 million [36]. Through the screening and recovery of novel metagenome-assembled genomes (MAGs) and single-cell amplified genomes (SAGs) from public archives by data-driven projects, such as the Genome Taxonomy Database (GDTB) [37], the Genomes from Earth's Microbiomes (GEM) [38], the OceanDNA MAG catalog [39] and the Ocean Microbiomics Database [40], the catalog of microbial diversity is steadily increasing. Genomic analyses conducted on unidentified microorganisms disclosed novel functions and illustrated the extent of their untapped metabolic potential [38, 40-43]. These studies provided first evidence of the wealth of information available in existing data on public archives.

Despite the evidence outlined above for the importance of the rare biosphere, rare taxa remain regularly overlooked in environmental metagenomic studies, since the most abundant taxa remain the focus of the research on key players in microbial communities [6]. Still, the generated sequencing data may hold valuable information about the phylogenetic and metabolic diversity of rare taxa, their habitats and their role in the Earth's environments, that can be accessed by group-targeted data mining of the SRA.

Especially archaea have been regarded as members of the rare biosphere, complicating the study of their diversity and function in biogeochemical cycles [34, 44]. While it has been estimated that half of the archaeal diversity remains unidentified, multiple archaeal groups were shown to play an important role in organic matter degradation [34, 45, 46]. To investigate the capabilities of uncultivated archaea involved in the degradation of organic matter, specifically proteins, we established enrichments amended with pure egg-white protein (further referred to as protein). After detecting a so far ecologically and metabolically uncharacterized and rare class of the phylum Thermoplasmatota within these protein-amended enrichments, we used the data available on the databases of the International Nucleotide Sequence Database Collaboration (INSDC) to conduct an extensive group-targeted data mining to describe this new class. By screening 11,479 publicly available genome assemblies, data from 8,287 metagenomic sequencing runs as well as the genomes of the Ocean Microbiomics Database (OMD), we generated an integrative dataset enabling the study of this new class. We investigated the phylogenetic diversity, biogeography and metabolic capabilities of the novel class and present three so far unknown orders within this group. Notably, by analyzing the metabolic potential within the novel class, we observed percentages of unknown genes higher than those found in any other class within the Thermoplasmatota.

3.1.3 Results

3.1.3.1 High abundance of an ecologically and metabolically uncharacterized class of the phylum Thermoplasmatota in protein enrichments

To investigate the potential for novel microorganisms involved in protein degradation, we established a series of slurry enrichments using sediment from the Helgoland mud area. In order to reduce, and eventually eliminate the sediment component, incubations were transferred into anoxic Widdel medium after 372 days for further enrichment (methods section 2). Subsamples of the second-generation enrichment were taken on days 98 (480 days total) and 157 (529 days total) to perform 16S rRNA amplicon sequencing and qPCR. Within one replicate incubation, amended with protein, sulfate, and antibiotics, we observed a high relative

abundance of EX4484-6 (90.4%), a so far undescribed class within the phylum Thermoplasmatota, which had increased from day 98 to day 157 (Fig. 1a, Table S1). The taxon was represented by four ASVs of which one ASV showed a relative abundance of 90.2% (Fig. S1b, Table S1). In contrast, a second replicate showed a high relative abundance of '*Candidatus* Prometheoarchaeum syntrophicum' strain MK-D1 on day 157 (86.56%), while the relative abundance of the EX4484-6 class was lower (12%) (Fig. S1a and S1c, Table S1, Results SI). The control samples without protein amendment showed high relative abundances in Bathyarchaeia and Lokiarchaeia, reaching up to 52% and 31%, respectively. The class EX4484-6 was not enriched in any of the control samples.

Using a newly designed qPCR primer set, we followed the abundance of 16S rRNA gene copies of the class EX4484-6 within the same enrichments. In the unenriched environmental samples and in control samples, gene copies of the EX4484-6 class were below the detection limit (Fig. 1b), providing first indication that in the environment this group is a member of the rare biosphere. However, in the protein-amended samples, the 16S rRNA gene copies of the EX4484-6 class increased 100-fold from day 98 to day 157, reaching a number of 1.2 x 10^7 gene copies per ml slurry (Fig. 1b, day 157). After 372 days (744 days total), gene copies decreased again (8.7 x 10^5 gene copies per ml slurry; Fig. S2). Based on these initial findings the uncultivated EX4484-6 class might be involved in protein metabolism. Metagenomic sequencing was performed to further analyze the metabolic potential of this class.

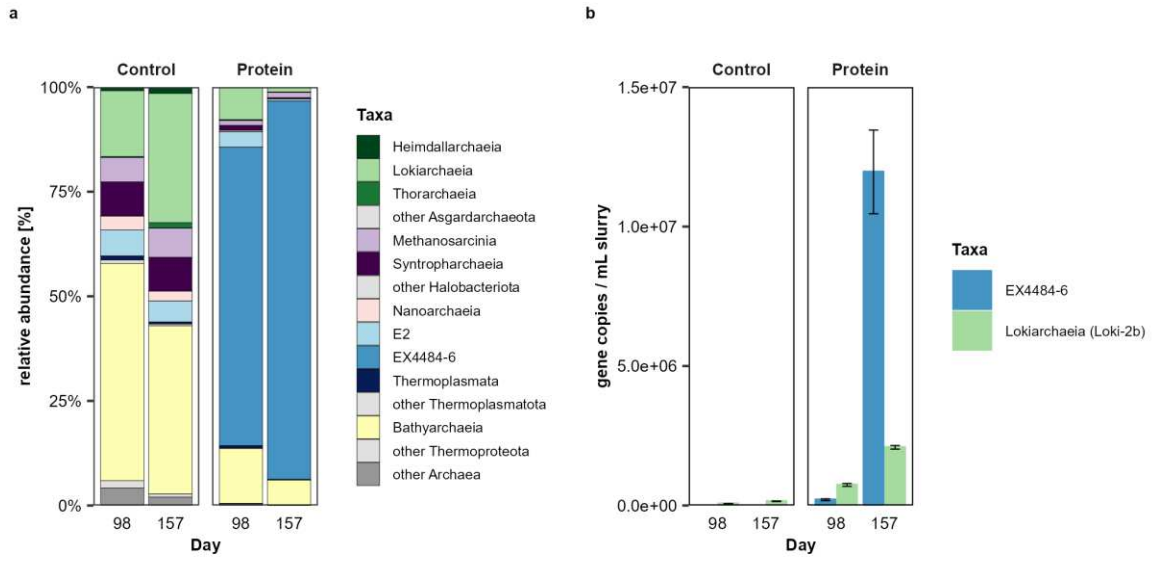

Figure 1 Abundance of archaeal 16S rRNA genes in second generation enrichments on day 98 and 157. (a) Relative 16S rRNA gene abundance within protein samples (amended with protein, sulfate and antibiotics) on day 98 and day 157 and control samples (amended with sulfate and antibiotics). **(b)** 16S rRNA gene copies per mL slurry of the classes EX4484-6 and Lokiarchaeia subgroup Loki-2b in protein-amended samples and control samples.

3.1.3.2 Data mining uncovers 35 MAGs representing the class EX4484-6

Through metagenomic sequencing of the protein and antibiotics amended enrichment from day 157, we retrieved one MAG classified as member of the class EX4484-6. To analyze the phylogeny between the 16S rRNA gene obtained from the EX4484-6 MAG and the highly abundant EX4484-6 ASV obtained through 16S rRNA gene amplicon sequencing, we calculated a 16S rRNA gene phylogenetic tree (Fig. S3). Both, the full length 16S rRNA gene from the MAG and the 16S rRNA gene sequence of the ASV fell into the same branch with a sequence identity of 100%. Both sequences are closest to the cultured representative *Methanomassiliicoccus luminyensis,* with a sequence identity of 83-84% and 100% of the 16S rRNA gene sequence of the ASV and 97% of the 16S rRNA gene from the MAG being covered (Table S2).

To further characterize the novel class, we performed a group-targeted data mining, investigating the phylogenetic breadth, biogeography, abundance and rarity in other habitats, as well as the metabolic capabilities. We increased the number of available genomes of the EX4484-6 group by searching 11,479 genome assemblies of the phylum Thermoplasmatota and additional unclassified genome assemblies (Fig. 2a) published in GenBank. Still, we could only find an additional five published EX4484-6 MAGs, prompting us to screen 57.8 TB of publicly available metagenomic short read data. Using the non-redundant marine EX4484-6 MAGs retrieved through the first screening, we searched 8,287 publicly available metagenomic runs of aquatic origin for the presence of EX4484-6 MAGs (Fig. 2b-e). Sampling bias was reduced by screening samples of 30 different categories, covering coastal and open ocean environments, along with extreme habitats known to host Thermoplasmatota [47-51] (Fig. 2e). In 30 metagenomic studies the class EX4484-6 was found with a cumulative coverage > 2 (Table S3). Medium and high quality EX4484-6 MAGs as per the minimum information about a metagenome-assembled genome (MIMAG) [52] could be reconstructed from the studies (PRJNA368391, PRJNA531756, PRJNA541421, PRJNA704804, PRJNA721298 and PRJNA889212 and were added to our dataset. Further, previously unpublished EX4484-6 MAGs from the Baltic Sea, Benguela upwelling system (Namibia), Cariaco basin (Venezuela), Arabic Sea, Qinghai Lake (China) and the Scotian Slope (Atlantic Ocean), were added to our dataset (Table S4), yielding a total of 35 medium to high quality EX4484-6 MAGs with a completeness of > 80% (80.3% to 97.9%) and contamination of < 5% (0.05% to 4.35%). Genome sizes of the single MAGs were adjusted based on their completeness and varied from the smallest genome size of 1.14 Mbp to the largest genome size of 3.94 Mbp (Table S5).

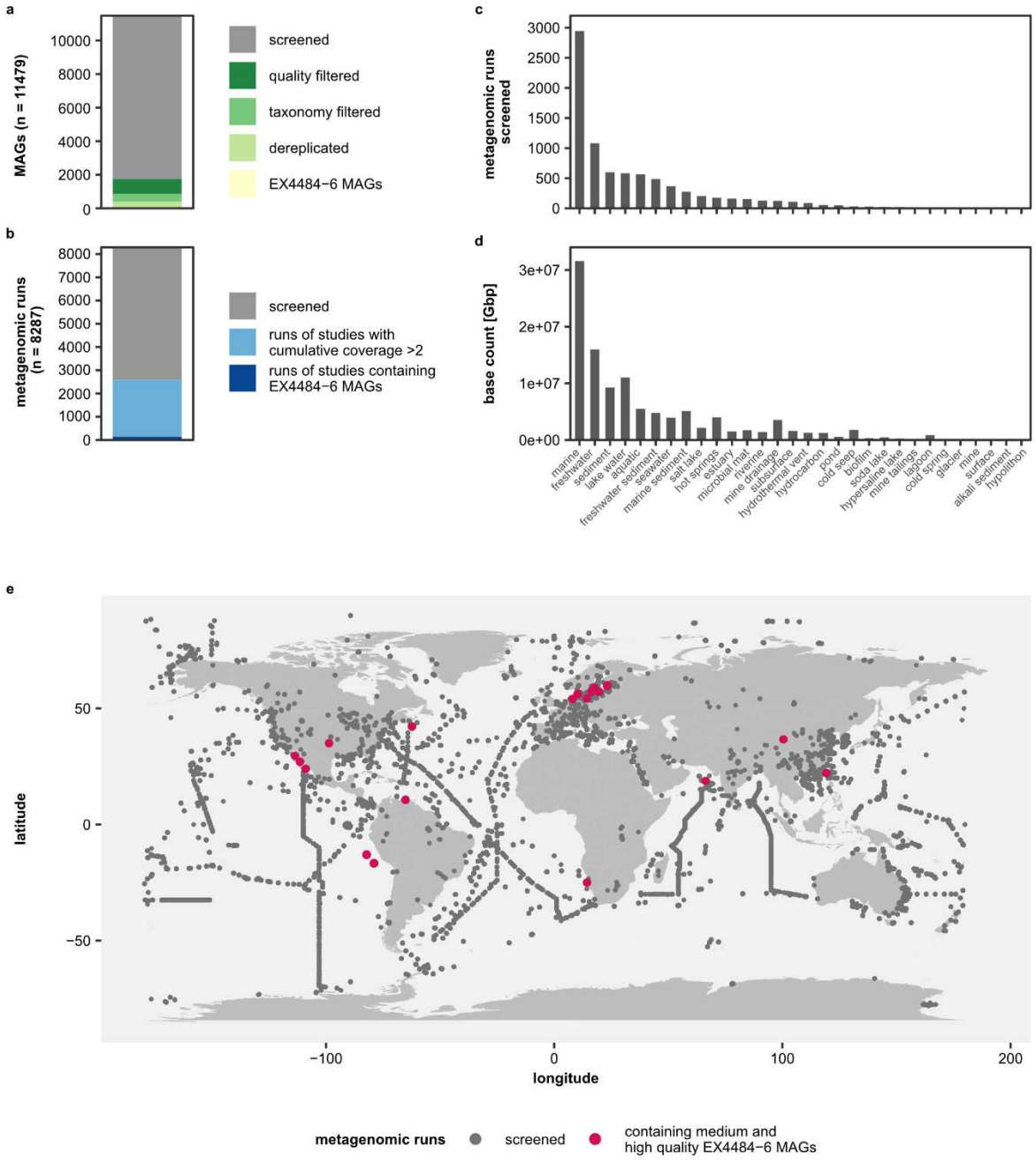

Figure 2 Screened data for retrieving novel EX4484-6 MAGs. Number of data sets and data volume of (**a**) screened published MAGs from GenBank, (**b**) screened metagenomic sequencing runs from the SRA, (**c**) metagenomic sequencing runs screened per environmental category based on scientific names (NCBI taxon id), (**d**) base counts in giga base pairs (Gbp) of metagenomic sequencing runs screened per environmental category based on scientific names (NCBI taxon id). (**e**) World map of all screened metagenomic runs with indication of those locations at which target MAGs were detected or reconstructed.

3.1.3.3 EX4484-6 forms a novel class with four orders

Next, we performed a phylogenomic analysis on 370 Thermoplasmatota species found within the screened 11,479 genome assemblies and all 35 EX4484-6 MAGs. The tree based on 53 archaeal marker genes showed one distinct cluster with high bootstrap support consisting of all

35 EX4484-6 MAGs (Fig. 3a). Taxonomic ranks were assigned to the EX4484-6 MAGs using the relative evolutionary divergenge (RED) rank normalization according to GTDB [37]. Based on the calculated RED, the EX4484-6 cluster can be grouped into four orders (orders 1-4) (Fig. S4), which consist of in total six families (order 1: family 1A, 1B; order 2: family 2; order 3: family 3A, 3B; order 4: family 4). From these only families 1A and 1B were represented in the GTDB database versions 207 and 214. Additional computed average nucleotide identity (ANI) and amino acid identity (AAI) supported the reported RED values (Fig. S5 and S6).

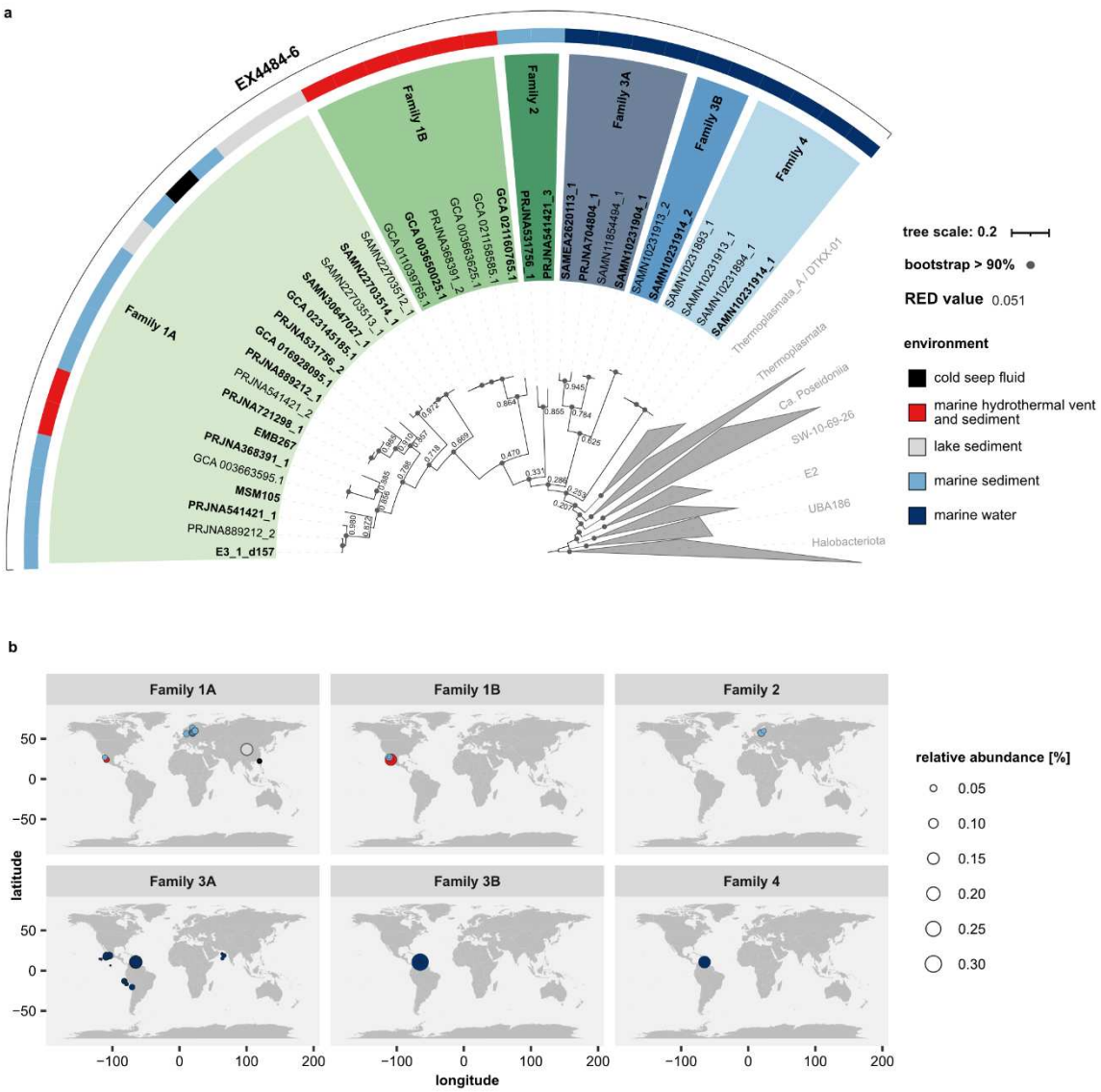

Figure 3 Phylogenomic tree of the Thermoplasmatota based on 53 archaeal marker genes. (a) Maximum-likelihood tree (RAxML, 100 bootstraps) of 370 Thermoplasmatota MAGs and 35 EX4484-6 MAGs obtained through data mining of genome assemblies and metagenomic short-read data sets. MAGs in bold define the non-redundant EX4484-6 MAGs, derived through MAG dereplication (methods section 7). Node labels indicate RED values, which were used to define the phylogeny of the EX4484-6 class into four orders, consisting of 6 families (1A, 1B, 2, 3A, 3B, 4). The environments from which single MAGs are derived are indicated by a colored strip. **(b)** Relative abundance of EX4484-6 families in the environment. The environments individual MAGs were found in are indicated by color.

3.1.3.4 EX4484-6 is globally distributed in coastal and organic rich areas

We aimed to further gain insights into the biogeography of the EX4484-6 MAGs and differences between the habitats of individual families within this class. For this, 8,573 metagenomic runs were mapped against 20 non-redundant EX4484-6 representatives and relative abundances of MAGs in these samples were calculated. From all searched metagenomic sequencing runs, EX4484-6 MAGs could only be detected in 128 runs (Table S6-7). We used stringent detection methods with a minimum breadth of coverage of 50% and a minimum percentage identity of 95% for computation of relative abundances to avoid false positives. Using these strict detection thresholds, three of the non-redundant EX4484-6 MAGs (MSM105, EMB267, GCA_016928095.1) were not found in any of the searched runs. These three MAGs did already cluster with a greater distance in our phylogenomic tree and thus might occupy habitats that were not included in our analysis or for which data was scarce.

Relative abundances of EX4484-6 MAGs in the environment ranged between 0% and a maximum of 0.32%, with on average < 0.05% (Fig. 3b and 4, Fig. S7). While at class level, EX4484-6 could be detected in continental shelf environments worldwide (Fig. S7), single families preferred very distinctive habitats (Fig. 3b and 4). MAGs of the families 1A and 2 were mostly detected in marine surface sediments (Table S7). Family 1A was additionally present in a variety of environments, such as lake and marine sediments, hydrothermal vents and cold seeps, and therefore could be found more globally distributed (Fig. 3b and 4). The MAGs of family 2 were present in marine sediments from the Baltic Sea (Fig. 3b). MAGs of family 1B were only found in hydrothermal sediments from the Guaymas and Pescadero basin (Fig. 3b). All MAGs from the families 3A, 3B and 4 were found in the marine water column, specifically oxygen minimum and deficient zones (Fig. 3b and 4). MAGs of families 3B and 4 were confined to the Cariaco Basin (Venezuela), while family 3A was additionally observed in the Arabic Sea and off the coast of Mexico, Peru and Chile (Fig. 3b). Two of the MAGs originating from the water column (SAMEA2620113_1 and SAMN10231904_1) were found predominantly in the free-living fraction of the seawater community (retained on filters with a pore size 0-5 µm), suggesting a habitat selection of these microorganisms (Fig. 4a). However, most other MAGs found within the water column were derived from bulk seawater samples (retained on filters with 0.2 µm pore size), thus it cannot be resolved whether they were part of the free-living or particle-associated community (Fig. 4a). The class EX4484-6 occurred in less than 1% of all screened data sets (Fig. S8a). Compared to other classes within the phylum of Thermoplasmatota, the class EX4484-6 showed the lowest average relative abundance in the

environment (Fig. S8b). Highest relative abundances among the Thermoplasmatota were found in the classes Thermoplasmata (up to 43%) and *Ca*. Poseidoniia (up to 6%).

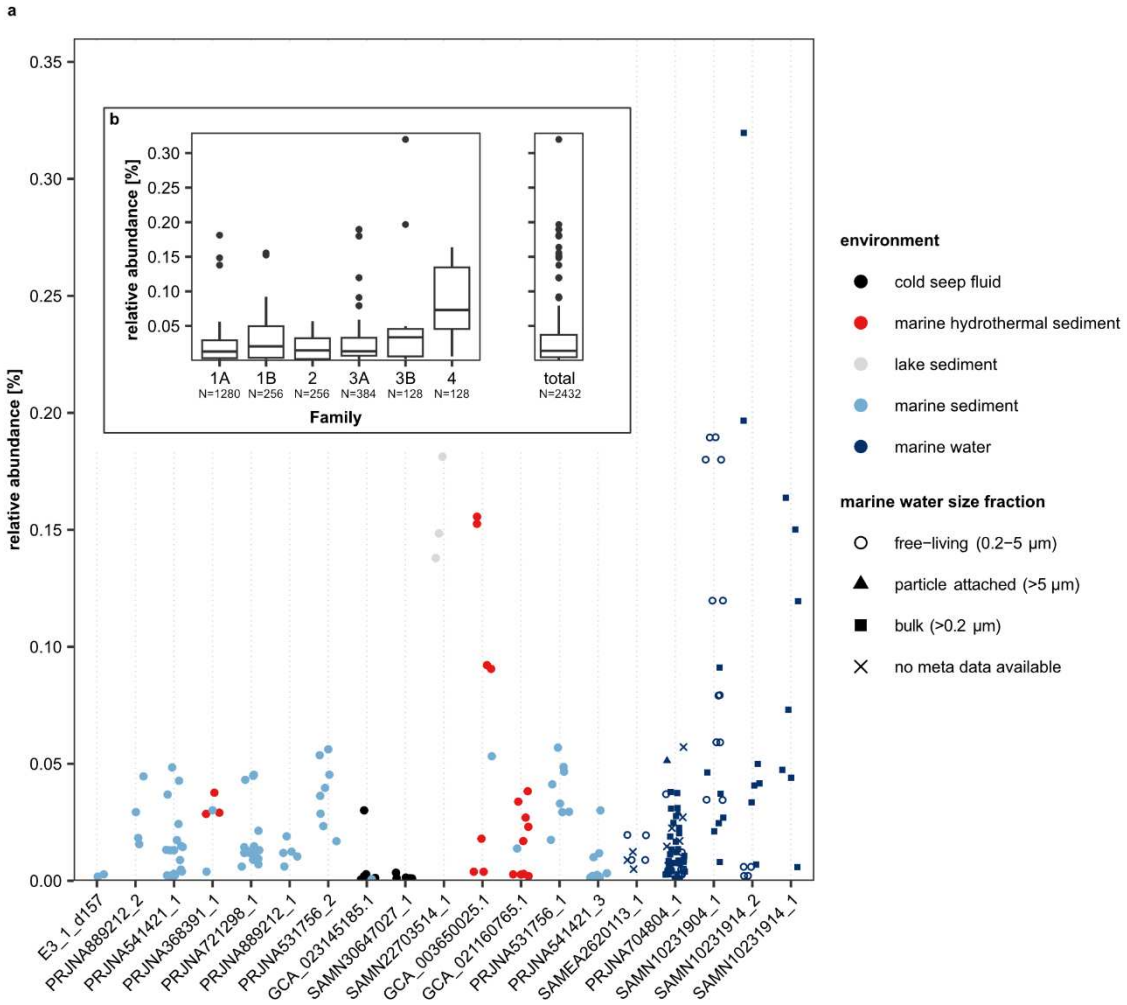

Figure 4 Relative abundances of EX4484-6 MAGs in screened samples for (**a**) individual non-redundant MAGs. For MAGs found in marine water samples, the filter size each sample was filtered through is indicated by shape. Colors indicate the environment of the samples in which the MAGs were found. (**b**) Relative abundance of EX4484-6 MAGs summarized by family and for the class as whole.

3.1.3.5 Housekeeping genes and shared metabolic potential encoded by the core genome of the class EX4484-6

To assess genomic differences among families, orthogroups shared between MAGs of the class EX4484-6 were studied. Orthogroups are defined as a set of genes, which descended from a single last common ancestor [53] and as such are considered to be homologous [54]. For this we performed an NMDS analysis based on orthogroups found within all EX4484-6 MAGs (Fig. S9a). The NMDS showed a clustering of MAGs, which resembled the structure of the phylogenomic tree (Fig. 3a) with MAGs derived from sediment samples and those derived

from water samples forming two well-defined distinct clusters. Generally, the number of orthogroups found per family resembled differences in genome sizes within the families (Fig. S9b, Table S5). Three of the families contained high numbers of family-specific orthogroups. In family 1A, with the highest number of orthogroups (n=2,961), 1,179 family-specific orthogroups (40% of the orthogroups in this family) could be identified (Fig. S9b). Families 3A and 4 from the water column contained 848 (32%) and 800 (40%) family-specific orthogroups, respectively. We further observed that more closely related families within the class EX4484-6 shared more orthogroups (Fig. S10). In total we found 477 orthogroups, which were present in at least 90 % of all 35 EX4484-6 MAGs and as such were defined as core genome of the class EX4484-6 (Table S8). Overall, the core genome encoded expected housekeeping genes involved in gene expression, such as translation, transcription, replication, DNA repair, tRNA biogenesis and ribosomal proteins. Further, genes affiliated with metabolic processes, such as transporters, the gluconeogenesis pathway, TCA cycle, pyruvate metabolism, fatty acid and amino acid degradation were found within the core genome (Tables S8-11). Moreover, genes needed for biosynthetic processes, such as the nucleotide and amino sugar metabolism as well as the biosynthesis of amino acids, fatty acids, lipids, glycans, vitamins and cofactors were encoded (Tables S8-10).

3.1.3.6 Heterotrophy and mixotrophy as main nutritional strategies of the class EX4484-6

As we found our initial EX4484-6 MAG within protein-amended enrichments, we analyzed all MAGs for the potential of protein degradation (Table S11). Metabolic capabilities included amino acid degradation for all MAGs of this novel class (details see Fig. 5). MAGs of the families 1A, 1B and 2 additionally encoded genes for extracellular peptidases of the families C11A, M14B and S08A (Fig. 5, Fig. S11a), required for the first step of protein polymer hydrolysis into smaller peptides, whereas families 3A, 3B and 4 lacked genes for these enzymes. However, MAGs of all families encoded genes for oligopeptide transporters, different aminopeptidases (*pepF, pepT, pepP, pepS, map*) and aminotransferases (Table S12, Fig. S11b). Furthermore, all MAGs encoded aspartate aminotransferase (*aspB*) and alanine aminotransferase (*alaA*) for deamination. Single MAGs encoded an alanine-glyoxylate transaminase (AGXT2), branched-chain amino acid aminotransferase (*ilvE*) and aromatic amino acid aminotransferase. After deamination, the resulting 2-oxoacids can be further converted to acetyl-CoA via pyruvate ferredoxin oxidoreductase (*por*), or to acyl-CoA via indolepyruvate ferredoxin oxidoreductase (*ior*), 2-oxoacid:ferredoxin oxidoreductase (*kor*) or

2-oxoisovalerate ferredoxin oxidoreductase (*vor*). Ferredoxin might serve as electron donor for hydrogen formation by the present NiFe hydrogenase group 3c as suggested by Imachi, et al. (2020) [55]. Energy-rich acyl-CoA can support ATP formation by an acetyl coenzyme A synthetase (*acdAB;* ADP-forming), which was found in all orders. In four families (1B, 2, 3A and 4) also genes for succinyl-CoA synthetase (*sucCD*) were found. The amino acid degradation results in the main products acetate and organic acids. The families 2, 3A and 4 additionally encoded all genes of the beta-oxidation pathway, including a butyryl-CoA dehydrogenase (ACADS), acyl-CoA dehydrogenase (ACADM), enoyl-CoA hydratase (*crt*), 3-hydroxyacyl-CoA dehydrogenase (*fadB*) and acetyl-CoA acyltransferase (*fadA*) to further degrade short and medium chain acyl-CoAs. Family 1B additionally contained a lactate dehydrogenase (*ldhA*), which forms lactate from pyruvate. The findings of protein, amino acid and fatty acid degradation among the families of the EX4484-6 suggest a heterotrophic lifestyle for this class. Moreover, the presence of genes of the Wood Ljungdahl pathway in family 3A could indicate a mixotrophic lifestyle for this family. A more detailed annotation of metabolic and assimilatory pathways can be found in supplementary results for all orders (Results SII, Fig. S12-15, Table S10-12).

3.1.3.7 EX4484-6 families are adapted to environmental stress

MAGs of all families encoded genes for the prevention of oxidative stress, including thioredoxin reductase (*trxR*), thioredoxin (*trxA*), and the desulfoferredoxin (*dfx*), acting as superoxide reductase [56]. Moreover, in MAGs of all families, except family 4, genes for the conversion of hydrogen peroxide to water, catalyzed by peroxiredoxin (*prxQ*), were found [57]. MAGs of the families 1A and 2 further contained genes for the reduction of arsenate via an arsenate reductase (*arsC*), which reduces arsenate As(V) to arsenite As(III). For removal of arsenite from the cell a gene encoding an arsenite transporter (*acr3* / *arsB*) was found in MAGs of family 1A, 2 and 3A. An arsenite methyltransferase (AS3MT) was additionally found in the MAGs of family 1A, 3A, 3B and 4. Lastly, a gene for the defense against antimicrobial drugs was encoded in all MAGs of family 1A, 2 and 4, annotated as a MATE family drug/sodium antiporter, which is driven by a sodium gradient [58].

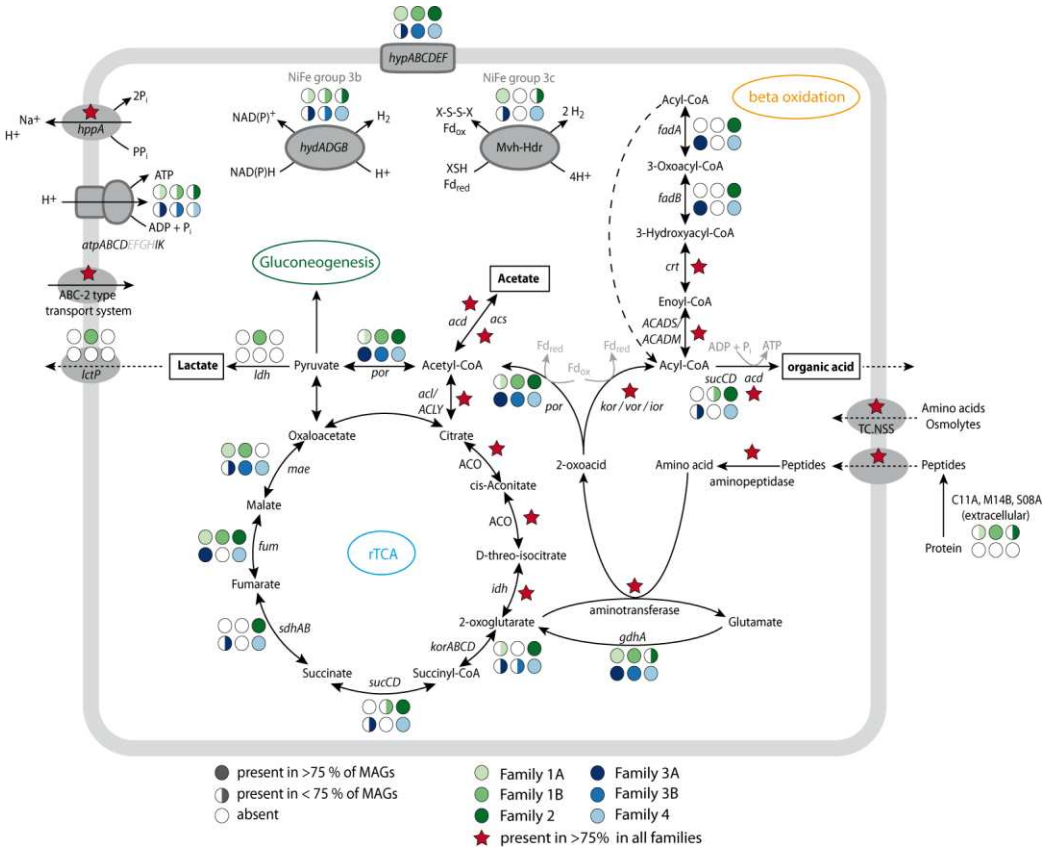

Figure 5 Metabolic reconstruction of the main metabolic features of the novel class EX4484-6. The presence of genes is indicated by full or half circles for each family or with red stars if present in >75% in all families. Amino acid degradation: *gdhA* glutamate dehydrogenase, *kor* 2-oxoacid:ferredoxin oxidoreductase, *vor* 2-oxoisovalerate ferredoxin oxidoreductase, *ior* indolepyruvate ferredoxin oxidoreductase, *por* pyruvate ferredoxin oxidoreductase, *acd* acetyl coenzyme A synthetase; beta-oxidation: ACADS butyryl-CoA dehydrogenase, ACADM acyl-CoA dehydrogenase, *crt* enoyl-CoA hydratase, *fadB* 3-hydroxyacyl-CoA dehydrogenase, *fadA* acetyl-CoA acyltransferase; rTCA: *acl*/ACLY ATP-citrate lyase, ACO aconitate hydratase, *idh* isocitrate dehydrogenase, *korABCD* 2-oxoacid:ferredoxin oxidoreductase, *sucCD* succinyl-CoA synthetase, *sdhAB* succinate dehydrogenase / fumarate reductase, *fum* fumarate hydratase, *mae* malate dehydrogenase (oxaloacetate-decarboxylating); pyruvate metabolism: *acd* acetyl coenzyme A synthetase, *acs* acetyl-CoA synthetase, *por* pyruvate ferredoxin oxidoreductase, *ldh* lactate dehydrogenase; hydrogenases: *hydADGB* sulfhydrogenase, *mvh* F420-non-reducing hydrogenase, *hdr* heterodisulfide reductase, *hypABCDEF* hydrogenase expression/formation protein, *atpABCDEFGHJK* V/A-type H+-transporting ATPase, *lctP* lactate permease, *hppA* K(+)-stimulated pyrophosphate-energized sodium pump, TC.NSS neurotransmitter:Na+ symporter family.

3.1.3.8 Rare MAGs in the phylum Thermoplasmatota hold higher numbers of functionally unknown genes

We observed high numbers of hypothetical and unknown genes in all MAGs of EX4484-6 (Table S14, Fig. S16), besides the low relative abundance of MAGs of this class in the screened metagenomic runs. As this group must be regarded as part of the rare biosphere, this raises the question of whether the number of hypothetical and unknown genes are unexpectedly high in this novel class. Genes were defined as unknown, if the genes could neither be annotated through the non-redundant RefSeq database (NR) nor by KEGG. The percentage of unknown genes among the EX4484-6 ranged between 1.6% to 63% in single MAGs with on average

26% and the highest percentages of unknown genes found in MAGs of the families 2, 3A, 3B and 4, most of which derived from the water column (Fig. S16). Additionally, AGNOSTOS was run to investigate novelty at protein domain level [43]. Based on the AGNOSTOS classification, 14% - 29% of the genes in the EX4484-6 MAGs were classified as genomic unknowns (genes with unknown function, derived from sequenced or draft genomes), and between 0.24% - 17% genes have been characterized as environmental unknown (genes with unknown function, found only in environmental metagenomes or MAGs) (Fig. S17) with no further functional assignment. While classification through AGNOSTOS showed a higher percentage of genes with known protein domains, KEGG and NR annotations could not give functional assignments for these genes, which were therefore regarded as novel genes. Novel genes in this study were defined as those genes that are orphan in function despite possibly being found in other microorganisms.

To test for a correlation between the percentage of unknown genes and the occurrence of the MAGs within the phylum of Thermoplasmatota, we defined a rarity index as median relative abundance of the MAG in the environment, weighted by the fraction of data sets the MAG occurred in across all screened data sets (fraction of occurrence). Based on the rarity index we differentiated between rare (rarity < median rarity) and common (rarity > median rarity) MAGs. MAGs in the category not detected were not found in any of the screened aquatic-derived metagenomic runs (Fig. 6), as these MAGs originated from soil, biodigesters, or human and animal-associated habitats (Table S15). Thus, low or missing relative abundances for these non-aquatic Thermoplasmatota do not necessarily classify these taxa as overall rare but rare in the screened environments. The percentages of unknown genes in the rare group were higher than the percentages of unknown genes in the common group (Fig. 6b). Notably, we found that most of our EX4484-6 MAGs (73%) were defined as rare by our definition (Fig. 6a, Table S16). Further, we observed that the three novel EX4484-6 orders (order 2, 3 and 4) hold a higher percentage of unknown genes compared to any other Thermoplasmatota order (Fig. 6c).

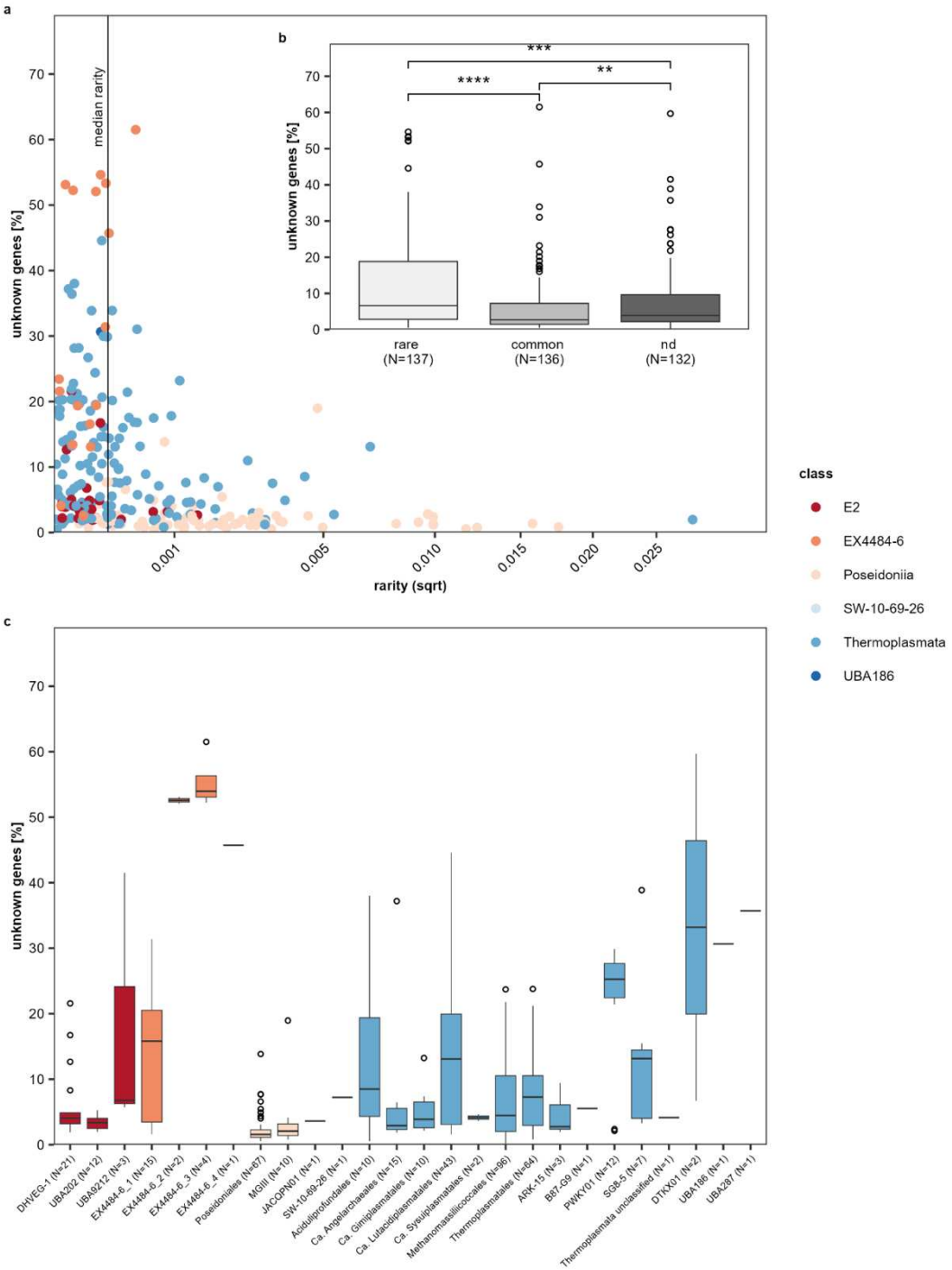

Figure 6 Relationship between gene novelty and MAG occurrence. (a) Percentage of unknown genes vs. rarity for each of the redundant MAGs within the phylum of Thermoplasmatota. The x-axis was square-root (sqrt)-transformed. The rarity index was defined as median relative abundance weighted by the fraction of occurrence. MAGs below the median rarity were defined as rare, MAGs above were defined as common. (**b)** Percentage of unknown genes in genomes of the three defined rarity groups: rare, common and not detected (nd), which contain those genomes, to which none of the screened metagenomic short read data mapped. Differences between groups were tested by Wilcoxon Signed Rank test, Bonferroni adjusted significance threshold: 0.0167, p-values are indicated by asterisks (**** p <= 0.0001, *** p <= 0.001, ** p <= 0.01). Number of observations N indicate number of genomes sorted into each of the defined rarity groups. (**c)** Percentage of unknown genes for each order found within the phylum Thermoplasmatota. Number of observations N represents the number of Thermoplasmatota genomes per order.

Following up on the hypothesis of the rare biosphere acting as gene pool for the community, we investigated if orthogroups found in the EX4484-6 MAGs were also present in other Thermoplasmatota. Results showed differences between orthogroups shared among EX4484-6 MAGs and those shared with other Thermoplasmatota (Wilcoxon Signed Rank test, Bonferroni adjusted *p*-value: 0.0167). While we observed significant differences between orthogroups shared among EX4484-6 MAGs and those shared with other Thermoplasmatota for orthogroups in all three categories (annotated, hypothetical and unknown), the absolute effect size was highest for hypothetical and unknown orthogroups (Impact Effect size test, annotated: 0.3496, hypothetical: 1.1955, unknown: 2.2668). Thus, hypothetical and unknown orthogroups found in the EX4484-6 MAGs were mostly confined to the EX4484-6 class (Fig. S18), while annotated orthogroups were more likely to be shared between EX4484-6 MAGs and other Thermoplasmatota.

3.1.4 Discussion

Only recently, the importance of the rare biosphere in the environment and its role in promoting community stability has been recognized [2, 6, 13, 59, 60]. Studies on global phylogenetic diversity identified not only abundant taxa but also members of the rare biosphere [37, 38, 40, 61]. Yet, studies on the rare biosphere in marine environments are very limited and the roles rare taxa play in these ecosystems remain elusive. We detected the novel and rare class EX4484-6 within the phylum of Thermoplasmatota in protein-amended enrichments. Using these first findings, we combined information gathered from the enrichment experiments and coupled these to extensive group-targeted data mining to investigate organic matter degradation by the EX4484-6. This class has been so far overlooked; it remains unexplored and scarcely represented in INSDC databases.

In this study, we identified the class EX4484-6 as organic matter degrader in sediments from the Helgoland mud area, likely able to survive on protein compounds as sole carbon source, enabled by the presence of genes encoding pathways for degradation of amino acids in all 35 EX4484-6 MAGs (Fig. 5). These results are consistent with findings on several other members of the phylum Thermoplasmatota, which live heterotrophically or mixotophically by degrading organic matter, i.e., utilizing fatty acids, carbohydrates, proteins, peptides and amino acids as their substrates in aquatic environments [45, 46, 62, 63]. Along with genes for the degradation of amino acids, MAGs of sediment inhabiting EX4484-6 families encoded genes for extracellular peptidases involved in protein polymer hydrolysis. In MAGs derived from the water column, genes for protein degradation were not found, contrasting other abundant

planktonic Thermoplasmatota that were shown to degrade proteins [62]. The pelagic EX4484-6 MAGs in our study might have lost the trait for protein degradation or thrive in close proximity to those microbes that host extracellular peptidases. Despite that missing genes in some families could also result from the incompleteness of the analyzed MAGs, our observation suggests different lifestyle preferences among the EX4484-6 families. Although the potential for fatty acid degradation was observed in some of the MAGs, the oxidation reaction is thermodynamically unfavorable and mechanisms for the consumption of reducing equivalents are required [64]. As we did not observe an alternative electron acceptor, electrons might undergo bifurcation via the detected heterodisulfide reductase [65], which was present in some EX4484-6 members with the potential of fatty acid oxidation. Alternatively, the fermentation of fatty acids could also become favorable by a low hydrogen partial pressure in the environment [64]. Therefore, despite showing the potential for fatty-acid degradation, the feasibility of this metabolism remains so far unresolved for members of EX4484-6.

Overall, all data sets containing reads of EX4484-6 MAGs were confined to continental shelf environments. These areas are regarded as organic carbon storage hotspots where large amounts of carbon are supplied through river discharge and land runoff, and large phytoplankton blooms in upwelling areas [66, 67]. Due to high organic matter input and high sedimentation rates, oxygen in such sediments can be depleted within the first millimeters, generating an anoxic environment [68]. Corroborating EX4484-6 presence in anoxic environments, all MAGs from the water column were predominantly found in oxygen minimum and oxygen deficient zones, in which high bioavailability of marine organic matter is prevalent due to high primary productivity in the overlying water column and high respiration causes local anoxia [69-72]. The presence of all MAGs only in such organic, carbon-rich and anoxic environments with high input of marine organic matter agrees with our findings that the class EX4484-6 is involved in amino acid and partially protein degradation, besides being protected against possible oxidative stress.

Additional to the core metabolism shared among the majority of EX4484-6 MAGs, we identified environment-specific adaptations. MAGs retrieved in this study were found in samples from marine sediments, lake sediments, hydrothermal sediments and vents, cold seeps as well as the marine water column. While most EX4484-6 families were found at distinct sites only and may therefore occupy very distinctive habitats, two families (1A and 3A), exhibited a more widespread distribution. We detected a high functional repertoire of the most heterogenous family 1A, reflected by a high number of family-specific orthogroups. Finding the highest number of MAGs in this family reflects its broad distribution.

Specific environmental adaptions could be detected in most of the families. For example, families 1A, 2 and 4 were found in locations of the Baltic Sea and the Cariaco Basin in Venezuela, which are in proximity to riverine input. As such, these locations might be experiencing input of wastewater, which could explain the defensive MATE family drug/sodium antiporter excreting antimicrobial drugs or naturally occurring antibiotics in these MAGs. The presence of such genes might indicate antimicrobial resistance in these human-influenced environments [73-75]. Further, MAGs found in the Baltic Sea carried genes against arsenic toxicity, which is in agreement with high arsenic concentrations reported in this environment [76].

Moreover, the class EX4484-6 did not only exhibit a habitat selection of specific continental shelf environments, but its relative abundance in the environment also did not exceed 0.32%, with an average relative abundance of < 0.05%, making this class a member of the rare biosphere based on currently used definitions [2, 6]. In contrast, other classes within the Thermoplasmatota, such as *Ca*. Poseidoniia and Thermoplasmata were found to have much higher relative abundances in the screened aquatic environments. Only recently, novel taxa, namely *Ca*. Sysuiplasmatales, *Ca*. Lunaplasmatales, *Ca*. Yaplasmales and *Ca*. Gimiplasmatales were described and are characterized by limited environmental distribution and low abundances [51, 77-79]. The continuous detection of novel rare taxa indicates the untapped diversity, which lies within the whole phylum of Thermoplasmatota and might lie within other prokaryotic phyla. Our findings highlight the importance of studying the rare biosphere to understand not only their metabolic functions, including so far undescribed functional diversity, but also evolutionary processes [34], such as the transition from an aerobic to anaerobic lifestyle among members of the Thermoplasmatota [51].

Besides being classified as rare biosphere member, all orders of the class EX4484-6 contained high percentages of unknown and hypothetical genes. An average of 26% of the genes in the EX4484-6 MAGs were unknown. Specifically, MAGs of the families found in the water column exhibited high percentages of unknown genes (43-63%). In general, rare MAGs as opposed to common MAGs within the phylum Thermoplasmatota hold higher percentages of unknown genes (Fig. 6a), with the three newly identified EX4484-6 orders (2, 3 and 4) having the highest percentages of unknown genes among the whole phylum. It is a common feature of uncultivated and rare taxa to contain high numbers of unknown protein families [2, 41] with up to 70% of genes lacking a functional annotation [38, 80]. These results could be expected as annotation databases, such as KEGG and NR, use complete genomes and functionally characterized genes as reference [81, 82] and lack information on uncultivated taxa.

One important hypothesis raised is whether rare biosphere organisms function as genetic seed bank under changing environmental conditions [2, 6, 7]. To investigate if the high numbers of hypothetical and unknown genes found within the EX4484-6 MAGs in our study could provide such a gene pool, we analyzed the presence of these genes in members of all Thermoplasmatota. Only few of the unknown and hypothetical genes found within the EX4484-6 were also shared among other Thermoplasmatota, indicating that lateral gene transfer of the uncharacterized genes among other Thermoplasmatota was absent or very limited. The lack of gene transfer of these uncharacterized genes suggests that these genes rather act as accessory genes than core metabolic functions [83]. Therefore, these unknown and hypothetical genes most likely do not serve as a gene pool for other members in this phylum. The absence of lateral gene transfer for these uncharacterized genes raises the question of which evolutionary pressures and vectors of lateral gene transfer might be inactive. As members of the rare biosphere the class EX4484-6 could undergo less phage predation and as such phage mediated gene transfer might be restricted, yet also genetic distance to other organisms could be indicative of a lack of lateral gene transfer by transduction [84]. However, so far our exploration of the potential for lateral gene transfer is limited to the EX4484-6 MAGs and the phylum Thermoplasmatota and thus, additional insights may be gained by extending the analysis to all prokaryotic lineages.

3.1.5 Conclusions

With our study, we demonstrated how to combine targeted enrichments with metagenomic sequencing and group-targeted data mining to investigate the metabolic potential of the rare biosphere, specifically focusing on the uncharacterized class EX4484-6 within the phylum Thermoplasmatota. We identified the class EX4484-6 as member of the rare biosphere, utilizing proteins and amino acids in organic matter rich and aquatic habitats with limited or no oxygen. Our study revealed habitat-specificity for all families in this class with low abundances in their environments. We used the group EX4484-6 to better understand features of rare microorganisms in the environment and could identify high percentages of hypothetical and unknown genes, compared to other classes of the phylum Thermoplasmatota. The limited lateral transfer of these uncharacterized genes offers an intriguing incentive to further explore the functional and genetic diversity among members of the rare biosphere. Temporally variable abundances and niche preferences of rare biosphere members can be further investigated by sequencing so far underrepresented environments and over a series of time. Our findings not only shed light on the metabolic potential and habitat-specificity of the class EX4484-6 but

also underscore the broader importance of exploring the rare biosphere. Building on this foundation, the data available in public archives present a valuable opportunity to target the generation of new data in areas of underrepresentation, such as the rare biosphere, by first thoroughly reusing what is available. With this study, we highlight the need of targeted and data-integrative approaches to gain further insights into the rare biosphere and unravel functions and metabolic potential that lie within these understudied taxa.

3.1.6 Methods

3.1.6.1 Sample collection and enrichments

Sediment was collected from the Helgoland mud area (54°05’15.5’’N, 7°58’05.5’’E) by gravity coring in 2017 during the RV HEINCKE cruise HE483 [85]. Sediments from a depth of 45-70 cm were selected from gravity core HE483/2-2 for initial slurry incubations. We selected sediments from the Helgoland mud area for these incubations as it is renowned for its high organic content [86, 87]. Anoxic slurry incubations were set up using sediment and sterilized artificial sea water (ASW; composition 26.4 g NaCl, 11.2 g $MgCl_2 \cdot 6\ H_2O$, 1.5 g $CaCl_2 \cdot 2\ H_2O$, 0.7 g KCl and 4.26 g Na_2SO_4 per liter) at a ratio of 1:4 (w/v). A total volume of 50 ml mixed slurry was dispensed in 120-ml serum bottles sealed with butyl rubber stoppers. To remove residual oxygen, the headspace of the serum bottles was exchanged with N_2. The slurry was preincubated for two days. Four replicates were set up for each initial treatment, containing either 1.86 g/l egg white protein or no additional substrate as control. Two of the four treatments were additionally amended with an antibiotic mix (D-cycloserin, kanamycin, vancomycin, ampicillin and streptomycin; 40 mg/l for each). Samples were incubated at 20°C. After 25 days, one replicate of the protein enrichment with antibiotics and one replicate without antibiotics was additionally amended with sodium 2-bromoethanesulfonate (BES; 4 mM) for the inhibition of methanogenesis.

3.1.6.2 Further enrichment of target microorganisms

To retrieve highly enriched Thermoplasmatota from the first generation of enrichments, 5 ml slurry were anoxically transferred into Widdel medium after 372 days of the initial incubations. A total volume of 25 ml medium was anaerobically dispensed into 50 ml serum bottles and the headspace was flushed with N_2:CO_2 (80:20, v/v) (BIOGON C20 E941/E290; Linde, Germany). As basal medium, anaerobic Widdel medium [88] was prepared using salt water medium (20.0 g NaCl, 3.0 g $MgCl_2 \cdot 6\ H_2O$, 0.2 g KH_2PO_4, 0.25 g NH_4Cl, 0.5 g KCl, 0.15 g $CaCl_2 \cdot 2\ H_2O$ per liter): After sterilization, the basal medium was cooled down under N_2:CO_2 (80:20, v/v) atmosphere. Sterilized 1 M $NaHCO_3$ (30 ml), trace element solution SL 10 (1 ml) [89], selenite-

tungstate solution (1 ml) [88], 7 vitamins solution (0.5 ml) [90] and sodium sulfide ($Na_2S \cdot 9 H_2O$; final concentration 0.7 M) were added into 1 l basal medium. As redox indicator, sterilized resazurin (0.2 ml; 0.2 % w/v) was added. The pH was adjusted from 7.2 to 7.4.
All samples were amended with 30 mM sulfate (Na_2SO_4) and an antibiotics mix (D-cycloserin, kanamycin, vancomycin, ampicillin and streptomycin; 50 mg/l for each). Control samples were incubated without an additional carbon source. Protein samples were amended with 2.33 g/l egg white protein as sole carbon source. Samples were incubated at 20°C.

3.1.6.3 Nucleic acid extraction

Two milliliters of slurry were sampled anaerobically from each of the treatments at day 98, 157 and 372 of the second-generation enrichment (section 2) for nucleic acids extraction according to Aromokeye, et al. (2018) [91]. DNA pellets were washed twice with 1 ml cold 70% ethanol and eluted in 100 µl diethylpyrocarbonate (DEPC) treated water (Carl Roth, Germany). The quality of nucleic acids was checked with a NanoDrop 1000 spectrophotometer (Peqlab Biotechnologie, Erlangen, Germany).

3.1.6.4 16S rRNA gene amplicon sequencing

Illumina amplicon sequencing libraries were prepared as described in Aromokeye, et al. (2018) [91] using 30 PCR cycles. Primers targeting the V4 region of the archaeal 16S rRNA were Arc519F (5'-CAGCMGCCGCGGTAA-3') [92] and Arc806R (5'-GGACTACVSGGGTATCTAAT-3') [93]. Thermal cycling conditions were as follows: initial denaturation at 95°C for 3 min, 30 cycles of denaturation at 95°C for 20 s, annealing at 60°C for 15 s, elongation at 72°C for 15 s, final elongation at 72°C for 1 min. Amplicons were generated at Novogene (Cambridge, UK) on the NovaSeq 6000 platform (2 x 250 bp, Illumina) in mixed orientation by ligation, therefore resulting in forward and reverse amplicon orientation in both forward (R1) and reverse read files (R2). Reads were demultiplexed and primer clipped using cutadapt v2.1. [94] and further processed using the package dada2 v1.16.0 [95] in R v4.0.2 [96]. R1 and R2 reads were trimmed to 150 and 160 bp with a maximum error rate of 2. Subsequently, error rates were learned and samples dereplicated and denoised independently for each library orientation, by pooling the data from all samples, using a modified loess function adapted for libraries with binned quality scores [97]. Error-corrected R1 and R2 reads were merged into amplicon sequence variants (ASVs) and sequence tables for forward and reverse orientations were combined by reorientation of the reverse ASVs. Chimeras, ASVs of unexpected lengths (< 249 bp and > 255 bp) and singletons were removed. A bootstrap cutoff of 80 was used to perform taxonomic classification with the assignTaxonomy function of dada2

with a newly formatted GTDB r214 reference database containing the full 16S rRNA gene set. For further processing of archaeal ASVs all non-archaeal taxa were removed.

3.1.6.5 Quantitative PCR (qPCR)

We quantified 16S rRNA gene copy numbers of Archaea, Bacteria and uncultured Thermoplasmatota in the second-generation enrichments. Reaction mixtures contained 1x Takyon Master Mix (Eurogentec, Seraing, Belgium), 4 µg bovine serum albumin (Roche, Mannheim, Germany), 300 nM primers, 1 ng of DNA template or 2 µl of standard in a total reaction volume of 20 µl. The primer pair used for archaea was 806F/912R (5'-GACTACHVGGGTATCTAATCC-3' / 5'-GTGCTCCCCCGCCAATTCCTTTA-3', annealing temperature 58°C) [98], for bacteria 8Fmod/338Rmod (5'-AGAGTTTGATYMTGGCTCAG-3' / 5'-GCWGCCWCCCGTAGGWGT-3', annealing temperature 58°C) [99, 100], for uncultured Thermoplasmatota (EX4484-6) the newly designed pair 472F/633R (5'-CGGTAAATCTCTGGGTAAATCG-3' / 5'-ACCCGTTCTGGTCGGACGCYTT-3', annealing temperature 64°C) and for Lokiarchaeia subgroup 2b a newly designed pair Loki2b_34F/Loki2b_278R (5'-TCCGACTGCTATCCGGGTAA-3' / 5'-TCACGGCCCTTATCGATCAT-3', annealing temperature 60°C). Thermal cycling conditions were as follows: initial denaturation at 95°C for 5 min, 40 cycles of denaturation at 95°C for 30 s, annealing at given temperatures for 30 s, elongation at 72°C for 40 s.

3.1.6.6 Metagenomic analysis

After analyzing the amplicon sequencing data, the protein and antibiotics amended replicate of day 157 (second generation), which showed a high proportion of Thermoplasmatota, was chosen for metagenomic sequencing. An amount of ~ 1 µg extracted DNA was used for metagenomic sequencing on the Illumina HiSeq 4000 platform (2 x 150 bp) at Novogene (Cambridge, UK). Metagenomic reads were adapter and quality trimmed using bbduk from bbmap v38.86 [101]. De novo assemblies were generated with SPAdes v.3.15.5 using the flag *meta* [102] and megahit v1.2.9 using the preset *meta-sensitive* [103]. For read mapping, fasta headers of each of the assemblies were simplified with anvio-7.1 [104]; contigs < 1000 bp were removed and subsequently the quality-trimmed reads were mapped back to the assemblies using bowtie2 v2.3.5.1 [105]. For each assembly, binning was performed with metabat2 v2.12.1 [106] and CONCOCT v1.0.0 [107], followed by bin refinement using the bin refinement module of metaWRAP v1.3.1 [108]. The refined bins of both assemblies were then dereplicated by dRep v3.0.0 using the ANImf algorithm with a primary ANI of 0.9 (mash v2.3

[109]) and a secondary ANI of 0.95 (fastANI v1.32 [110]) with a minimum aligned fraction of 0.5 [111]. Dereplicated bins were reassembled using the bin reassembly module of metaWRAP. Quality of obtained MAGs was calculated with checkM2 v0.1.3 [112] and taxonomic classification was assigned through gtdbtk v2.1.0 [113], based on the GTDB database v207 [37].

3.1.6.7 Retrieval of EX4484-6 MAGs from public archives

To increase the number of EX4484-6 MAGs for further analysis, a list of in total 19,931 genome assemblies, consisting of all Thermoplasmatota as well as ecological and unclassified MAGs according to their NCBI taxon ID, was retrieved through the ENA advanced search interface (date accessed: 22.08.2022; Supplementary Fig. 19). Genome assemblies were filtered based on their scientific names, removing those environments affiliated with anthropogenic activities. The remaining 11,479 MAGs were downloaded from RefSeq [82] or GenBank [114], quality assessed using checkM2 v0.1.3 [112] and filtered based on completeness (> 80%) and contamination (< 5%). Quality filtered MAGs (1,602) were taxonomically classified using gtdbtk v2.1.0 [113]. Only 714 MAGs classified as Thermoplasmatota were kept for further analyses. Additionally, this MAG collection was supplemented by 132 quality filtered Thermoplasmatota MAGs from GTDB [37], which were not yet included in the ENA search output. Furthermore, 34 additional Thermoplasmatota MAGs from previous studies of *Ca*. Lutacidiplasmatales [50], *Ca.* Gimiplasmatales [78], *Ca*. Sysuiplasmatales [51] and uncultured Thermoplasmatota [63] were included in the dataset. The total remaining set of 844 MAGs was dereplicated using dRep v3.0.0 as described above (section 6), resulting in a total number of 388 MAGs. Within this data set, five new non-redundant MAG clusters of EX4484-6 were found, of which four clusters were represented by one single MAG, while one cluster was represented by four MAGs, all of which derived from Guaymas basin hydrothermal sediments [48, 115, 116]. The cluster representatives of four clusters were of marine origin and further used as input for the following mining of metagenomic short read data for EX4484-6.

3.1.6.8 Mining of metagenomic short read data from public archives

Data mining of environmental metagenomes was conducted to increase the number of MAGs for the group EX4484-6 by re-analysis of metagenomic short read data (Fig. S19). For this, a list containing raw reads of 44,968 ecological metagenomes was downloaded from the ENA advanced search (date accessed: 10.11.2022) using the parameters tax_tree(410657), library strategy “WGS”, instrument platform “ILLUMINA”, library source “METAGENOMIC” and

library selection “RANDOM”. Samples without location information, unspecified instrument model, missing ftp links and a base count below 3 Gbp were removed from the selection. Through manual filtering, non-aquatic samples were further removed, leaving a data set of 8,287 metagenomes.

In a first screening step all 8,287 metagenomes were downloaded, quality trimmed using bbduk from bbmap version 38.86 [101] and mapped against the previously identified EX4484-6 clusters (MAG from enrichments, section 6; dereplicated marine EX4484-6 clusters, section 7) using bwa v0.7.17 [117]. The per genome coverage was estimated from the number of mapped reads using coverm v0.6.1 [118] with a minimum read percent identity of 50%, minimum aligned percentage of 50% and a minimum read overlap of 30 bp between read and reference. Cumulative coverage was calculated for all data sets within a given study accession. Studies with a cumulative coverage for EX4484-6 MAGs > 2 were used for de-novo co-assembly and binning, unless MAGs associated with the study had already been published. In these cases, published MAGs were screened for EX4484-6 MAGs along with the catalog of Earth’s microbiome [38], the Ocean DNA MAG catalog [39] and the Ocean Microbiomics Database OMDv2 [40, 119]. MAGs from the OMDv2 were contributed by the co-authors and produced as previously published in Paoli, et al. (2022) [40]. A more detailed pipeline description is included in supplementary methods. Further, high quality EX4484-6 MAGs from unpublished work were contributed to this study by co-authors. Raw reads from studies without published MAGs were downloaded and adapter and quality trimmed using bbduk v38.86 [101]. For each study, de-novo co-assemblies were computed using megahit v1.2.9 [103] with either the preset meta-large, starting at the lowest possible kmer size (27 or 37) for large sample sizes, or a modified preset meta-sensitive starting at kmer size 23. Then, quality trimmed reads were mapped onto the co-assembly using bowtie2 v2.3.5.1 [105], binned, refined and subsequently classified using gtdbtk v2.1.0 [113] with GTDB database v207 as previously described (section 6). If bins of the class EX4484-6 were present, all Thermoplasmatota MAGs within the bin set were reassembled using the metaWRAP bin reassembly module followed again by taxonomic classification. All newly found EX4484-6 MAGs were additionally manually refined using anvio v7.1 [104].

3.1.6.9 Annotation of EX4484-6 MAGs

Genes of EX4484-6 MAGs were predicted using prodigal v2.6.3 [120]. Predicted genes were annotated with the non-redundant RefSeq database (NR) (accessed 23.06.2023) [121] and KEGG release 104 [81, 122] using diamond blastp v2.0.15 [123]. Additionally, we retrieved a

full annotation of all MAGs using InterProScan v5.65-97.0 [124, 125]. Peptidases were identified by diamond blastp against the MEROPS database v12.4 [126]. From these, extracellular peptidases (signal peptides) were determined using signalp v6.0 [127]. Signal peptides were only counted as such, if they were annotated as 'SP' and contained an additional MEROPS annotation. CAZymes were annotated using dbCAN v3.0.7 [128]. To reduce false positive CAZymes in the annotation, all predicted CAZymes were manually validated with the KEGG and NR annotation. Only those CAZymes with more than one annotation according to dbCAN that were also represented in KEGG or NR were counted as positive hits. Lastly, transporters were annotated with the Transporter Classification Database (accessed 12/2023) [129] using diamond blastp v2.0.15 [123] to improve information on transporters found through NR and KEGG. For all blastp searches, a blast score ratio [130] threshold of 0.4 was applied. If metabolic pathways were incomplete, also annotations below the blast score ratio were analyzed.

From all predicted genes for each of the redundant EX4484-6 MAGs, genes were sorted into the categories annotated, hypothetical and unknown based on their annotation status. Annotated genes were defined as genes, which had a functional annotation from KEGG or NR, hypothetical genes were defined from the remaining genes without functional annotation, if they had at least one hypothetical classification and genes without any annotation were defined as functionally unknown genes. To further characterize the unknown genes in the EX4484-6 MAGs, we applied AGNOSTOS v1.1 [43] with default parameters.

To analyze clusters of orthologous genes within the class of EX4484-6, all predicted genes were clustered into orthogroups using orthofinder v2.5.5 with multiple sequence alignment for tree inference [53]. From all orthogroups a Non-Metric Multidimensional Scaling (NMDS) was computed using the R function metaMDS from the package vegan v2.6.4 [131] with Jaccard dissimilarities based on shared orthogroups between MAGs. For visualization of family specific orthogroups an upset plot was created using the package UpSetR v1.4.0 [132]. Orthogroups, which were present in at least 90% of all redundant EX4484-6 MAGs, were defined as core genome and annotations of genes therein were analyzed.

For contextualization of the proportions of unknown genes in the EX4484-6 MAGs, we similarly annotated all Thermoplasmatota genomes (section 7) using NR and KEGG, and sorted predicted genes into the three previously mentioned categories: annotated, hypothetical, unknown. Additionally, orthogroups across all Thermoplasmatota were computed using the rooted phylogenomic tree (section 10) as input tree and sorted based on their annotation status as described previously. The proportion of EX4484-6 orthogroups shared within EX4484-6

MAGs and between Thermoplasmatota and EX4484-6 MAGs was calculated (section 7). Wilcoxon Rank Sum Tests were then applied to test for differences in that proportion between EX4484-6 and other Thermoplasmatota for each annotation status and the effect size for each difference was computed using the package ImpactEffectsize in R [133]. The *p*-value was corrected for executing three pairwise comparisons using the Bonferroni correction.

3.1.6.10 Marker gene tree and phylogenomic analyses

All dereplicated Thermoplasmatota MAGs (section 7), together with all redundant EX4484-6 MAGs (section 8), were used as input for a phylogenomic marker gene tree. In total, the tree contained 419 genomes, including an outgroup consisting of 14 medium to good quality (completeness > 80%, contamination < 5%), randomly selected Halobacteriota genomes. A multiple sequence alignment (MSA) with 53 marker genes was generated using the de novo workflow of gtdbtk v2.1.0 [113]. With the resulting MSA a suitable model (LG+I+R10+F) was identified using ModelFinder [134] as implemented within iqtree2 v2.2.2.7 [135]. The marker gene tree was calculated using raxml-ng v1.1.0 [136] with 20 starting trees. Bootstrap convergence with a bootstrap cutoff at 0.02 was reached after 100 trees. The tree was rooted at the outgroup and manually collapsed based on the GTDB v207 taxonomy [37] using iTol v6.8.1 [137]. Further, relative evolutionary divergence (RED) was calculated using the function getreds from the R package castor [138]. Additionally, average nucleotide identity (ANI) was calculated using fastANI v1.32 [110] and amino acid identity (AAI) was computed using fastAAI v0.1.18 [139].

3.1.6.11 Relative MAG abundance in the environment

All redundant EX4484-6 MAGs were dereplicated using dRep v3.0.0 as described above (section 6), resulting in 20 non-redundant MAGs. Among these clusters, three of the initial four marine cluster representatives (section 7) were found. The fourth cluster was represented by a MAG from a different study, showing higher completeness and lower contamination than the previously obtained cluster representative, while still deriving from the same sediment (Guaymas basin). To quantify EX4484-6 in the environment, a competitive mapping index was constructed from the dereplicated MAGs of the EX4484-6 class and other representatives of the Thermoplasmatota phylum (section 7) using bowtie2 v2.3.5 [105]. For obtaining relative abundances of EX4484-6 MAG representatives in the environment, the initial 8,287 metagenomic runs (section 8) were downloaded along with 286 additional TARA ocean runs, which we initially excluded from our metagenomic run screening as unlikely to contain sufficient EX4484-6 affiliated reads for MAG recovery (section 8). PhiX sequences and

adapters were removed, followed by quality trimming using a minimum read length equal to 2/3 of the initially generated read length or 100 bp for sequencing runs with more than 160 cycles, a trimming quality of 20 and a minimum average quality of 10 with bbduk from bbmap version 38.86 [101]. Quality trimmed reads were aligned to the competitive mapping index using bowtie2. Mean coverage, breadth of coverage and relative abundances of single MAGs were computed using coverm genome v0.6.1 [118] with a minimum breadth of coverage of 50% and a percentage identity of at least 95%.

3.1.6.12 Evaluation of rarity in the phylum Thermoplasmatota

Rarity for each non-redundant EX4484-6 MAG was defined as the median of its relative abundance in the metagenomic runs, where it was detected (section 11), weighted by the fraction of the runs it occurred in of all screened runs. The median rarity across all Thermoplasmatota MAGs was calculated and then defined as cutoff between rare (rarity < median rarity) and common MAGs (rarity > median rarity) in our dataset. Additionally, a third group was contributed by those MAGs that were not detected in any of the screened metagenomic runs. The percentages of unknown genes (section 9) were sorted by rarity groups (common, rare, not detected) and Wilcoxon Rank Sum Tests were applied to test for differences between these three groups. The *p*-value was corrected for executing three pairwise comparisons using the Bonferroni correction.

3.1.7 Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All code used to perform analyses in this study is available on https://github.com/Microbial-Ecophysiology/EX4484-6_data_mining. Amplicon and metagenomic raw reads for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB80318 (https://www.ebi.ac.uk/ena/data/view/PRJEB80318), using the data brokerage service of the German Federation for Biological Data (GFBio [140]), in compliance with the Minimal Information about any (X) Sequence (MIxS) standard [141]. EX4484-6 MAGs obtained and generated in this study and all data tables used for figure

generation are available on zenodo https://zenodo.org/records/10813815. Clone sequences were deposited at GenBank under the accession numbers PQ255994-PQ256084.

Competing interests

The authors declare that they have no competing interests.

Funding statement

This study was supported by DFG under Germany's excellence Strategy, no. EXC-2077-390741603.

Author contributions

XY, MDM and MWF conceived the initial idea for enrichment experiments; MDM and XY carried out enrichment, cloning and characterization experiments; MDM, LCW and TRH performed analyses of amplicon sequencing data; CH conceived the initial idea for data-mining; MDM, CH, SMV, HJR, SS and CV performed bioinformatic analyses for EX4484-6 MAG generation; JF and JP supplied the study with additional genomes; MDM created all illustrations; MWF, CH and XY provided constructive feedback and guided the execution of the project; MDM wrote the original draft; all authors revised the manuscript.

Acknowledgments

We acknowledge funding from the start-up research fund (project-ID XJ2300006031 of Hainan University to XY. We acknowledge funding from ETH Zurich and the Swiss National Science Foundation [205320_215395] to SS, and SMV acknowledges funding from the Human Frontier Science Program through a postdoctoral fellowship [LT0050/2023-L]. For the data contributed by JP and JF we acknowledge the projects MGF-Ostsee and EVAR funded by the German Federal Ministry of Education and Research (BMBF) grant numbers 03F0848A and 03F0814, respectively.

3.1.8 References

1. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR et al. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc. Natl. Acad. Sci. U. S. A. 2006; 103(32):12115-12120.

2. Pascoal F, Costa R, Magalhães C. The microbial rare biosphere: current concepts, methods and ecological principles. FEMS Microbiol. Ecol. 2020; 97(1).

3. Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. Ecology of the rare microbial biosphere of the Arctic Ocean. Proc. Natl. Acad. Sci. U. S. A. 2009; 106(52):22427-22432.

4. Rabinowitz D, Rapp JK, Dixon PM. Competitive abilities of sparse grass species: means of persistence or cause of abundance. Ecology. 1984; 65(4):1144-1154.

5. Rabinowitz D. Seven forms of rarity. In: Synge H, editors. The biological aspects of rare plant conservation. New York: John Wiley & Sons, Inc.; 1981. p. 205-217.

6. Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V et al. Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J. 2017; 11(4):853-862.

7. Lynch MDJ, Neufeld JD. Ecology and exploration of the rare biosphere. Nat. Rev. Microbiol. 2015; 13(4):217-229.

8. Musat N, Halm H, Winterholler B, Hoppe P, Peduzzi S, Hillion F et al. A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. Proc. Natl. Acad. Sci. U. S. A. 2008; 105(46):17861-17866.

9. Pester M, Bittner N, Deevong P, Wagner M, Loy A. A 'rare biosphere' microorganism contributes to sulfate reduction in a peatland. ISME J. 2010; 4(12):1591-1602.

10. Hausmann B, Pelikan C, Rattei T, Loy A, Pester M. Long-term transcriptional activity at zero growth of a cosmopolitan rare biosphere member. mBio. 2019; 10(1):e02189-02118.

11. Dell'Anno A, Beolchini F, Rocchetti L, Luna GM, Danovaro R. High bacterial biodiversity increases degradation performance of hydrocarbons during bioremediation of contaminated harbor marine sediments. Environ. Pollut. 2012; 167:85-92.

12. Griffiths BS, Kuan HL, Ritz K, Glover LA, McCaig AE, Fenwick C. The relationship between microbial community structure and functional stability, tested experimentally in an upland pasture soil. Microb. Ecol. 2004; 47(1):104-113.

13. van Elsas JD, Chiurazzi M, Mallon CA, Elhottovā D, Krištůfek V, Salles JF. Microbial diversity determines the invasion of soil by a bacterial pathogen. Proc. Natl. Acad. Sci. U. S. A. 2012; 109(4):1159-1164.

14. Campbell BJ, Yu L, Heidelberg JF, Kirchman DL. Activity of abundant and rare bacteria in a coastal ocean. Proc. Natl. Acad. Sci. U. S. A. 2011; 108(31):12776-12781.

15. Gilbert JA, Field D, Swift P, Newbold L, Oliver A, Smyth T et al. The seasonal structure of microbial communities in the Western English Channel. Environ. Microbiol. 2009; 11(12):3132-3139.

16. Hugoni M, Taib N, Debroas D, Domaizon I, Jouan Dufournel I, Bronner G et al. Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. Proc. Natl. Acad. Sci. U. S. A. 2013; 110(15):6004-6009.

17. Vergin KL, Done B, Carlson CA, Giovannoni SJ. Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean. Aquat. Microb. Ecol. 2013; 71(1):1-13.

18. Caporaso JG, Paszkiewicz K, Field D, Knight R, Gilbert JA. The Western English Channel contains a persistent microbial seed bank. ISME J. 2012; 6(6):1089-1093.

19. Hamasaki K, Taniguchi A, Tada Y, Kaneko R, Miki T. Active populations of rare microbes in oceanic environments as revealed by bromodeoxyuridine incorporation and 454 tag sequencing. Gene. 2016; 576(2):650-656.

20. Mo Y, Zhang W, Yang J, Lin Y, Yu Z, Lin S. Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes. ISME J. 2018; 12(9):2198-2210.

21. Royo-Llonch M, Ferrera I, Cornejo-Castillo FM, Sánchez P, Salazar G, Stepanauskas R et al. Exploring microdiversity in novel *Kordia* sp. (Bacteroidetes) with proteorhodopsin from the Tropical Indian Ocean via single amplified genomes. Front. Microbiol. 2017; 8.

22. Vavourakis CD, Andrei A-S, Mehrshad M, Ghai R, Sorokin DY, Muyzer G. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. Microbiome. 2018; 6(1):168.

23. Zinger L, Amaral-Zettler LA, Fuhrman JA, Horner-Devine MC, Huse SM, Welch DBM et al. Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. PLoS One. 2011; 6(9):e24570.

24. Crespo BG, Wallhead PJ, Logares R, Pedrós-Alió C. Probing the rare biosphere of the north-west Mediterranean Sea: an experiment with high sequencing effort. PLoS One. 2016; 11(7):e0159195.

25. Donachie SP, Foster JS, Brown MV. Culture clash: challenging the dogma of microbial diversity. ISME J. 2007; 1(2):97-99.

26. Hu B, Xu B, Yun J, Wang J, Xie B, Li C et al. High-throughput single-cell cultivation reveals the underexplored rare biosphere in deep-sea sediments along the Southwest Indian Ridge. Lab Chip. 2020; 20(2):363-372.

27. Rego A, Raio F, Martins TP, Ribeiro H, Sousa AGG, Séneca J et al. Actinobacteria and Cyanobacteria diversity in terrestrial Antarctic microenvironments evaluated by culture-dependent and independent methods. Front. Microbiol. 2019; 10:1018.

28. Lewis WH, Tahon G, Geesink P, Sousa DZ, Ettema TJG. Innovations to culturing the uncultured microbial majority. Nat. Rev. Microbiol. 2021; 19(4):225-240.

29. Saw JHW. Characterizing the uncultivated microbial minority: towards understanding the roles of the rare biosphere in microbial communities. mSystems. 2021; 6(4).

30. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin. Microbiol. Infect. 2012; 18(12):1185-1193.

31. Pascoal F, Magalhães C, Costa R. The link between the ecology of the prokaryotic rare biosphere and its biotechnological potential. Front. Microbiol. 2020; 11:231.

32. Loman NJ, Constantinidou C, Chan JZM, Halachev M, Sergeant M, Penn CW et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat. Rev. Microbiol. 2012; 10(9):599-606.

33. Pedrós-Alió C. The rare bacterial biosphere. Ann. Rev. Mar. Sci. 2012; 4(4):449-466.

34. Spang A, Caceres EF, Ettema TJG. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science. 2017; 357(6351):eaaf3883.

35. ENA - European Nucleotide Archive: Statistics. https://www.ebi.ac.uk/ena/browser/about/statistics (2024). Accessed 13 May 2024.

36. Sequence Read Archive (SRA) summary. https://ftp.ncbi.nlm.nih.gov/genomes/genbank/assembly_summary_genbank.txt (2024). Accessed 13 May 2024.

37. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 2018; 36(10):996-1004.

38. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F et al. A genomic catalog of Earth's microbiomes. Nat. Biotechnol. 2021; 39(4):499-509.

39. Nishimura Y, Yoshizawa S. The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originated from various marine environments. Sci. Data. 2022; 9(1):305.

40. Paoli L, Ruscheweyh H-J, Forneris CC, Hubrich F, Kautsar S, Bhushan A et al. Biosynthetic potential of the global ocean microbiome. Nature. 2022; 607(7917):111-118.

41. Rodríguez del Río Á, Giner-Lamia J, Cantalapiedra CP, Botas J, Deng Z, Hernández-Plaza A et al. Functional and evolutionary significance of unknown genes from uncultivated taxa. Nature. 2024; 626(7998):377-384.

42. Pavlopoulos GA, Baltoumas FA, Liu S, Selvitopi O, Camargo AP, Nayfach S et al. Unraveling the functional dark matter through global metagenomics. Nature. 2023; 622(7983):594-602.

43. Vanni C, Schechter MS, Acinas SG, Barberán A, Buttigieg PL, Casamayor EO et al. Unifying the known and unknown microbial coding sequence space. eLife. 2022; 11:e67667.

44. Medina-Chávez NO, Travisano M. Archaeal communities: the microbial phylogenomic frontier. Front. Genet. 2022; 12.

45. Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD et al. Predominant archaea in marine sediments degrade detrital proteins. Nature. 2013; 496(7444):215-218.

46. Lazar CS, Baker BJ, Seitz KW, Teske AP. Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches. ISME J. 2017; 11(4):1058.

47. Baker BJ, Banfield JF. Microbial communities in acid mine drainage. FEMS Microbiol. Ecol. 2003; 44(2):139-152.

48. Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. Nat. Commun. 2018; 9(1):4999.

49. Reysenbach AL, Liu Y, Banta AB, Beveridge TJ, Kirshtein JD, Schouten S et al. A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. Nature. 2006; 442(7101):444-447.

50. Sheridan PO, Meng Y, Williams TA, Gubry-Rangin C. Recovery of Lutacidiplasmatales archaeal order genomes suggests convergent evolution in Thermoplasmatota. Nat. Commun. 2022; 13(1):4110.

51. Yuan Y, Liu J, Yang TT, Gao SM, Liao B, Huang LN. Genomic insights into the ecological role and evolution of a novel Thermoplasmata order, "*Candidatus* Sysuiplasmatales". Appl. Environ. Microbiol. 2021; 87(22):e0106521.

52. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. 2017; 35(8):725-731.

53. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019; 20(1):238.

54. Fitch WM. Distinguishing homologous from analogous proteins. Syst. Biol. 1970; 19(2):99-113.

55. Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y et al. Isolation of an archaeon at the prokaryote-eukaryote interface. Nature. 2020; 577(7791):519-525.

56. Riebe O, Fischer RJ, Bahl H. Desulfoferrodoxin of *Clostridium acetobutylicum* functions as a superoxide reductase. FEBS Lett. 2007; 581(29):5605-5610.

57. Perkins A, Nelson KJ, Parsonage D, Poole LB, Karplus PA. Peroxiredoxins: guardians against oxidative stress and modulators of peroxide signaling. Trends Biochem. Sci. 2015; 40(8):435-445.

58. Krah A, Huber RG, Zachariae U, Bond PJ. On the ion coupling mechanism of the MATE transporter ClbM. Biochim. Biophys. Acta Biomembr. 2020; 1862(2):183137.

59. Gobet A, Böer SI, Huse SM, van Beusekom JEE, Quince C, Sogin ML et al. Diversity and dynamics of rare and of resident bacterial populations in coastal sands. ISME J. 2011; 6(3):542-553.

60. Wang Y, Hatt JK, Tsementzi D, Rodriguez-R LM, Ruiz-Pérez CA, Weigand MR et al. Quantifying the importance of the rare biosphere for microbial community response to organic pollutants in a freshwater ecosystem. Appl. Environ. Microbiol. 2017; 83(8):e03321-03316.

61. 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(5):16048.

62. Orsi WD, Smith JM, Liu S, Liu Z, Sakamoto CM, Wilken S et al. Diverse, uncultivated bacteria and archaea underlying the cycling of dissolved protein in the ocean. ISME J. 2016; 10(9):2158-2173.

63. Yin X, Zhou G, Cai M, Zhu Q-Z, Richter-Heitmann T, Aromokeye DA et al. Catabolic protein degradation in marine sediments confined to distinct archaea. ISME J. 2022.

64. Monetti MA, Scranton MI. Fatty acid oxidation in anoxic marine sediments: the importance of hydrogen sensitive reactions. Biogeochemistry. 1992; 17(1):23-47.

65. Buckel W, Thauer RK. Flavin-Based Electron Bifurcation, A New Mechanism of Biological Energy Coupling. Chemical Reviews. 2018; 118(7):3862-3886.

66. Atwood TB, Witt A, Mayorga J, Hammill E, Sala E. Global patterns in marine sediment carbon stocks. Front. Mar. Sci. 2020; 7.

67. Burdige DJ. Preservation of organic matter in marine sediments: controls, mechanisms, and an imbalance in sediment organic carbon budgets? Chem. Rev. 2007; 107(2):467-485.

68. Glud RN. Oxygen dynamics of marine sediments. Mar. Biol. Res. 2008; 4(4):243-289.

69. Scranton MI, Astor Y, Bohrer R, Ho T-Y, Muller-Karger F. Controls on temporal variability of the geochemistry of the deep Cariaco Basin. Deep-Sea Res. Part I Oceanogr. Res. Pap. 2001; 48(7):1605-1625.

70. Rabalais NN, Turner RE, Jr. WJW. Gulf of Mexico Hypoxia, A.K.A. "The Dead Zone". Annu. Rev. Ecol. Syst. 2002; 33(1):235-263.

71. Helly JJ, Levin LA. Global distribution of naturally occurring marine hypoxia on continental margins. Deep-Sea Res. Part I Oceanogr. Res. Pap. 2004; 51(9):1159-1168.

72. Wyrtki K. The oxygen minima in relation to ocean circulation. Deep-Sea Res. Oceanogr .Abstr. 1962; 9(1):11-23.

73. Bouki C, Venieri D, Diamadopoulos E. Detection and fate of antibiotic resistant bacteria in wastewater treatment plants: a review. Ecotoxicol. Environ. Saf. 2013; 91:1-9.

74. Schijven JF, Blaak H, Schets FM, de Roda Husman AM. Fate of extended-spectrum β-lactamase-producing *Escherichia coli* from faecal sources in surface water and probability of human exposure through swimming. Environ. Sci. Technol. 2015; 49(19):11825-11833.

75. Munk P, Brinch C, Møller FD, Petersen TN, Hendriksen RS, Seyfarth AM et al. Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. Nat. Commun. 2022; 13(1):7251.

76. Szubska M, Bełdowski J. Spatial distribution of arsenic in surface sediments of the southern Baltic Sea. Oceanol. 2023; 65(2):423-433.

77. Zheng P-F, Wei Z, Zhou Y, Li Q, Qi Z, Diao X et al. Genomic evidence for the recycling of complex organic carbon by novel Thermoplasmatota clades in deep-sea sediments. mSystems. 2022; 7(3):e00077-00022.

78. Hu W, Pan J, Wang B, Guo J, Li M, Xu M. Metagenomic insights into the metabolism and evolution of a new Thermoplasmata order (*Candidatus* Gimiplasmatales). Environ. Microbiol. 2021; 23(7):3695-3709.

79. Zinke LA, Evans PN, Santos-Medellín C, Schroeder AL, Parks DH, Varner RK et al. Evidence for non-methanogenic metabolisms in globally distributed archaeal clades basal to the Methanomassiliicoccales. Environ. Microbiol. 2021; 23(1):340-357.

80. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh HJ, Cuenca M et al. Gene expression changes and community turnover fifferentially shape the global ocean metatranscriptome. Cell. 2019; 179(5):1068-1083.e1021.

81. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023; 51(D1):D587-d592.

82. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016; 44(D1):D733-745.

83. Dmitrijeva M, Tackmann J, Matias Rodrigues JF, Huerta-Cepas J, Coelho LP, von Mering C. A global survey of prokaryotic genomes reveals the eco-evolutionary pressures driving horizontal gene transfer. Nat. Ecol. Evol. 2024; 8(5):986-998.

84. Popa O, Landan G, Dagan T. Phylogenomic networks reveal limited phylogenetic range of lateral gene transfer by transduction. ISME J. 2017; 11(2):543-554.

85. Bijma J. Station list and links to master tracks in different resolutions of HEINCKE cruise HE483, Bremerhaven - Bremerhaven, 2017-04-19 - 2017-04-26. PANGAEA; 2017.

86. Hebbeln D, Scheurle C, Lamy F. Depositional history of the Helgoland mud area, German Bight, North Sea. Geo-Mar. Lett. 2003; 23(2):81-90.

87. Oni OE, Schmidt F, Miyatake T, Kasten S, Witt M, Hinrichs K-U et al. Microbial communities and organic matter composition in surface and subsurface sediments of the Helgoland mud srea, North Sea. Front. Microbiol. 2015; 6.

88. Widdel F. Anaerober Abbau von Fettsäuren und Benzoesäure durch neu isolierte Arten Sulfat-reduzierender Bakterien: Georg-August-Universität zu Göttingen; 1980.

89. Widdel F, Kohring G-W, Mayer F. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. Arch. Microbiol. 1983; 134(4):286-294.

90. Widdel F, Pfennig N. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. Arch. Microbiol. 1981; 129(5):395-400.

91. Aromokeye DA, Richter-Heitmann T, Oni OE, Kulkarni A, Yin X, Kasten S et al. Temperature controls crystalline iron oxide utilization by microbial communities in methanic ferruginous marine sediment incubations. Front. Microbiol. 2018; 9.

92. Ovreås L, Forney L, Daae FL, Torsvik V. Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. Appl. Environ. Microbiol. 1997; 63(9):3367-3373.

93. Takai K, Horikoshi K. Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. Appl. Environ. Microbiol. 2000; 66(11):5066-5072.

94. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. 2011; 17(1):3.

95. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods. 2016; 13(7):581-583.

96. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2020.

97. Salazar G. https://github.com/benjjneb/dada2/issues/938#issuecomment-589657164 (2020). Accessed 15 Jan 2024.

98. Yu Y, Lee C, Kim J, Hwang S. Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. Biotechnol. Bioeng. 2005; 89(6):670-679.

99. Satokari RM, Vaughan EE, Akkermans AD, Saarela M, de Vos WM. Bifidobacterial diversity in human feces detected by genus-specific PCR and denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. 2001; 67(2):504-513.

100. Daims H, Brühl A, Amann R, Schleifer KH, Wagner M. 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. 1999; 22(3):434-444.

101. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Laboratory. 2014; LBNL Report #: LBNL-7065E.

102. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr. Protoc. Bioinformatics. 2020; 70(1):e102.

103. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015; 31(10):1674-1676.

104. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS et al. Community-led, integrated, reproducible multi-omics with anvi'o. Nat. Microbiol. 2021; 6(1):3-6.

105. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 2012; 9(4):357-359.

106. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019; 7:e7359.

107. 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(11):1144-1146.

108. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP - a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018; 6(1):158.

109. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016; 17(1):132.

110. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 2018; 9(1):5114.

111. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017; 11(12):2864-2868.

112. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. PREPRINT bioRxiv. 2022.

113. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the Genome Taxonomy Database. PREPRINT bioRxiv. 2022.

114. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. GenBank. Nucleic Acids Res. 2019; 48(D1):D84-D86.

115. Speth DR, Yu FB, Connon SA, Lim S, Magyar JS, Peña-Salinas ME et al. Microbial communities of Auka hydrothermal sediments shed light on vent biogeography and the evolutionary history of thermophily. ISME J. 2022; 16(7):1750-1764.

116. Zhou Z, Liu Y, Xu W, Pan J, Luo Z-H, Li M. Genome- and community-level interaction insights into carbon utilization and element cycling functions of Hydrothermarchaeota in hydrothermal sediment. mSystems. 2020; 5(1).

117. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754-1760.

118. Woodcraft BJ. CoverM. https://github.com/wwood/CoverM (2007). Accessed 30 Nov 2023.

119. Ocean Microbiomics Database. https://microbiomics.io/ocean2 (2024). Accessed 05 Sept 2024.

120. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 2010; 11:119.

121. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022; 50(D1):D20-d26.

122. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28(1):27-30.

123. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods. 2021; 18(4):366-368.

124. 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. 2020; 49(D1):D344-D354.

125. 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(9):1236-1240.

126. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 2014; 42(Database issue):D503-509.

127. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. 2022; 40(7):1023-1025.

128. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023; 51(W1):W115-W121.

129. Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V et al. The Transporter Classification Database (TCDB): 2021 update. Nucleic Acids Res. 2021; 49(D1):D461-d467.

130. Rasko DA, Myers GSA, Ravel J. Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinform. 2005; 6(1):2.

131. Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P et al. vegan: community ecology package. 2022.

132. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017; 33(18):2938-2940.

133. Lötsch J, Ultsch A. A non-parametric effect-size measure capturing changes in central tendency and data distribution shape. PLoS One. 2020; 15(9):e0239623.

134. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods. 2017; 14(6):587-589.

135. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 2020; 37(5):1530-1534.

136. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019; 35(21):4453-4455.

137. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007; 23(1):127-128.

138. Louca S, Doebeli M. Efficient comparative phylogenetics on large trees. Bioinformatics. 2017; 34(6):1053-1055.

139. Konstantinidis K, Ruiz Pérez C, Gerhardt K, Rodríguez-R L, Jain C, Tiedje J et al. FastAAI: efficient estimation of genome average amino acid identity and phylum-level relationships using tetramers of universal proteins. PREPRINT Research Square. 2022.

140. Diepenbroek M, Glöckner F, Grobe P, Güntsch A, Huber R, König-Ries B et al. Towards an Integrated Biodiversity and Ecological Research Data Management and Archiving Platform: The German Federation for the Curation of Biological Data (GFBio); 2014.

141. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat. Biotechnol. 2011; 29(5):415-420.

3.2 Supplementary

3.2.1 Supplementary Methods

3.2.1.1 Clone library construction

In order to retrieve long sequences for phylogenetic analysis and quantitative PCR (qPCR) standard preparation, a clone library of archaeal 16S rRNA gene fragments (~800 bp) of the protein-amended samples from the first generation enrichments (main methods section 1) was constructed. The primers 109F (5'-ACKGCTCAGTAACACGT-3') [1] and 912R (5'-GTGCTCCCCCGCCAATTCCTTTA-3') [2] were used for PCR with the ALLin RPH polymerase Kit (highQu, Kraichtal, Germany) following the manufacturer's protocol. Thermal cycling conditions included initial denaturation at 95°C for 10 min, followed by 28 cycles with denaturation at 95°C for 30 s, annealing at 52°C for 45 s, amplification at 72°C for 90 s, and final amplification at 72°C for 5 min. PCR products were purified using the Monarch PCR & DNA Cleanup kit (New England Biolabs, Frankfurt am Main, Germany), ligated into the pGEM-t vector (Promega, Mannheim, Germany) and transformed into *Escherichia coli* JM109 competent cells (Promega, Mannheim, Germany) according to the manufacturer's protocol. DNA of randomly selected white colonies was extracted and subsequently PCR amplified with M13 primers (M13F-40; 5'-GTTTTCCCAGTCACGAC-3' and M13b; 5'-CAGGAAACAGCTATG-3') at the following thermal cycling conditions: initial denaturation at 95°C for 5 min, 30 cycles at 95°C for 30 s, 55°C for 45 s, 72°C for 45 s, followed by final amplification at 72°C for 10 min. The PCR was performed using AmpliTaq PCR kit (Applied biosystems, Carlsbad, USA). Amplicons of 64 clones were submitted to LGC Genomics (Berlin, Germany) for Sanger sequencing.

3.2.1.2 Standard preparation for qPCR

Selected clone sequences affiliated with different archaeal groups were used for qPCR standard preparation. The 16S rRNA gene of previously extracted colonies (section I) was amplified with the primer set 109F (5'-ACKGCTCAGTAACACGT-3') and 912R (5'-GTGCTCCCCCGCCAATTCCTTTA-3') using the AmpliTaq PCR kit. PCR was performed using the following thermal cycling conditions: initial denaturation at 95°C for 5 min, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 58°C for 45 s and elongation at 72°C for 60 s, followed by final amplification at 72°C for 5 min. The resulting ~800 bp PCR product was analyzed by agarose gel electrophoresis and purified using the Monarch PCR & DNA Cleanup kit. Concentrations of standard DNA were quantified using the Quant-iT PicoGreen

dsDNA assay kit (Thermo Fisher Scientific, USA). Each qPCR standard was diluted to a concentration of 0.5 ng/µl for further use.

3.2.1.3 qPCR primer design

qPCR primers were designed for quantification of specific uncultured Thermoplasmatota groups present in our first- and second generation enrichments. Clone and sequences of amplicon sequence variants (ASV) of first (main methods section 1) and second (main methods section 2) generation enrichments were aligned using the Silva Incremental Aligner (SINA) v1.2.11 [3], imported into the ARB software v6.0.2 [4] and added to the SILVA 16S rRNA gene phylogenetic base tree (database version 138.1 Ref NR 99; [5, 6]) using the ARB Parsimony tool to retrieve their taxonomic placement. Based on their taxonomic placement, added clone sequences of the target group (class EX4484-6) were selected in the SILVA 16S rRNA gene phylogenetic base tree. Additionally, clone sequences from other archaeal phyla were selected as outgroup. For qPCR primer design, Primer Prospector v1.0.1 [7] was used to calculate possible primer sequences. Calculated primers were checked in ARB for their theoretical specificity against sequences of other closely related groups, selecting only primer sequences with at least two mismatches for non-EX4484-6 taxa in the forward and at least 1 mismatch in the reverse primer. Resulting primer sequences 472f (5'-CGGTAAATCTCTGGGTAAATCG-3') and 633r (5'-ACCCGTTCTGGTCGGA CGCYTT-3') were selected. The primer pair was validated for high specificity against the newly prepared Thermoplasmatota standard (positive control) and other standards (negative controls) of uncultured Thermoplasmata C1, Lokiarchaeia (Loki-2b), Bathyarchaeia (Bathy-8, Bathy-15), ANME-1, SG8-5 and MBGD. Further, a test for efficiency of the new Thermoplasmatota standard at annealing temperatures ranging from 58°C to 64°C was performed, followed by a melting curve stage after PCR, determining the best annealing temperature at 64°C with a final primer concentration of 300 nM. Similarly, a new primer was created for Lokiarchaeia subgroup 2b. The primer pair was validated for high specificity against a Loki-2b standard (positive control) obtained in a previous study [8] and other standards (negative controls) of Lokiarchaeia (Loki-2c, Loki-3), Bathyarchaeia (Bathy-8, Bathy-15), ANME-1, SG8-5 and MBGD.

3.2.1.4 16S rRNA gene phylogenetic tree

For the 16S rRNA gene phylogenetic tree, a total of 10,769 pre-aligned Thermoplasmatota sequences with a minimum sequence length of 1,300 bp, pintail quality > 30, sequence quality > 50 and alignment quality > 50 were downloaded from SILVA (database version 138.1 Ref

NR 99; [5, 6]). Further, 16S rRNA genes were extracted from all redundant Thermoplasmatota MAGs (main methods section 7) and all found EX4484-6 MAGs (main methods section 8) using barrnap v0.9 [9] and aligned using SINA v1.2.11 [3]. Aligned sequences were manually refined in the built-in SINA alignment tool in ARB v7.1. All additional sequences were added to the SILVA 16S rRNA gene phylogenetic base tree (database version 138.1 Ref NR.99) in the ARB software v7.1 [4-6] using ARB parsimony to augment underrepresented Thermoplasmatota groups. Additional to the sequences retrieved from the data-mined Thermoplasmatota and EX4484-6 MAGs, a total of 354 reference sequences with an alignment length of 1,303 bp were selected for de-novo phylogenetic tree reconstruction from the base tree. As outgroup 10 sequences from Halobacteriota were selected. A GTR+I+G4 model for the 16S rRNA gene phylogenetic tree was determined using modeltest-ng v0.1.7 [10], which was further used for calculation of the tree with raxml-ng v1.1.0 [11]. In total 50 starting trees were inferred; bootstrap convergence at a cutoff of 0.03 was reached after 1,300 trees. Shorter 16S rRNA gene sequences found in EX4484-6 MAGs, along with ASVs from the enrichments were added after tree calculation. Short sequences were aligned to the tree alignment with mothur v1.45.3 [12] and placed into the existing tree using epa-ng v0.3.8 [13] and gappa v0.7.1 [14]. Nucleotide blast [15] was used to calculate the similarity between the EX4484-6 16S rRNA gene ASV sq2 and the 16S rRNA gene sequence found in the EX4484-6 MAG retrieved from the enrichment on day 157 (main methods section 6, sequences provided on zenodo (https://zenodo.org/records/10813815).

3.2.1.5 Data collection, processing and MAG reconstruction in OMDv2

A total of 209 publicly available studies were gathered from selected marine metagenomics literature and matched with BioProject identifiers from the European Nucleotide Archive (ENA; Table S17). For this collection, raw read data were downloaded from ENA, and metagenomic data processing was performed as described in Paoli, et al. (2022) [16]. Briefly, sequencing raw reads were filtered using bbmap v.38.06 [17] by removing sequencing adapters from the reads, filtering out reads that mapped to quality control sequences (PhiX library), and discarding low quality reads using the parameters *trimq* = 14, *maq* = 20, *maxns* = 1, and *minlength* = 45. Additionally, read sets from Tara expeditions and from samples that required > 2TB of RAM in the subsequent assembling step were normalized with *bbnorm.sh target* = 40 and *mindepth* = 0. All metagenomes were individually assembled with metaSPAdes (versions from 3.11 to 3.15, depending on when the assembly was performed) [18]. For MAG reconstruction, quality-controlled metagenomic reads from at least 50 samples in OMDv2 were

individually mapped against the scaffolds (≥ 1 kbp) of each sample. Reads were mapped with BWA v.0.7.17-r1188 [19], allowing reads to map at secondary sites (with the *-a* flag). Then, alignments were filtered by length (≥ 45 bp), with identity and coverage of the read sequence values of ≥ 97% and ≥ 80%, respectively. The resulting BAM files were processed using the *jgi_summarize_bam_contig_depths* script of MetaBAT 2 v.2.12.1 [20] to provide within- and between-sample coverages for each scaffold. The scaffolds were finally binned by running MetaBAT 2 on all samples individually with parameters *--minContig* 2000 and *--maxEdges* 500 for increased sensitivity. The quality of each metagenomic bin was evaluated using both the 'lineage workflow' of CheckM v.1.1.3 [21] and anvi'o v.7.1 [22]. Resulting MAGs with CheckM completeness ≥ 50% and contamination ≤ 10% or anvi'o completion ≥ 50% and redundancy ≤ 10% were taxonomically annotated using gtdbtk v.2.1.0 [23] with the default parameters against the GTDB v207 release [24]. These MAGs are available at [25].

3.2.2 Supplementary Results and Discussion

I. Additional qPCR results

During the incubation time of 157 days (second generation), gene copies of Lokiarchaeia increased to 5.02 x 10^7 gene copies per ml slurry. In the second biological replicate, Lokiarchaeia remained the most abundant group throughout the whole incubation period (Fig. S2). With prolonging incubation time, EX4484-6 decreased in the second generation once the abundance of *Lokiarchaeia* had increased.

II. Annotation of the class EX4484-6

An overview of the full annotation can be found in Table S11. We further provide metabolic reconstructions based on annotated genes for each single order (Fig. S12-15).

Carbon metabolism

All orders encoded genes affiliated with peptide and amino acid degradation (Table S11). Families 1A, 1B and 2 encoded genes for extracellular peptidases (peptidase families C11A, M14B, S08A) (Fig. S11a). Moreover, all families encoded genes for oligopeptide transporters, the neurotransmitter:Na^+ symporter of the NSS family, which can catalyze the uptake of nitrogenous substances, such as amino acids and osmolytes [26], different aminopeptidases (*pepF, pepT, pepP, pepS, map*) and aminotransferases (Table S12, Fig. S11b). All orders encoded the aspartate aminotransferase (*aspB*) and alanine aminotransferase (*alaA*). Single MAGs encoded an alanine-glyoxylate transaminase (AGXT2), branched-chain amino acid aminotransferase (*ilvE*) and aromatic amino acid aminotransferase. After deamination, the

resulting 2-oxoacids could be further converted to acetyl-CoA via pyruvate ferredoxin oxidoreductase (*por*), or to acyl-CoA via indolepyruvate ferredoxin oxidoreductase (*ior*), 2-oxoacid:ferredoxin oxidoreductase (*kor*) or 2-oxoisovalerate ferredoxin oxidoreductase (*vor*). Acyl-CoA could further be hydrolyzed in substrate-level phosphorylation by an ADP-forming acetyl-CoA synthetase (*acdAB*) to form corresponding organic acids. The genes for ADP-forming acetyl coenzyme A synthetase were found in all orders. In four of the families (1B, 2, 3A and 4) also genes for succinyl-CoA synthetase (*sucCD*) were found, which could catalyze the conversion of succinyl-CoA to succinate.

The families 2, 3A and 4 additionally encoded all genes of the beta-oxidation pathway (Fig. S17), including a butyryl-CoA dehydrogenase (ACADS), acyl-CoA dehydrogenase (ACADM), enoyl-CoA hydratase (*crt*), 3-hydroxyacyl-CoA dehydrogenase (*fadB*) and acetyl-CoA acyltransferase (*fadA*) to further degrade short and medium chain acyl-CoAs. The amino acid degradation would result in the main products being organic acids. Family 1B additionally contained a lactate dehydrogenase (*ldhA*), with which lactate could be formed from pyruvate.

MAGs of the class EX4484-6 contained between 2 to 14 different carbohydrate active enzymes (CAZymes), with most CAZymes being annotated as glycosyl transferases of family GT2 and GT4 (Table S12). Glycosyltransferases are involved in biosynthesis of glycosidic bonds and as such not involved in the metabolic degradation of sugars [27]. Some MAGs also encoded glycoside hydrolases, which hydrolase glycosidic bonds and might therefore be involved in carbohydrate degradation. Within our dataset only 21 of 35 MAGs encoded different glycoside hydrolases. Most of the MAGs contained between one and three glycoside hydrolases (Table S13). Overall, these low CAZyme counts for archaea have been observed before [28] and demonstrate the limited potential of archaea in carbohydrate degradation. While some MAGs found in this study might take part in the breakdown of carbohydrates, we did not observe a common feature for all families.

All orders within the class EX4484-6 encoded genes for a partial reverse citric acid (rTCA) cycle, including an ATP-citrate lyase (*aclAB* / ACLY), aconitate hydratase (ACO) and isocitrate dehydrogenase (*idh*) (Table S11). The ATP-citrate lyase, catalyzing the cleavage of citrate into acetyl-CoA and oxaloacetate, is regarded as key enzyme for the rTCA cycle [29]. Additionally, most of the families encoded genes for fumarate hydratase (*fum*) and malate dehydrogenase (*mae*). Only all MAGs in families 2 and 4 encoded genes for 2-oxoglutarate ferredoxin oxidoreductase (*korABCD*), succinyl-CoA synthetase (*sucCD*) and succinate dehydrogenase (*sdhAB*), therefore possessing a complete rTCA cycle. During the rTCA cycle,

CO_2 is fixed in conversions of succinyl-CoA to 2-oxoglutarate and in conversion of 2-oxoglutarate to isocitrate [30]. For other carbon fixation pathways, namely the 3-Hydroxypropionate cycle, 3-Hydroxypropionate/4-Hydroxybutyrate cycle or Dicarboxylate/4-Hydroxybutyrate cycle no complete pathways were encoded in any of the families (Table S11). The presence of inorganic carbon fixation via the rTCA was previously suggested for other groups within the Thermoplasmatota, such as Thermoplasmatota_A [30], which clustered closest to the EX4484-6 class in our phylogenomic tree (Fig. 3a) and the *Ca*. Proteinoplasmatales [31].

In contrast to any other family, genes involved in the Wood Ljungdahl pathway could only be detected in family 3A (Fig. S14). All of the MAGs contained genes for acetyl-CoA decarbonylase/synthase, CODH/ACS complex subunit beta, gamma and delta (*cdhCDE*). Besides, genes for methylenetetrahydrofolate reductase (*metF*), methylenetetrahydrofolate dehydrogenase (*folD*), formate--tetrahydrofolate ligase (*fhs*), 5-methyltetrahydrofolate corrinoid/iron sulfur protein methyltransferase (*acsE*) and the alpha subunit of formate dehydrogenase (*fdhA*) were present. However, the catalytic subunit of anaerobic carbon-monoxide dehydrogenase, catalyzing the reduction of CO_2 to CO could not be detected in any of the MAGs. As all MAGs within this family have a completeness between 80-91%, these missing genes in the Wood Ljungdahl pathway might be due to incomplete genomes.

Carbon assimilation

Carbon might be assimilated through gluconeogenesis and the pentose phosphate pathway (PPP) to form nucleic acids via phosphoribosylpyrophosphate (PRPP). Pyruvate formed through degradation of amino acids or formed from acetyl-CoA during amino acid breakdown via the rTCA cycle might be converted via the gluconeogenesis pathway to glyceraldehyde-3P. All families encoded genes for pyruvate dikinase (*pps*), enolase (*eno*), phosphoglycerate mutase (*gpmA*) or 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (*gpmM*), phosphoglycerate kinase (*pgk*) and glyceraldehyde-3-phosphate dehydrogenase (*gap*) (Table S11, Fig. S12-15). Further, all families, except family 3B, encoded all genes of the non-oxidative PPP, including a transketolase (*tkt*), transaldolase (*tal*), ribose-5-phosphate isomerase (*rpi*), ribulose-phosphate 3-epimerase (*rpe*) and ribose-phosphate pyrophosphokinase (*prsA*). Family 3B lacked genes of the transketolase, transaldolase and ribulose-phosphate 3-epimerase. However, family 3B was the only family, which encoded a glucokinase (*glk*), suggesting that the family might use either the glycolysis or the gluconeogenesis pathway. Since neither genes for the pyruvate kinase (*pyk*) nor the fructose-1,6-bisphosphatase (*fbp*)

could be found, no clear prediction could be made for the carbon assimilation potential of this family.

Hydrogenases and energy conservation

Multiple subunits of the oxidative phosphorylation complex I were only encoded in family 1A. Genes for two of four subunits of the succinate dehydrogenase/fumarate reductase (*sdhAB*) were found in families 2 and 4, those families which encoded a full rTCA cycle. Of complex V multiple subunits of the V/A-type H+-transporting ATPase (A_1A_0-ATP synthase, *atpABCDEFGIK*) for energy conservation through a sodium gradient were encoded by all families, with most families lacking subunits *atpE* and *atpG*. None of the families encoded genes involved in the oxidative phosphorylation complexes III and IV (Table S11). The lack of most genes involved in oxidative phosphorylation suggests an anaerobic lifestyle for the class EX4484-6.

All families encoded genes for a K(+)-stimulated pyrophosphate-energized sodium pump (*hppA*), which could couple the hydrolysis of pyrophosphate to the transport of sodium across the membrane against an electrochemical gradient [32, 33]. The sodium gradient could further be used to form ATP via the detected ATPase [34]. Moreover, all MAGs of families 3A, 3B and 4 contained genes for the H+/Na+-translocating ferredoxin:NAD+ oxidoreductase (*rnfABCDEG*), which oxidizes reduced ferredoxin and reduces NAD for energy conservation, while transporting ions across the cytoplasmic membrane [35]. The sodium gradient resulting from the *rnf* complex could be further used by the encoded V/A-type H+-transporting ATPase (A_1A_0-ATP synthase), as has been hypothesized previously [35].

All families encoded genes for hydrogenase nickel incorporation proteins and hydrogenase expression proteins (*hypABCDEF*) required for biosynthesis and Ni insertion of NiFe hydrogenases [36]. Genes for NiFe group 3b sulfhydrogenases were present in most of the families. However, in most families only genes for two subunits, which function as hydrogen dehydrogenase (*hydAD*) were found. Specifically, family 4 lacked the beta and gamma subunits (*hydBG*), which function as sulfur reductase [37], and family 2 did not encode any subunit of the sulfhydrogenase. The sulfhydrogenase was shown to catalyze the reduction of elemental sulfur or polysulfides to hydrogen sulfide and the oxidation or reverse reaction of hydrogen with NAD(P)+ as electron acceptor in the absence of sulfur [37, 38]. As no other sulfur related genes were found in any of the MAGs, it is most likely that the sulfhydrogenase might function as a hydrogen dehydrogenase only.

MAGs of families 1A, 2, 3A and 4 further encoded the NiFe group 3c F420-non-reducing hydrogenase (*mvhADG*) and heterodisulfide reductase (*hdrABC*). While this hydrogenase is usually associated with methanogenic archaea [39-42], in the EX4484-6 class it might rather be involved in the reduction of ferredoxin and an unknown disulfide reducer, coupled to the oxidation of hydrogen, as was also suggested for *Ca*. Lokiarchaeum prometheoarchaeum [43] and the hyperthermophilic *Panguiarchaeum symbiosum* [44].

Lastly, MAGs of the families 1A and 1B encoded three of the fourteen subunit containing *NiFe* group 4d membrane bound hydrogenase (*mbh*), which was suggested to function as a redox-driven ion pump, generating a proton motive force through reduction of protons with a low-potential ferredoxin, thereby producing hydrogen [45]. Along with the three found hydrogenase subunits, MAGs of families 1A and 1B encoded multicomponent Na^+:H^+ antiporter subunits (*mnhBCDEFG*), and gene homologs of NADH-quinone oxidoreductase (*nuoBCDEFH*), which were shown to have similarities to multiple subunits of the NiFe group 4d membrane bound hydrogenase [45]. Genes of multicomponent Na^+:H^+ antiporter subunits, NADH-quinone oxidoreductase and the found membrane bound hydrogenase subunits were located along single contigs, possibly forming operons, which suggests that MAGs of order 1 might contain a *mbh* like hydrogenase.

Transporters

Besides various ABC transporters, all families contained genes for ABC-2 type transport system ATP-binding proteins and ABC-2 type transport system permease proteins, for which no function could be assigned (Table S11). Iron, an essential nutrient in microorganisms, is required for enzymatic processes and as cofactor for proteins involved in respiration, oxidative stress resistance or gene regulation [46]. Genes encoding the iron (II) transport system (*feoAB)* could be found in families 1A, 1B, 3B and 4. MAGs of families 2 and 3A only encoded subunit *feoB*. Ferrous iron in its reduced form is only present in oxygen-limiting, anoxic and low pH conditions [47]. Since these families only encoded a ferrous iron transport system and are present in marine sediments, in which neutral pH is prevalent [48-50], the MAGs found in the EX4484-6 class most likely live in anaerobic conditions [51].

Similarly to iron, zinc functions as cofactor in enzymatic reactions [52]. MAGs of families 1A, 1B and 4 encoded the cytoplasmic membrane zinc transporter *zupT*, which is responsible for the uptake of zinc and other metal ions [53]. Magnesium, another cofactor in enzymatic reactions and involved in the catalysis of ribozymes [54], RNA splicing and stabilization of proteins and RNA [55], is transported into the cell via the magnesium transporter *corA* [56,

57]. Genes for *corA* were found in families 1A and 2. Genes encoding a transporter for cobalt and nickel (*cbiMOQ*) could only be found in family 4. Nickel and cobalt also function as cofactors in enzymes, such as in NiFe hydrogenases [58] or in the corrin ring of coenzyme B12 [59].

Moreover, all families except family 4 encoded genes for a low-affinity inorganic phosphate transporter of the PiT family. Inorganic phosphate acts as a key nutrient in cells, as it is important for cellular building blocks, such as nucleic acids, phospholipids, teichoic acids and membranes. MAGs of family 1A and 1B additionally encoded genes for a high affinity phosphate transport system (*pstSABC*), which is induced in low phosphate concentrations. An accompanying *phoU* regulon was encoded and is required for the repression of the phosphate transport system *pstSABC* at high phosphate conditions to avoid uncontrolled phosphate uptake [60-62], therefore actively controlling phosphate transport. Furthermore, the *trk* system potassium uptake protein (*trkAH*) was encoded in all families, except family 4. Potassium plays a role in cellular homeostasis, osmotic tolerance, pH stress response and membrane potential maintenance [63, 64]. The *trkAH* was shown to act as an ATP- and ADP-gated ion channel, with *trkH* being classified as potassium transport protein in low external potassium concentrations [65]. MAGs of family 1B and 3A additionally contained genes for the high affinity molybdate/tungstate transport system ATP-binding protein *wtpABC* [66], while MAGs of family 4 encoded genes for the lower affinity tungstate transport system *tupABC* [67]. A gene for a tungstate transport system ATP-binding protein (*tupC*) was missing in all MAGs of family 4. Both, molybdate and tungstate are known to function as cofactors in enzymatic reactions [68, 69].

Along with a controlled import, cells actively export ions from the cell to maintain homeostasis. The families 1A, 2, 3A, 3B and 4 contained genes encoding a P-type calcium transporter (ATP2C). These P-type ion transporting ATPases transport cations across membranes, while coupling the transport to the hydrolysis of ATP [70].

Besides, MAGs of family 1A, 1B, 3A and 3B encoded an acetate uptake transporter (*satP*), which has an affinity for acetate and succinate [71]. Acetate can have a regulatory function for processes, such as motility and stress response [72]. Furthermore, acetate could be used as a carbon source for metabolic functions [72].

Stress response

All families of the class EX4484-6 contained genes related to environmental stress (Table S11). All families encoded genes for thioredoxin reductase (*trxR*), thioredoxin (*trxA*), and

desulfoferredoxin (*dfx*), acting as superoxide reductase [73], all of which prevent oxidative stress. Moreover, in MAGs of all families, except family 4, genes for peroxiredoxin (*prxQ*) were found, which catalyzes the conversion of hydrogen peroxide to water [74]. Additionally, MAGs of all families encoded DnaK. This enzyme functions as Heat Shock Protein-70 (Hsp70), which protects cells from heat and oxidation [75, 76].

To maintain cellular homeostasis, MAGs encoded a sodium:calcium antiporter (*yrbG*), which actively removes calcium ions from the cell by using a sodium gradient [77]. Genes for *yrbG* were found in all families, except family 3A.

Additionally, MAGs of all families contained genes for the P-type Cu+ transporter (*copA*) involved in maintaining copper homeostasis [78, 79]. While copper, in low concentrations, is required for enzymatic function, excess copper can become toxic for the cell and therefore needs to be exported [80, 81]. Another enzyme involved in heavy metal resistance is the cobalt-zinc-cadmium efflux system protein (*czcD*), for which genes were found in families 1A, 1B, 3B and 4 [82].

Besides active transport of heavy metal ions from the cell, microorganisms developed other functions to reduce toxicity. Toxic arsenate, as a structural homolog of phosphate, can be imported into the cell through phosphate transporters [83]. MAGs of the families 1A and 2 contained genes for the reduction of arsenate via an arsenate reductase (*arsC*), which reduces arsenate As(V) to arsenite As(III). A gene encoding an arsenite transporter (acr3 / *arsB*) to remove arsenite from the cell was found in MAGs of family 1A, 2 and 3A. An arsenite methyltransferase (AS3MT) was additionally found in the MAGs of family 1A, 3A, 3B and 4. This enzyme has the capability to detoxify arsenic by methylation of the toxic compound. Most of arsenic found in the environments derives from anthropogenic activity and can be introduced into environments through rivers, surface run-off and munition [84-87]. Genes against arsenic toxicity were mostly found in MAGs deriving from the Baltic Sea, which was shown to contain localized high arsenic concentrations [88].

Lastly, all MAGs of family 1A, 2 and 4 contained genes annotated as the MATE family drug/sodium antiporter, a defensive mechanism against antimicrobial drugs, driven by a sodium force [89]. Besides antimicrobial resistance against naturally occurring antibiotics, MAGs of these families were found in near-shore environments, such as the Baltic Sea or the Cariaco Basin in Venezuela, which are experiencing input of waste water and with this, antimicrobial drugs, might be causing antimicrobial resistance in these environments [90, 91].

3.2.3 Supplementary Figures

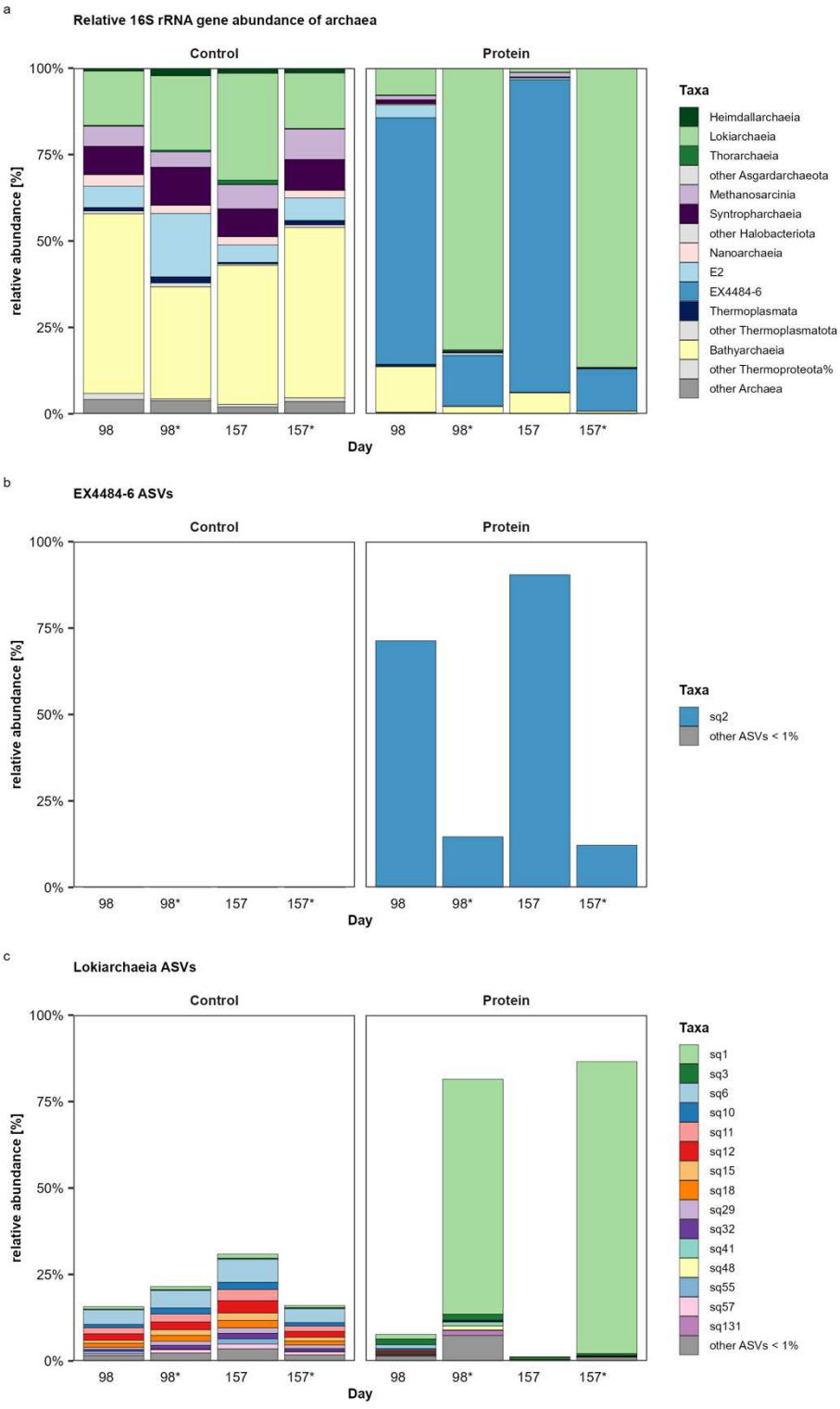

Figure S1 Microbial community composition of second-generation enrichments. Relative abundances of **(a)** archaeal 16S rRNA genes, (**b**) EX4484-6 ASVs and (**c**) Lokiarchaeia ASVs at day 98 and 157 in two replicates each of control samples and protein amended samples, both amended with 30 mM sulfate (Na_2SO_4) and an antibiotics mix (D-cycloserin, kanamycin, vancomycin, ampicillin and streptomycin; 50 mg/l for each). Protein samples were additionally amended with 2.33 g/l egg white protein.

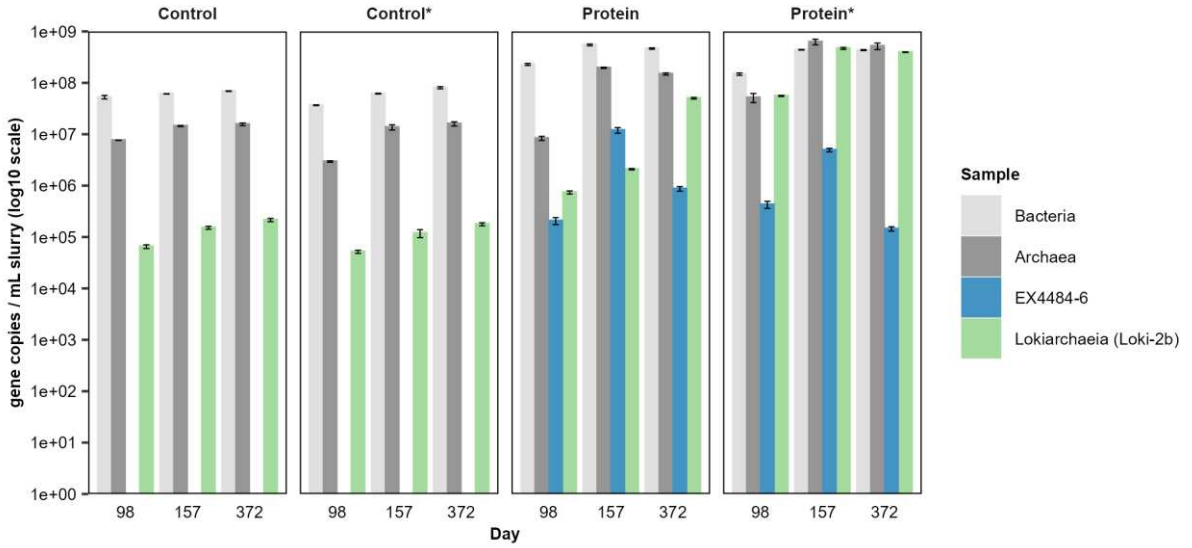

Figure S2 Microbial community composition of second-generation enrichments. 16S rRNA gene copies per ml slurry of Bacteria, Archaea, the class EX4484-6 and Lokiarchaeia subgroup Loki-2b. Gene copies are shown for both replicate enrichments (Replicate 1, Replicate 2*) of control samples and protein amended samples, both amended with 30 mM sulfate (Na_2SO_4) and an antibiotics mix (D-cycloserin, kanamycin, vancomycin, ampicillin and streptomycin; 50 mg/l for each). Protein samples were additionally amended with 2.33 g/l egg white protein. Error bars represent standard deviations of the three technical qPCR replicates per sample.

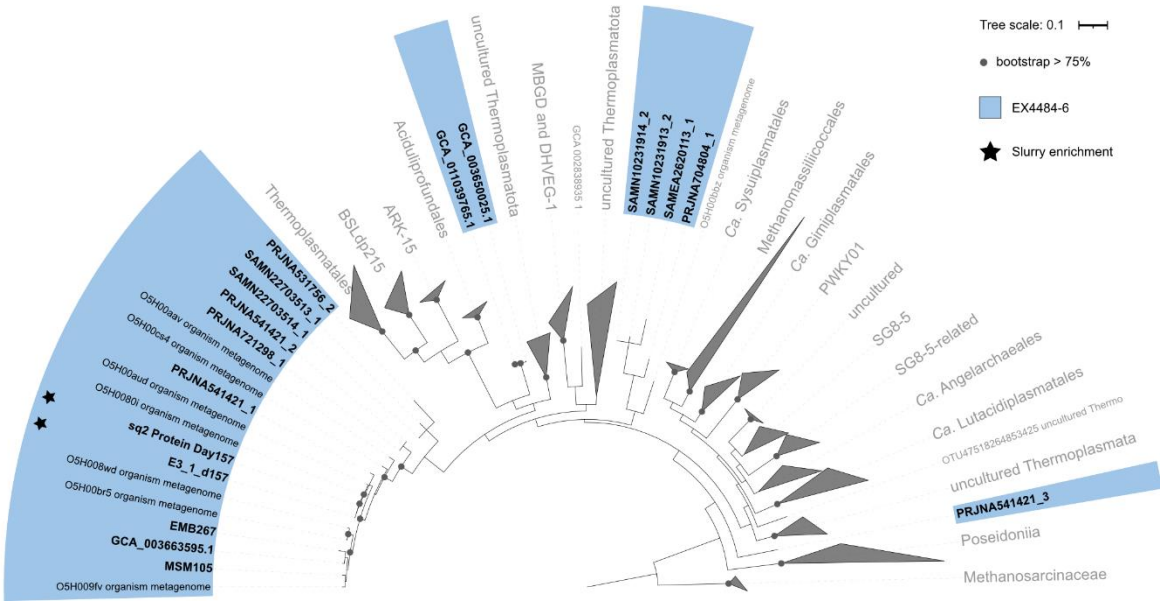

Figure S3 16S rRNA gene phylogenetic tree. Maximum-likelihood tree (RAxML, convergence reached after 1300 bootstraps) of 354 full length Thermoplasmatota sequences, including all retrieved full length EX4484-6 16S rRNA genes from single MAGs. Shorter 16S rRNA gene sequences of ASVs and those found in EX4484-6 MAGs were added to the existing tree.

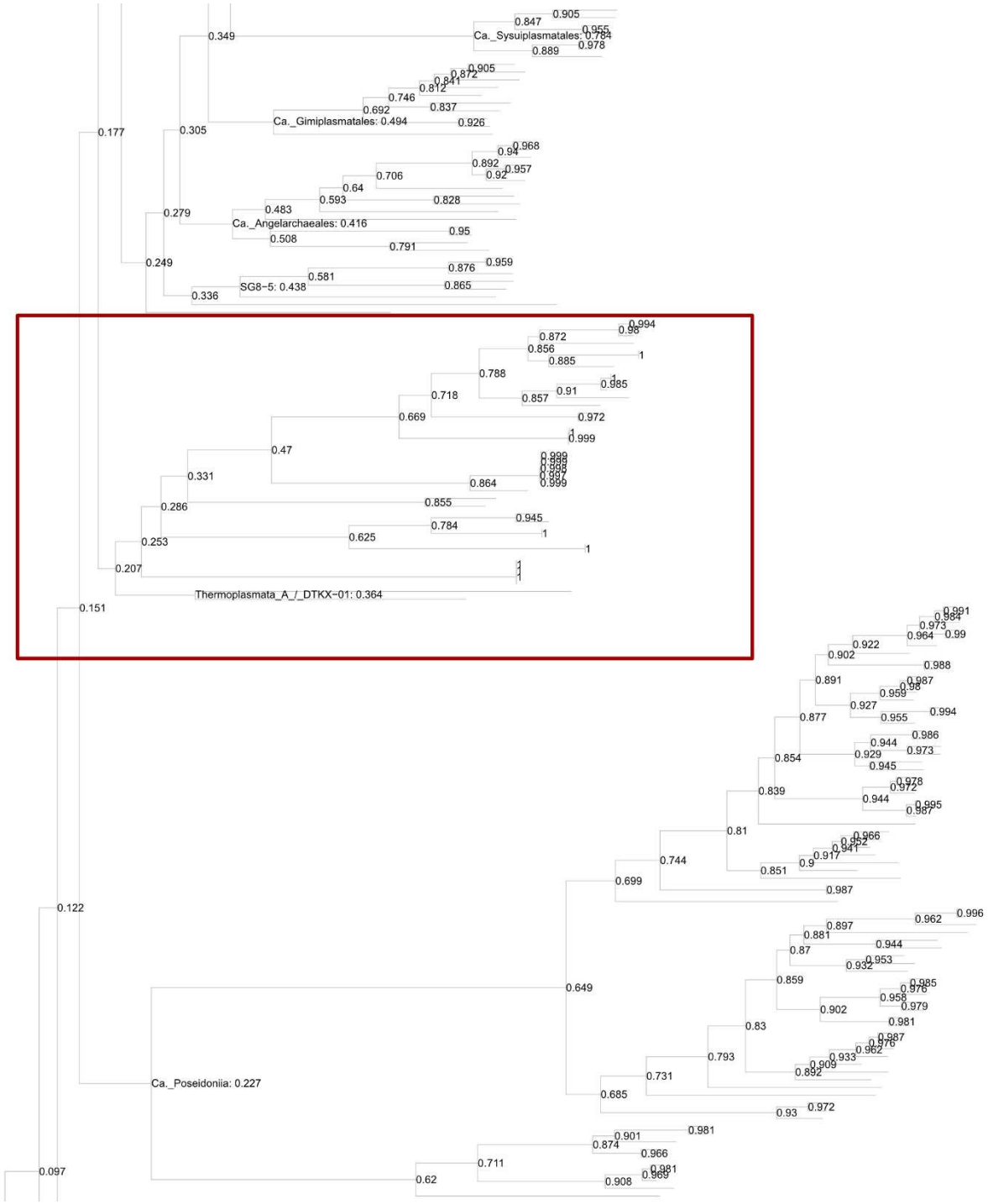

Figure S4 Partial phylogenomic tree of the Thermoplasmatota including RED. Maximum-likelihood tree (RAxML, 100 bootstraps) of 370 Thermoplasmatota MAGs and 35 EX4484-6 MAGs obtained through data mining of genome assemblies and metagenomic short read data sets. Node labels display taxonomic affiliation on class level. The class EX4484-6 clusters with Thermoplasmata_A/DTKX_01 according to the marker gene tree and is indicated by a red box (Fig. 3a). Node values indicate relative evolutionary divergence (RED) (whole tree provided as separate PDF)

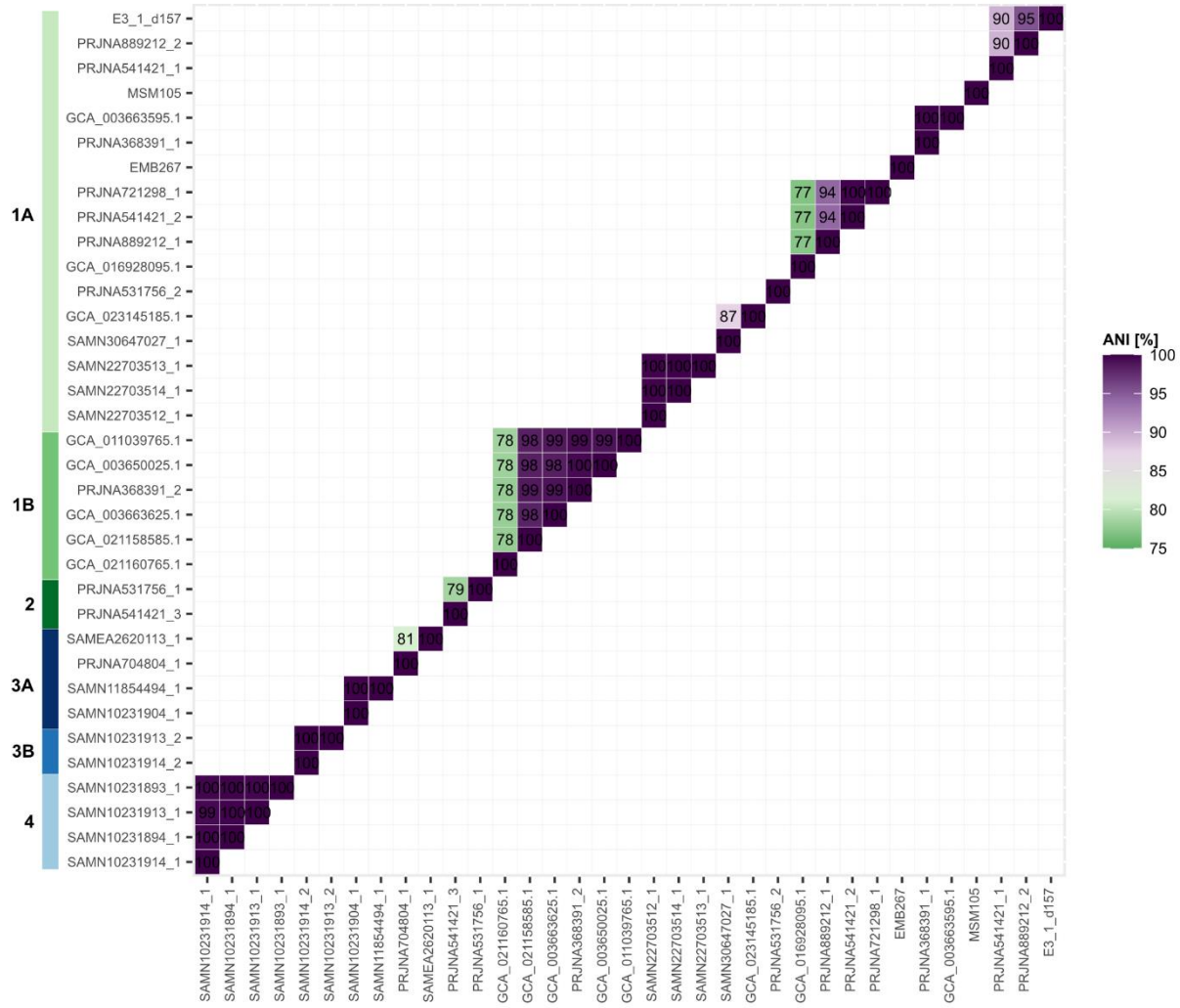

Figure S5 Average nucleotide identity (ANI). Heatmap of ANI between all 35 EX4484-6 MAGs obtained during our data mining. Numbers in colored tiles indicate the ANI percentage between the compared MAGs (threshold ANI > 75%). MAGs were ordered according to their taxonomy in the marker gene tree (Figure 3).

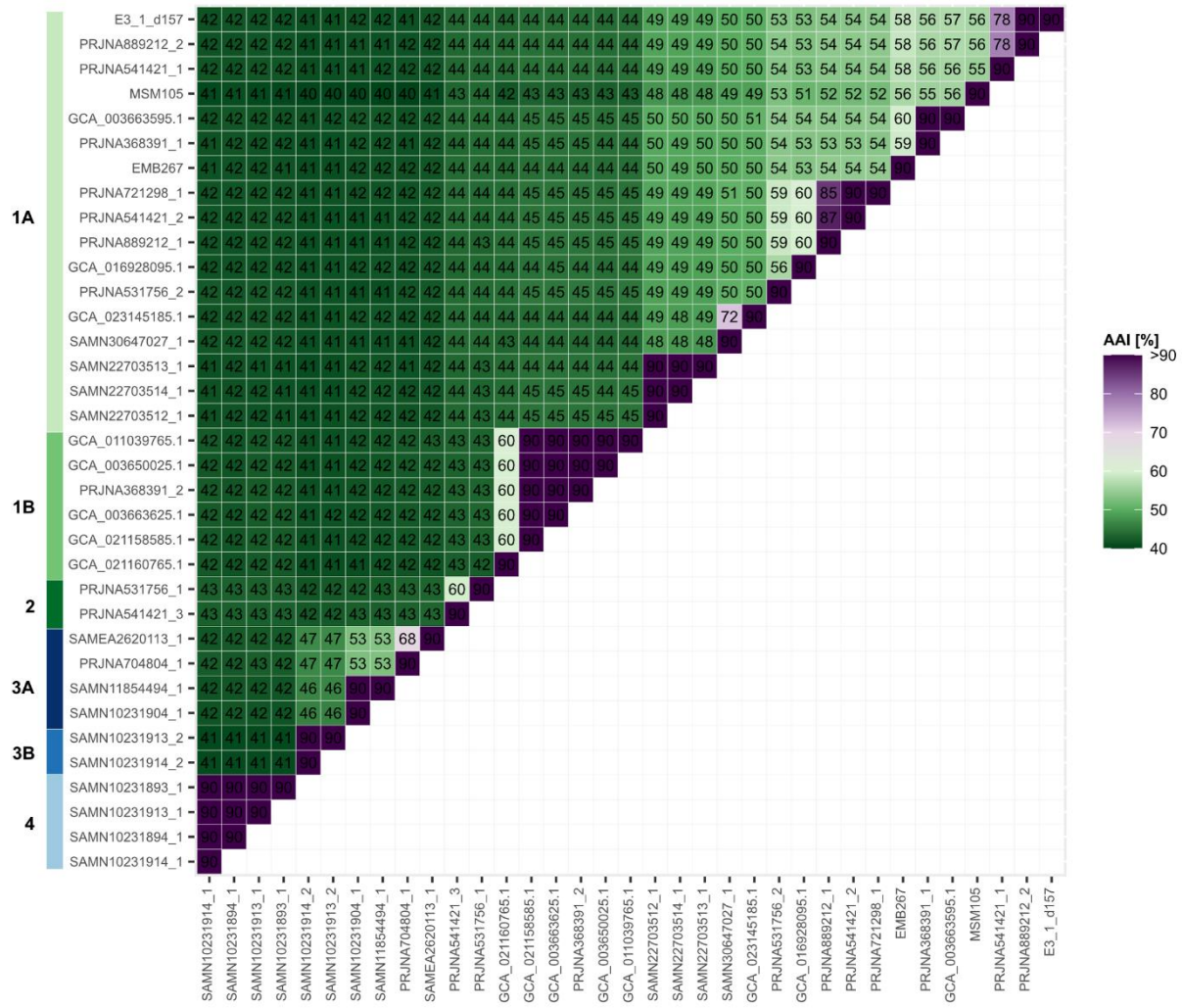

Figure S6 Amino acid identity (AAI). Heatmap of AAI between all 35 EX4484-6 MAGs obtained during our data mining. Numbers in colored tiles indicate the AAI percentage between the compared MAGs. MAGs were ordered according to their taxonomy in the marker gene tree (Figure 3).

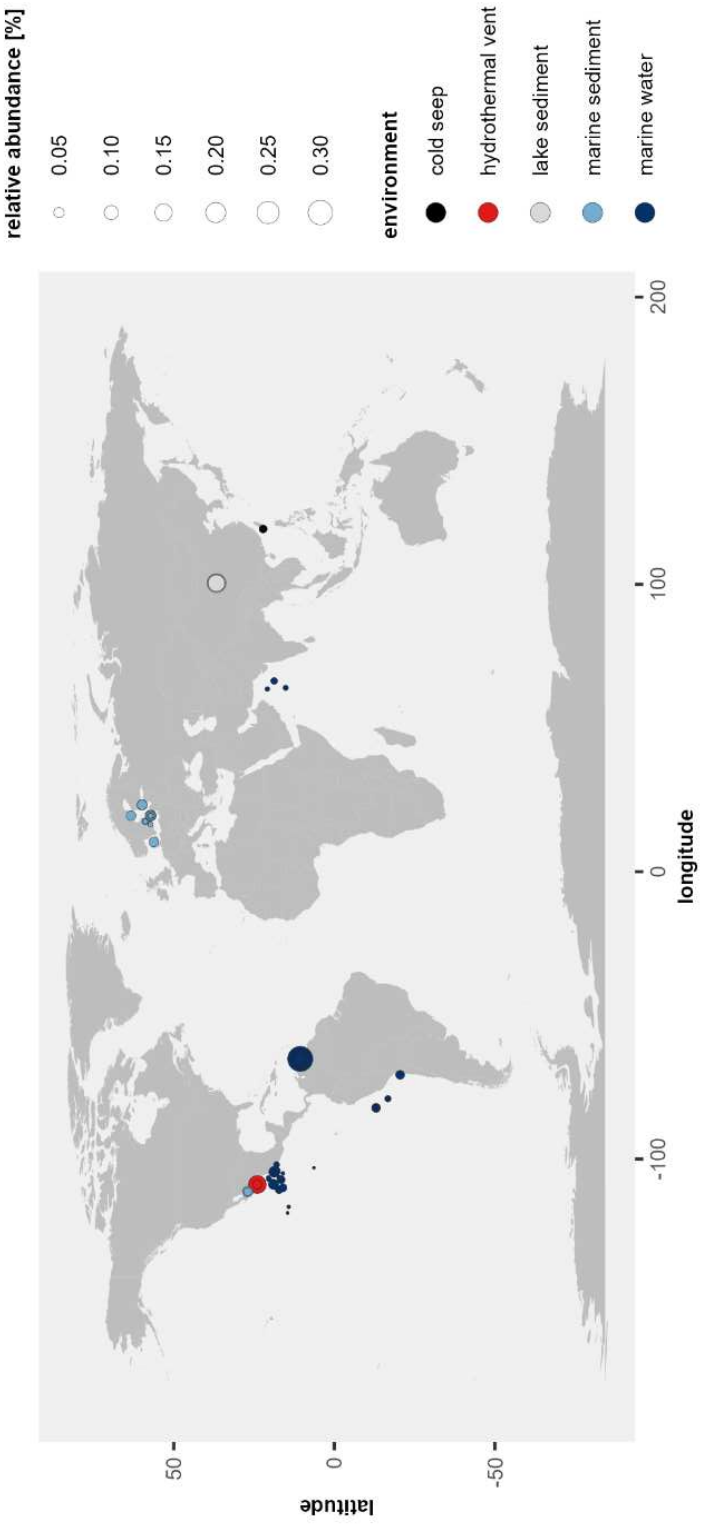

Figure S7 Distribution of the class EX4484-6. World map showing the distribution, relative abundance and environment of all observed EX4484-6 detected in 128 samples. Relative abundances were calculated by aligning quality trimmed reads of 8573 metagenomic sequencing runs to a competitive mapping index containing 20 non-redundant EX4484-6 MAGs. Point colors indicate the environment, point sizes indicate relative abundance.

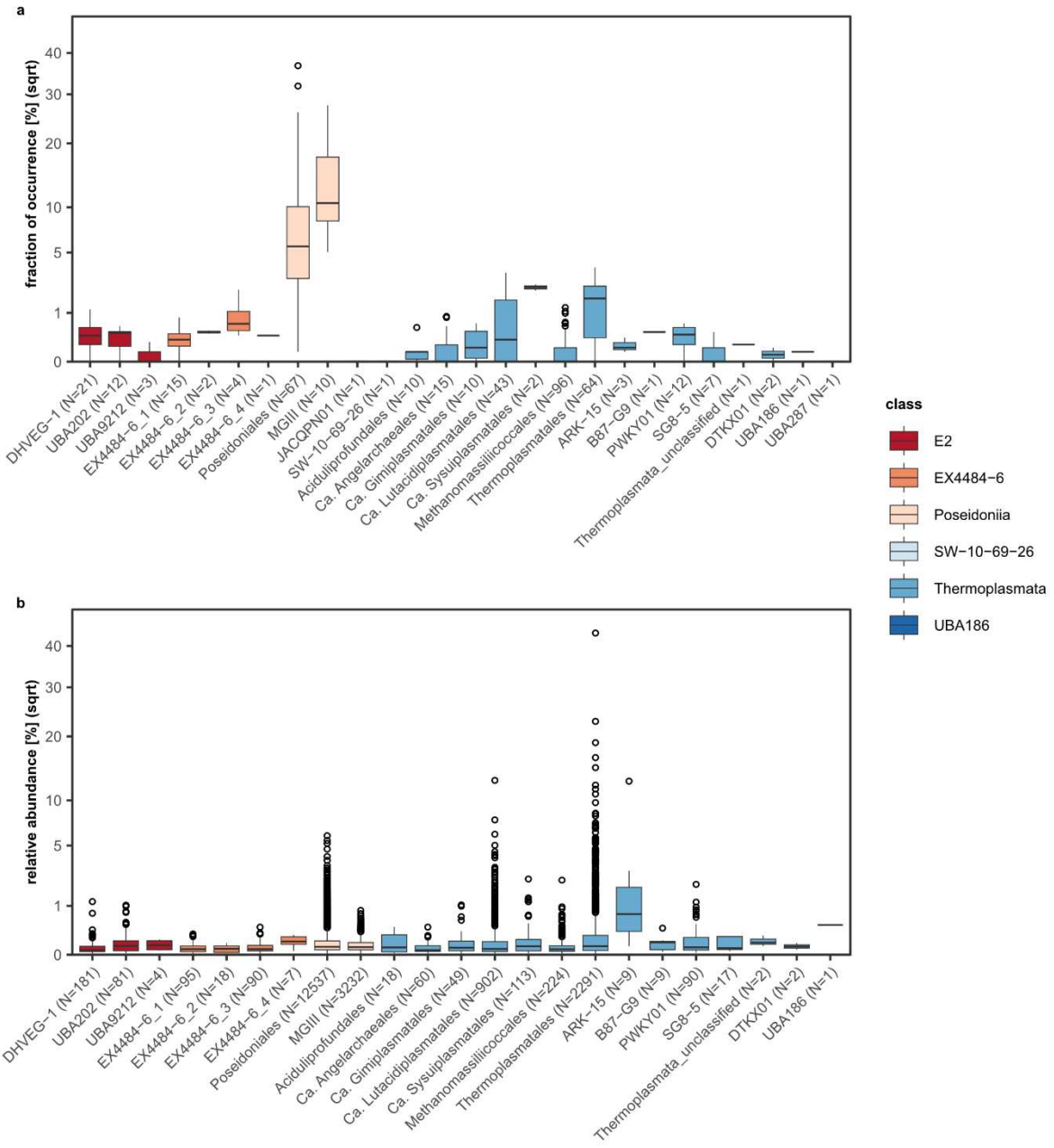

Figure S8 Abundance of non-redundant Thermoplasmatota in the environment. (a) As fraction of occurrence for each order found within the non-redundant Thermoplasmatota data set. The fraction of occurrence was defined as fraction of data sets MAGs occurred in across all screened data sets. Number of observations N represents the number of Thermoplasmatota genomes per order. (**b)** As relative abundance of each order in samples they occurred in. Number of observations N represents the number of metagenomic runs, in which the genome was detected.

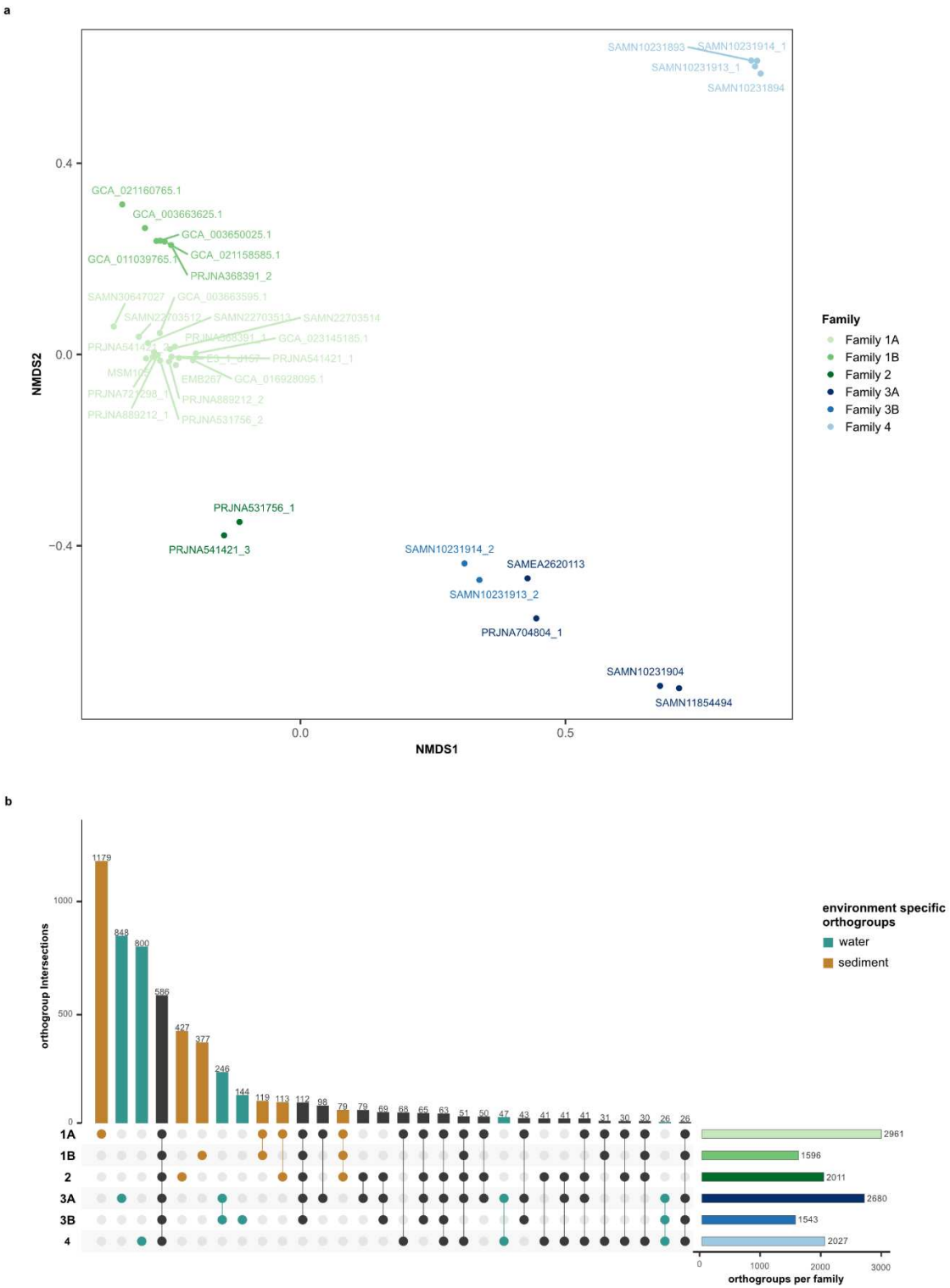

Figure S9 Clustering of EX4484-6 MAGs based on orthogroups. (a) NMDS based on all orthogroups found in all 35 EX4484-6. The NMDS was computed using the function metaMDS from the package vegan v2.6.4 with Jaccard dissimilarities and two dimensions. Single families are indicated by different colors. (**b)** Upset plot showing intersections of orthogroups for all six families. Intersections shown reflect 95% of all orthogroups within the data set. Numbers of orthogroups per intersection are indicated above the bars. Orthogroup intersections, which are only present in sediment or water are indicated by color. Total number of different orthogroups per family are indicated by horizontal bars next to the upset plot.

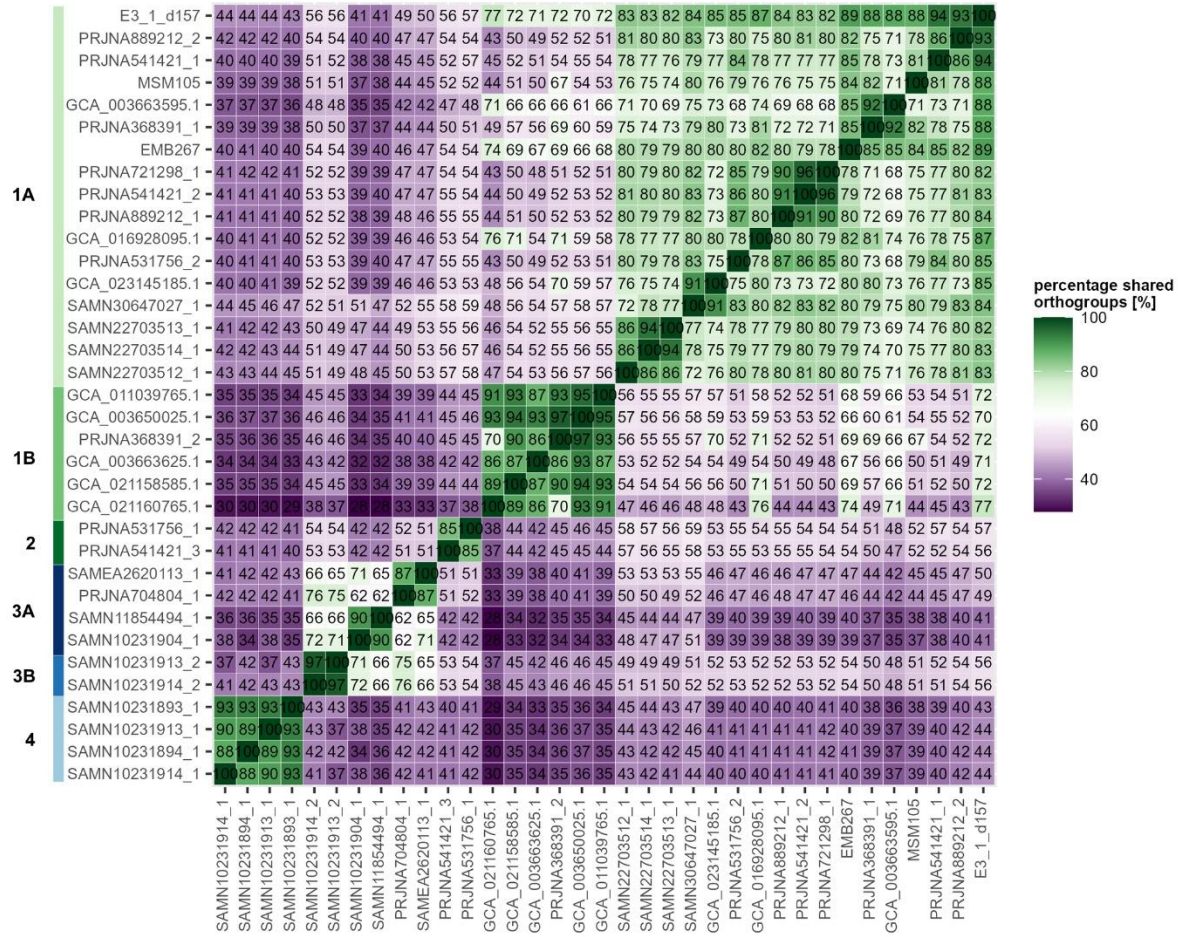

Figure S10 Shared orthogroups between EX4484-6 MAGs. Heatmap of shared orthogroups between all 35 EX4484-6 MAGs. Numbers in colored tiles in the heatmap indicate the percentage of shared orthogroups between the compared MAGs. MAGs were ordered according to their taxonomy in the marker gene tree (Figure 3).

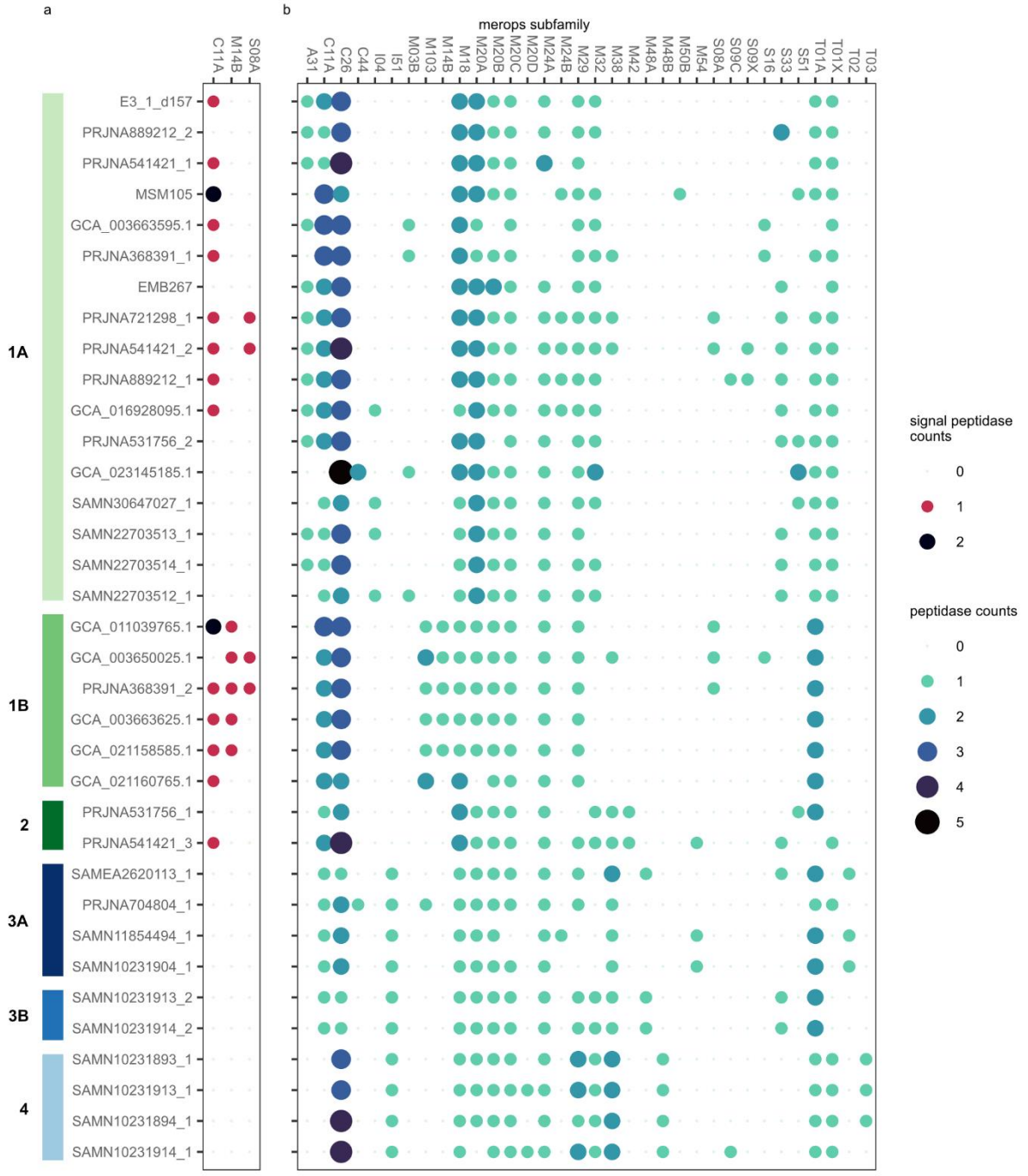

Figure S11 Peptidases in EX4484-6 MAGs. (a) Extracellular peptidase homologs within MAGs of the class EX4484-6. (**b)** Peptidase homologs within MAGs of the class EX4484-6. MAGs were ordered according to their taxonomy in the marker gene tree (Figure 3).

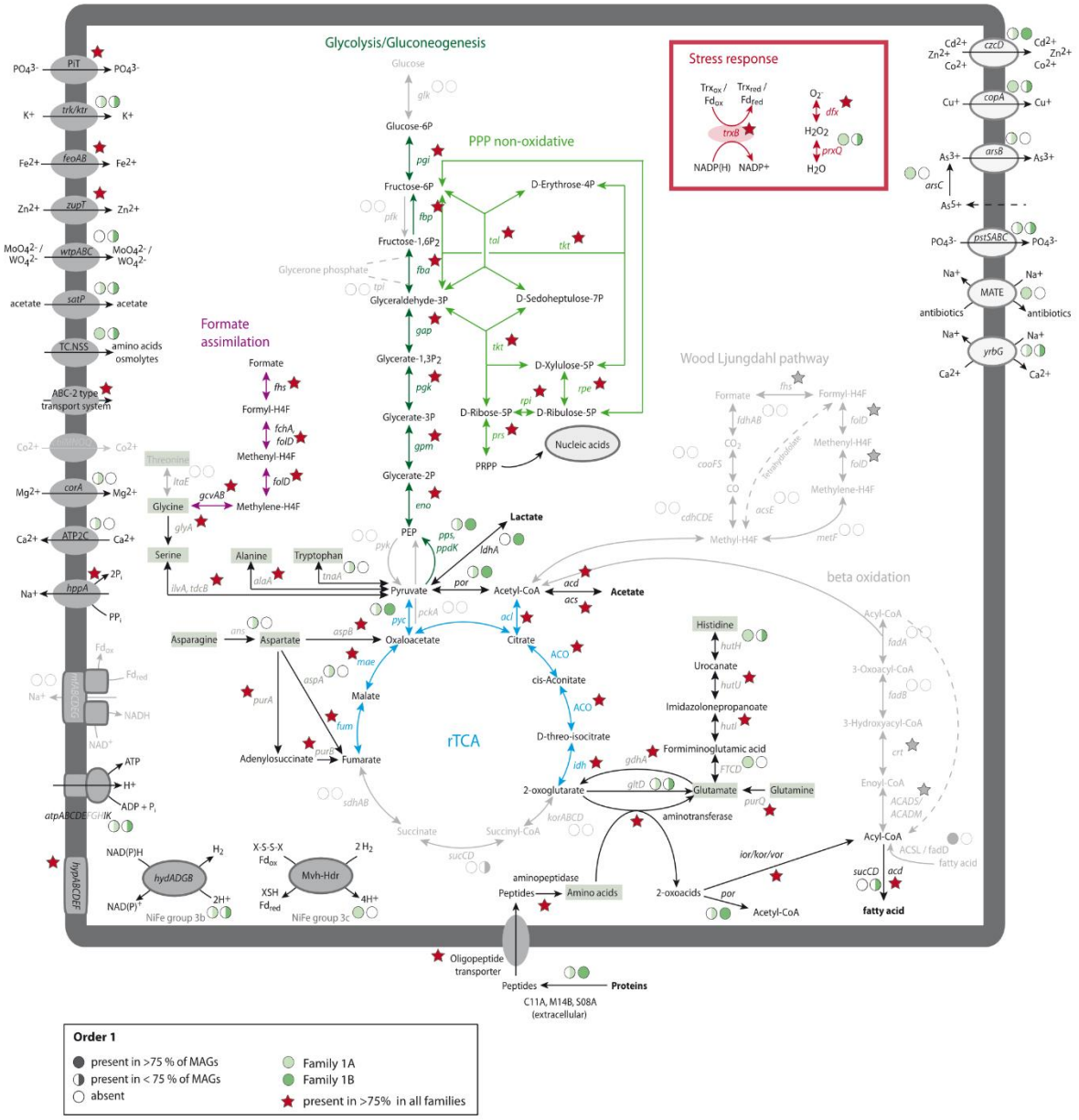

Figure S12 Metabolic reconstruction of order 1 (Family 1A, Family 1B). Pathways in grey represent pathways with missing genes that are therefore not functional. Gene abbreviations can be found in Supplementary Table 9. The presence of genes is indicated by full or half circles for each family. Red stars indicate the presence of genes in all families.

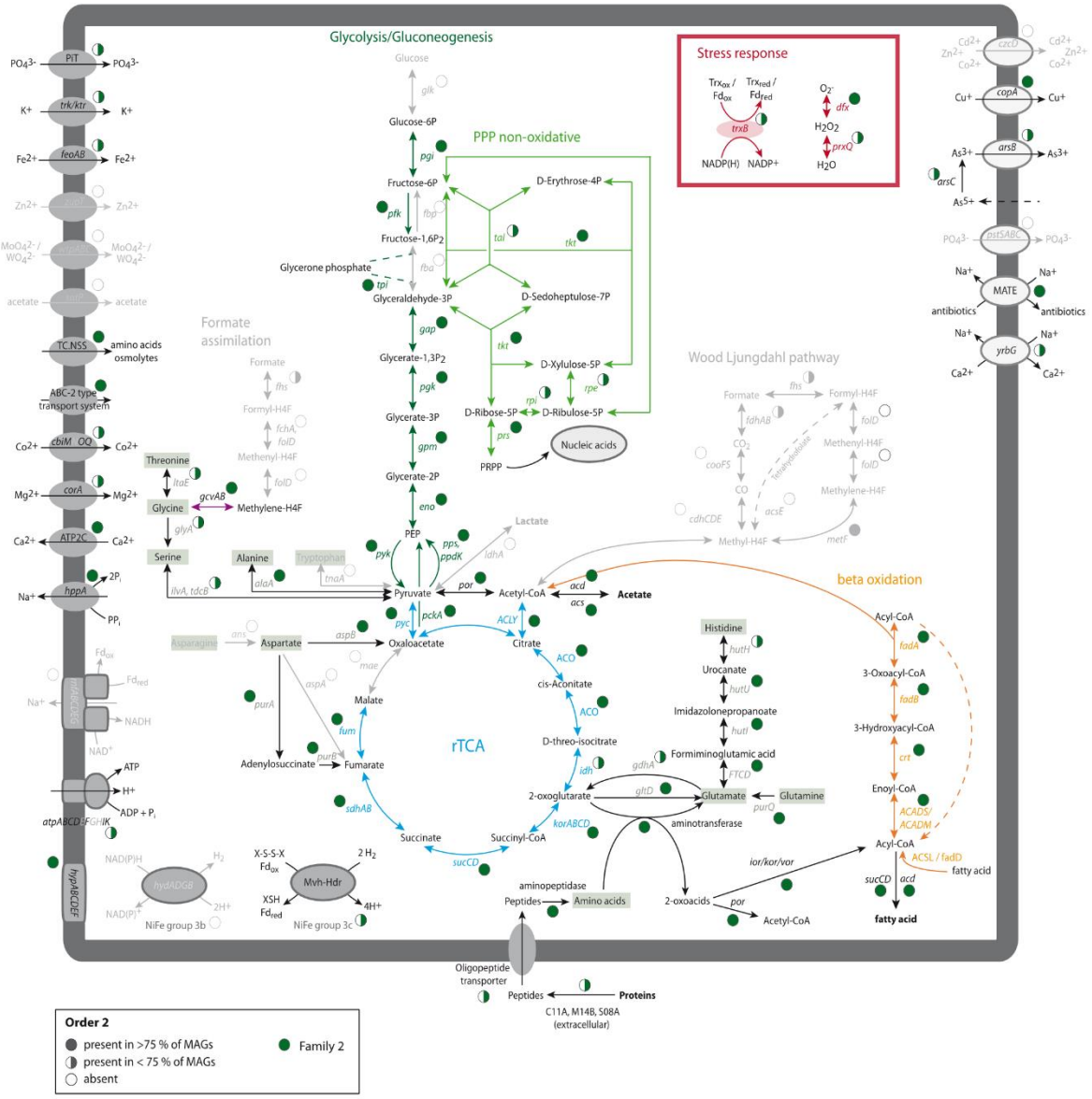

Figure S13 Metabolic reconstruction of order 2 (Family 2). Pathways in grey represent pathways with missing genes that are therefore not functional. Gene abbreviations can be found in Supplementary Table 9. The presence of genes is indicated by full or half circles.

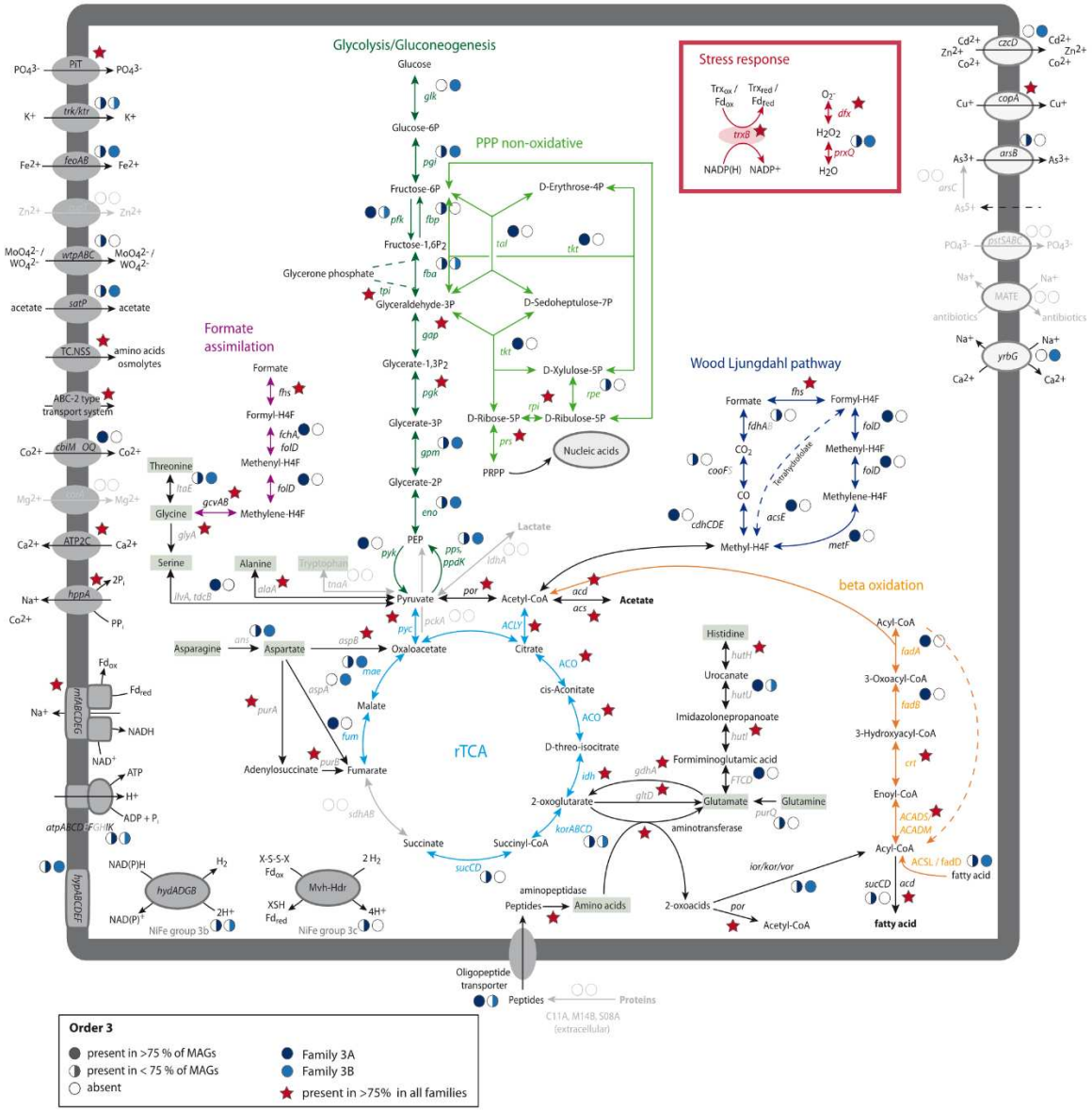

Figure S14 Metabolic reconstruction of order 3 (Family 3A, Family 3B). Pathways in grey represent pathways with missing genes that are therefore not functional. Gene abbreviations can be found in Supplementary Table 9. The presence of genes is indicated by full or half circles for each family. Red stars indicate the presence of genes in all families.

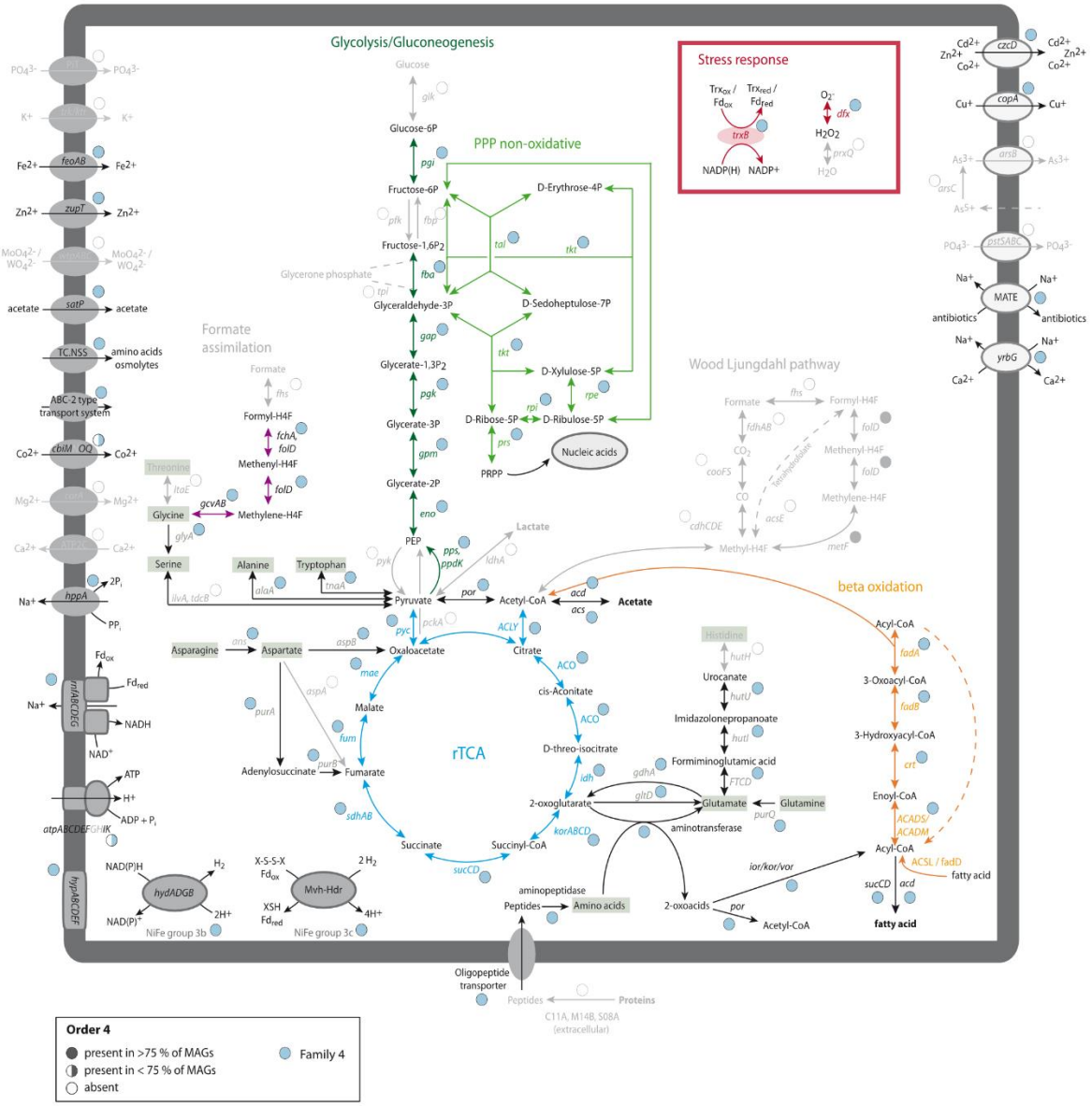

Figure S15 Metabolic reconstruction of order 4 (Family 4). Pathways in grey represent pathways with missing genes and are therefore not functional. Gene abbreviations can be found in Supplementary Table 9. The presence of genes is indicated by full or half circles.

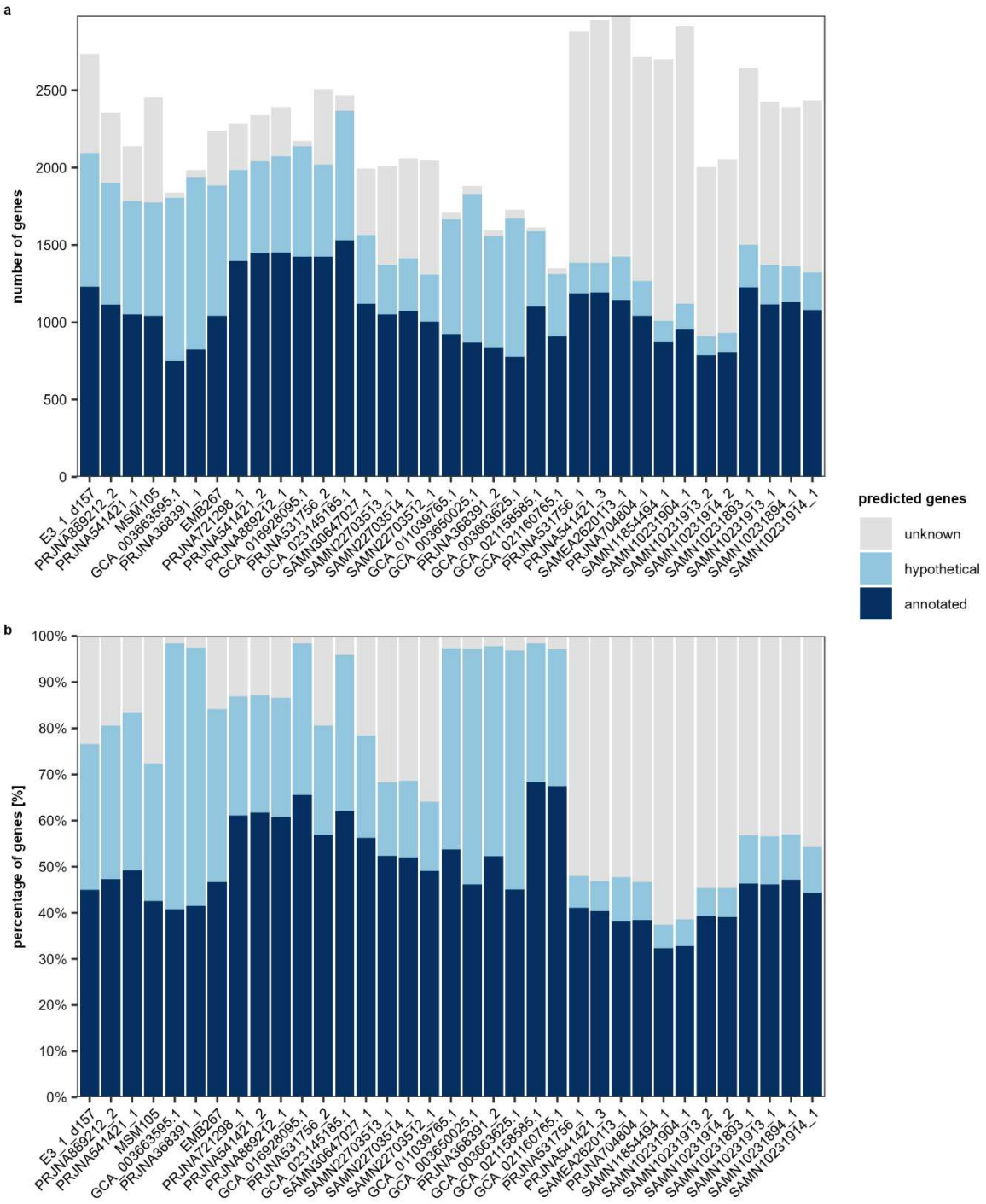

Figure S16 Annotation status of predicted genes in EX4484-6 MAGs. Genes classified as annotated, hypothetical and unknown based on NR and KEGG annotations shown as (**a**) number of genes and (**b**) percentage of genes.

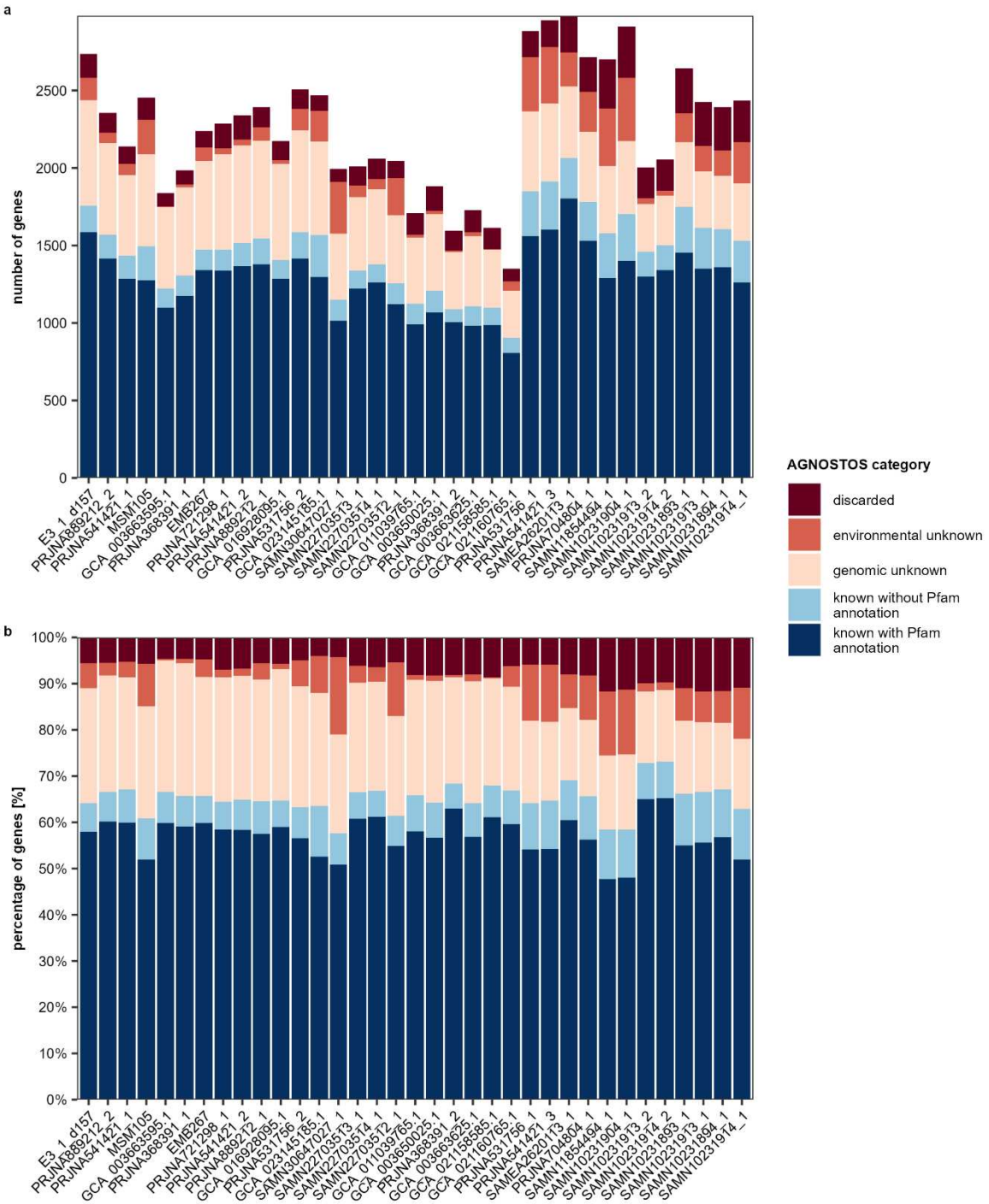

Figure S17 AGNOSTOS annotation category of genes in EX4484-6 MAGs. Gene classification based on AGNOSTOS. Gene classifications are shown as (**a**) number of genes and (**b**) percentage of genes.

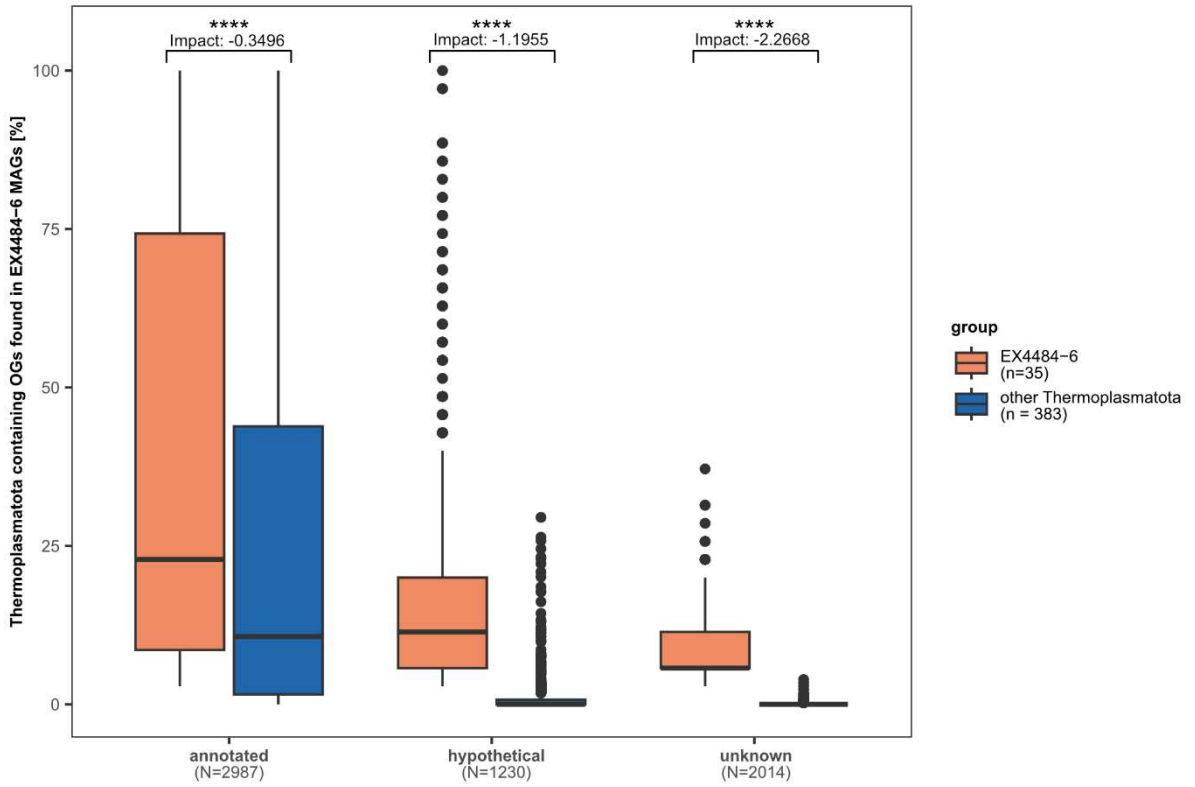

Figure S18 Shared orthogroups among Thermoplasmatota. Boxplot of Thermoplasmatota genomes sharing orthogroups (OGs) present within EX4484-6 MAGs separated by annotation category: annotated, hypothetical and unknown, according to their NR and KEGG annotation. Differences between groups were tested by Wilcoxon Signed Rank test, *p*-values are indicated by asterisks (**** p <= 0.0001). The measure of impact describes the effect size. Number of observations N indicates the number of orthogroups in each category. Number of genomes per group is indicated by number n.

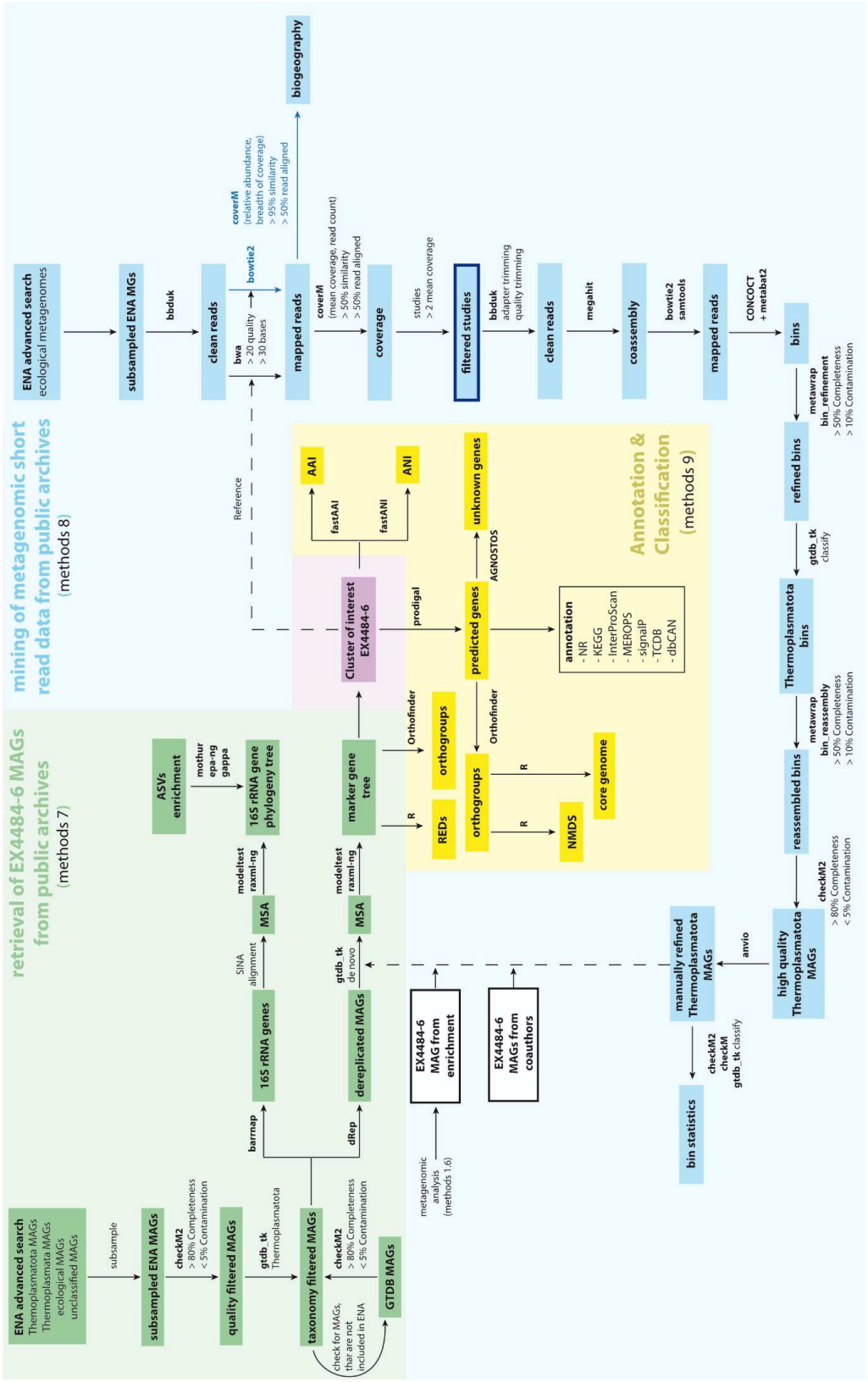

Figure S19 Flowchart of methods applied during the data mining conducted in this study. Retrieval of EX4484-6 MAGs from public archives (methods section 7) is displayed in green, the subsequent mining of metagenomic short read data from public archives is displayed in blue (methods section 8). Annotation and classification (methods section 9) was conducted on both, the MAGs retrieved from public archives and MAGs reconstructed from metagenomic short read data (yellow). Programs and settings used are indicated next to or below arrows.

3.2.4 Supplementary References

1. Grosskopf R, Janssen PH, Liesack W. Diversity and structure of the methanogenic community in anoxic rice paddy soil microcosms as examined by cultivation and direct 16S rRNA gene sequence retrieval. Appl. Environ. Microbiol. 1998; 64(3):960-969.

2. Lueders T, Friedrich MW. Effects of amendment with ferrihydrite and gypsum on the structure and activity of methanogenic populations in rice field soil. Appl. Environ. Microbiol. 2002; 68(5):2484-2494.

3. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012; 28(14):1823-1829.

4. 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(4):1363-1371.

5. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 2013; 42(D1):D643-D648.

6. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013; 41(Database issue):D590-596.

7. Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, Knight R. PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics. 2011; 27(8):1159-1161.

8. Yin X, Cai M, Liu Y, Zhou G, Richter-Heitmann T, Aromokeye DA et al. Subgroup level differences of physiological activities in marine Lokiarchaeota. ISME J. 2020; 15(3):848-861.

9. Seemann T. barrnap 0.9 : rapid ribosomal RNA prediction. https://github.com/tseemann/barrnap 2018.

10. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol. Biol. Evol. 2019; 37(1):291-294.

11. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019; 35(21):4453-4455.

12. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 2009; 75(23):7537-7541.

13. Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T et al. EPA-ng: massively parallel evolutionary placement of genetic sequences. Syst. Biol. 2018; 68(2):365-369.

14. Czech L, Barbera P, Stamatakis A. Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. Bioinformatics. 2020; 36(10):3263-3265.

15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J. Mol .Biol. 1990; 215(3):403-410.

16. Paoli L, Ruscheweyh H-J, Forneris CC, Hubrich F, Kautsar S, Bhushan A et al. Biosynthetic potential of the global ocean microbiome. Nature. 2022; 607(7917):111-118.

17. Bushnell B. BBMap: A fast, accurate, splice-aware aligner. Lawrence Berkeley National Laboratory. 2014; LBNL Report #: LBNL-7065E.

18. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr. Protoc. Bioinformatics. 2020; 70(1):e102.

19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754-1760.

20. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019; 7:e7359.

21. 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(7):1043-1055.

22. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS et al. Community-led, integrated, reproducible multi-omics with anvi'o. Nat. Microbiol. 2021; 6(1):3-6.

23. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the Genome Taxonomy Database. PREPRINT bioRxiv. 2022.

24. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 2018; 36(10):996-1004.

25. Ocean Microbiomics Database. https://microbiomics.io/ocean2 (2024). Accessed 05 Sept 2024.

26. Rudnick G, Krämer R, Blakely RD, Murphy DL, Verrey F. The SLC6 transporters: perspectives on structure, functions, regulation, and models for transporter dysfunction. Pflug Arch. Eur. J. Phy. 2014; 466(1):25-42.

27. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 2009; 37(Database issue):D233-238.

28. López-Mondéjar R, Tláskal V, da Rocha UN, Baldrian P. Global distribution of carbohydrate utilization potential in the prokaryotic tree of life. mSystems. 2022; 7(6):e0082922.

29. Mall A, Sobotta J, Huber C, Tschirner C, Kowarschik S, Bačnik K et al. Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. Science. 2018; 359(6375):563-567.

30. Garritano AN, Song W, Thomas T. Carbon fixation pathways across the bacterial and archaeal tree of life. PNAS Nexus. 2022; 1(5).

31. Yin X, Zhou G, Cai M, Zhu Q-Z, Richter-Heitmann T, Aromokeye DA et al. Catabolic protein degradation in marine sediments confined to distinct archaea. ISME J. 2022.

32. Baltscheffsky M, Schultz A, Baltscheffsky H. H+-PPases: a tightly membrane-bound family. FEBS Lett. 1999; 457(3):527-533.

33. Belogurov GA, Malinen AM, Turkina MV, Jalonen U, Rytkönen K, Baykov AA et al. Membrane-bound pyrophosphatase of *Thermotoga maritima* requires sodium for activity. Biochem. 2005; 44(6):2088-2096.

34. Malinen AM, Belogurov GA, Baykov AA, Lahti R. Na+-Pyrophosphatase: A novel primary sodium pump. Biochem. 2007; 46(30):8872-8878.

35. Kuhns M, Trifunović D, Huber H, Müller V. The Rnf complex is a Na+ coupled respiratory enzyme in a fermenting bacterium, *Thermotoga maritima*. Commun. Biol. 2020; 3(1):431.

36. Watanabe S, Sasaki D, Tominaga T, Miki K. Structural basis of [NiFe] hydrogenase maturation by Hyp proteins. Biol. Chem. 2012; 393(10):1089-1100.

37. Ma K, Weiss R, Adams MWW. Characterization of hydrogenase II from the hyperthermophilic archaeon *Pyrococcus furiosus* and assessment of its role in sulfur reduction. J. Bacteriol. 2000; 182(7):1864-1871.

38. Ma K, Schicho RN, Kelly RM, Adams MW. Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. Proc. Natl. Acad. Sci. U. S. A. 1993; 90(11):5341-5344.

39. Hedderich R, Berkessel A, Thauer RK. Purification and properties of heterodisulfide reductase from *Methanobacterium thermoautotrophicum* (strain Marburg). Eur. J. Biochem. 1990; 193(1):255-261.

40. Setzke E, Hedderich R, Heiden S, Thauer RK. H2: heterodisulfide oxidoreductase complex from *Methanobacterium thermoautotrophicum*. Composition and properties. Eur. J. Biochem. 1994; 220(1):139-148.

41. Thauer RK, Kaster AK, Seedorf H, Buckel W, Hedderich R. Methanogenic archaea: ecologically relevant differences in energy conservation. Nat. Rev. Microbiol. 2008; 6(8):579-591.

42. Kaster AK, Moll J, Parey K, Thauer RK. Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. Proc. Natl. Acad. Sci. U. S. A. 2011; 108(7):2981-2986.

43. Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y et al. Isolation of an archaeon at the prokaryote-eukaryote interface. Nature. 2020; 577(7791):519-525.

44. Qu Y-N, Rao Y-Z, Qi Y-L, Li Y-X, Li A, Palmer M et al. *Panguiarchaeum symbiosum*, a potential hyperthermophilic symbiont in the TACK superphylum. Cell Rep. 2023; 42(3):112158.

45. Sapra R, Verhagen MFJM, Adams MWW. Purification and characterization of a membrane-bound hydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. 2000; 182(12):3423-3428.

46. Lau CK, Ishida H, Liu Z, Vogel HJ. Solution structure of *Escherichia coli* FeoA and its potential role in bacterial ferrous iron transport. J. Bacteriol. 2013; 195(1):46-55.

47. Schulz H, Zabel M (eds.). Marine geochemistry, 2nd ed. edn. Berlin Heidelberg, Germany: Springer-Verlag Berlin Heidelberg; 2006.

48. Silburn B, Kröger S, Parker ER, Sivyer DB, Hicks N, Powell CF et al. Benthic pH gradients across a range of shelf sea sediment types linked to sediment characteristics and seasonal variability. Biogeochemistry. 2017; 135(1):69-88.

49. Von Damm KL, Edmond JM, Measures CI, Grant B. Chemistry of submarine hydrothermal solutions at Guaymas Basin, Gulf of California. Geochim. Cosmochim. Acta. 1985; 49(11):2221-2237.

50. Ulfsbo A, Hulth S, Anderson LG. pH and biogeochemical processes in the Gotland Basin of the Baltic Sea. Mar. Chem. 2011; 127(1):20-30.

51. Kammler M, Schön C, Hantke K. Characterization of the ferrous iron uptake system of *Escherichia coli*. J Bacteriol. 1993; 175(19):6212-6219.

52. Andreini C, Banci L, Bertini I, Rosato A. Zinc through the three domains of life. J. Proteome Res. 2006; 5(11):3173-3178.

53. Grass G, Franke S, Taudte N, Nies DH, Kucharski LM, Maguire ME et al. The metal permease ZupT from *Escherichia coli* is a transporter with a broad substrate spectrum. J. Bacteriol. 2005; 187(5):1604-1611.

54. Pyle AM. Role of metal ions in ribozymes. Met. Ions. Biol. Syst. 1996; 32:479-520.

55. Wolf FI, Cittadini A. Chemistry and biochemistry of magnesium. Mol. Aspects Med. 2003; 24(1-3):3-9.

56. Smith RL, Banks JL, Snavely MD, Maguire ME. Sequence and topology of the CorA magnesium transport systems of *Salmonella typhimurium* and *Escherichia coli*. Identification of a new class of transport protein. J. Biol. Chem. 1993; 268(19):14071-14080.

57. Smith RL, Maguire ME. Microbial magnesium transport: unusual transporters searching for identity. Mol. Microbiol. 1998; 28(2):217-226.

58. Boer JL, Mulrooney SB, Hausinger RP. Nickel-dependent metalloenzymes. Arch. Biochem. Biophys. 2014; 544:142-152.

59. Kobayashi M, Shimizu S. Cobalt proteins. Eur. J. Biochem. 1999; 261(1):1-9.

60. Nakata A, Amemura M, Shinagawa H. Regulation of the phosphate regulon in *Escherichia coli* K-12: regulation of the negative regulatory gene phoU and identification of the gene product. J. Bacteriol. 1984; 159(3):979-985.

61. Gardner SG, Johns KD, Tanner R, McCleary WR. The PhoU Protein from *Escherichia coli* interacts with PhoR, PstB, and metals to form a phosphate-signaling complex at the membrane. J. Bacteriol. 2014; 196(9):1741-1752.

62. Muda M, Rao NN, Torriani A. Role of PhoU in phosphate transport and alkaline phosphatase regulation. J. Bacteriol. 1992; 174(24):8057-8064.

63. Stautz J, Hellmich Y, Fuss MF, Silberberg JM, Devlin JR, Stockbridge RB et al. Molecular mechanisms for bacterial potassium homeostasis. J. Mol. Biol. 2021; 433(16):166968.

64. Beagle SD, Lockless SW. Unappreciated roles for K(+) channels in bacterial physiology. Trends Microbiol. 2021; 29(10):942-950.

65. Cao Y, Pan Y, Huang H, Jin X, Levin EJ, Kloss B et al. Gating of the TrkH ion channel by its associated RCK protein TrkA. Nature. 2013; 496(7445):317-322.

66. Bevers LE, Hagedoorn PL, Krijger GC, Hagen WR. Tungsten transport protein A (WtpA) in *Pyrococcus furiosus*: the first member of a new class of tungstate and molybdate transporters. J. Bacteriol. 2006; 188(18):6498-6505.

67. Makdessi K, Andreesen JR, Pich A. Tungstate Uptake by a highly specific ABC transporter in *Eubacterium acidaminophilum*. J. Biol. Chem. 2001; 276(27):24557-24564.

68. Hille R. Molybdenum and tungsten in biology. Trends Biochem. Sci. 2002; 27(7):360-367.

69. Kletzin A, Adams MW. Tungsten in biological systems. FEMS Microbiol. Rev. 1996; 18(1):5-63.

70. Møller JV, Juul B, le Maire M. Structural organization, ion transport, and energy transduction of P-type ATPases. Biochim. Biophys. Acta Rev. Biomembr. 1996; 1286(1):1-51.

71. Sá-Pessoa J, Paiva S, Ribas D, Silva IJ, Viegas SC, Arraiano CM et al. SATP (YaaH), a succinate-acetate transporter protein in *Escherichia coli*. Biochem. J. 2013; 454(3):585-595.

72. Bernal V, Castaño-Cerezo S, Cánovas M. Acetate metabolism regulation in *Escherichia coli*: carbon overflow, pathogenicity, and beyond. Appl. Microbiol. Biotechnol. 2016; 100(21):8985-9001.

73. Riebe O, Fischer RJ, Bahl H. Desulfoferrodoxin of *Clostridium acetobutylicum* functions as a superoxide reductase. FEBS Lett. 2007; 581(29):5605-5610.

74. Perkins A, Nelson KJ, Parsonage D, Poole LB, Karplus PA. Peroxiredoxins: guardians against oxidative stress and modulators of peroxide signaling. Trends Biochem. Sci. 2015; 40(8):435-445.

75. Thompson AD, Bernard SM, Skiniotis G, Gestwicki JE. Visualization and functional analysis of the oligomeric states of *Escherichia coli* heat shock protein 70 (Hsp70/DnaK). Cell Stress Chaperones. 2012; 17(3):313-327.

76. Delaney JM. Requirement of the *Escherichia coli* dnaK gene for thermotolerance and protection against H_2O_2. Microbiol. 1990; 136(10):2113-2118.

77. Besserer GM, Nicoll DA, Abramson J, Philipson KD. Characterization and purification of a Na+/Ca2+ exchanger from an archaebacterium. J. Biol. Chem. 2012; 287(11):8652-8659.

78. Fan B, Rosen BP. Biochemical characterization of CopA, the *Escherichia coli* Cu(I)-translocating P-type ATPase*. J. Biol. Chem. 2002; 277(49):46987-46992.

79. Rensing C, Fan B, Sharma R, Mitra B, Rosen BP. CopA: An *Escherichia coli* Cu(I)-translocating P-type ATPase. Proc. Natl. Acad. Sci. U. S. A. 2000; 97(2):652-656.

80. Peña MM, Lee J, Thiele DJ. A delicate balance: homeostatic control of copper uptake and distribution. J. Nutr. 1999; 129(7):1251-1260.

81. Linder MC. Introduction and overview of copper as an element essential for life. In: Biochemistry of Copper. Boston, MA: Springer US; 1991. p. 1-13.

82. Anton A, Grosse C, Reissmann J, Pribyl T, Nies DH. CzcD is a heavy metal ion transporter involved in regulation of heavy metal resistance in *Ralstonia* sp. strain CH34. J. Bacteriol. 1999; 181(22):6876-6881.

83. Rosen BP, Liu Z. Transport pathways for arsenic and selenium: a minireview. Environ. Int. 2009; 35(3):512-515.

84. Cullen WR, Reimer KJ. Arsenic speciation in the environment. Chem. Rev. 1989; 89(4):713-764.

85. Kalia K, Khambholja DB. 28 - Arsenic contents and its biotransformation in the marine environment. In: Flora SJS, editors. Handbook of Arsenic Toxicology. Oxford: Academic Press; 2015. p. 675-700.

86. Knobloch T. Chemical munitions dumped in the Baltic Sea: Report of the ad hoc expert group to update and review the existing information on dumped chemical munitions in the Baltic Sea (HELCOM MUNI): Helsinki Commission, Baltic Marine Environment Protection Commission; 2014.

87. Missiaen T, Söderström M, Popescu I, Vanninen P. Evaluation of a chemical munition dumpsite in the Baltic Sea based on geophysical and chemical investigations. Sci. Total Environ. 2010; 408(17):3536-3553.

88. Szubska M, Bełdowski J. Spatial distribution of arsenic in surface sediments of the southern Baltic Sea. Oceanol. 2023; 65(2):423-433.

89. Krah A, Huber RG, Zachariae U, Bond PJ. On the ion coupling mechanism of the MATE transporter ClbM. Biochim. Biophys. Acta Biomembr. 2020; 1862(2):183137.

90. Bouki C, Venieri D, Diamadopoulos E. Detection and fate of antibiotic resistant bacteria in wastewater treatment plants: a review. Ecotoxicol. Environ. Saf. 2013; 91:1-9.

91. Schijven JF, Blaak H, Schets FM, de Roda Husman AM. Fate of extended-spectrum β-lactamase-producing *Escherichia coli* from faecal sources in surface water and probability of human exposure through swimming. Environ. Sci. Technol. 2015; 49(19):11825-11833.

Chapter IV

Diverse methoxydotrophic acetogens facilitate terrestrial carbon in anoxic marine sediments

Mara D. Maeke, Xiuran Yin, Charlotte Recke, Lea Berger, Lea C. Wunder, Christiane Hassenrück, Jan Tebben, Tilmann Harder, Jenny Wendt, Marcus Elvert, Tim Richter-Heitmann, Michael W. Friedrich

Manuscript in preparation

Running title:
MAC degradation by Clostridia

Contribution to the manuscript:

Experimental concept and design	70%
Acquisition of experimental data	70%
Data analysis and interpretation	90%
Preparation of figures and tables	100%
Drafting of manuscript	100%

Diverse methoxydotrophic acetogens facilitate terrestrial carbon in anoxic marine sediments

Mara D. Maeke[1], Xiuran Yin[1,2*], Charlotte Recke[1], Lea Berger[1], Lea C. Wunder[1], Christiane Hassenrück[3], Jan Tebben[4,5], Tilmann Harder[5], Jenny Wendt[6], Marcus Elvert[6], Tim Richter-Heitmann[1], Michael W. Friedrich[1,6*]

[1]Microbial Ecophysiology Group, Faculty of Biology/Chemistry, University of Bremen, Bremen, Germany

[2]State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou, China

[3]Biological Oceanography, Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Rostock, Germany

[4]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

[5]Marine Chemistry Group, Faculty of Biology/Chemistry, University of Bremen, Bremen, Germany

[6]MARUM – Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany

*Correspondence:
Prof. Dr. Michael W. Friedrich
University of Bremen, Microbial Ecophysiology Group
James-Watt-Straße 1 28359 Bremen, Germany
Phone: +49 (0) 421 218-63060
Email: michael.friedrich@uni-bremen.de

Dr. Xiuran Yin
Hainan University, State Key Laboratory of Marine Resource Utilization in South China Sea
Renmin Ave. No.58 570228 Haikou, China
Email: 996383@hainanu.edu.cn

4.1.1 Abstract

One-carbon metabolic processes, including methylotrophy, methanotrophy and methoxydotrophy, are essential to marine carbon cycling. However, microbial contributors and the ecological roles of methoxydotrophy in marine environments remain poorly understood despite their significant role in the degradation of recalcitrant compounds across diverse coastal ecosystems. Through a combination of RNA stable isotope probing (RNA-SIP), enrichment culturing, metagenomics, and data-mining, we identified more than 20 amplicon sequence variants (ASVs) capable of metabolizing MACs. These ASVs include unclassified genera in the Alkalibacteraceae (7 ASVs) and Oscillospiraceae (2 ASVs), and a novel family in the Acetivibrionales (4 ASVs) of the Clostridia class, which were not previously recognized for this function. The identified taxa harbor distinct gene clusters encoding methyltransferases acting on methoxy groups, the core enzymes for MAC metabolism, and demonstrate the ability to fix CO_2. The diversity, unique metabolic traits, and ecological roles of these methoxydotrophs involved in degrading recalcitrant compounds are widespread and critical for regulating carbon transformation in anoxic marine environments.

4.1.2 Introduction

Marine sediments constitute up to one third of terrestrial organic carbon (Burdige 2005). About 0.5 petagrams (Pg) of terrestrial carbon is transported into coastal oceans annually (Bianchi 2011). As part of terrestrial carbon lignin, the second most abundant biopolymer on earth, is deposited in marine sediments (Gargulak et al. 2015, Boerjan et al. 2003, Burdige 2005). Yet only about half of the terrestrial organic matter transported by rivers is ultimately buried in marine sediments (Bianchi 2011), indicating the remineralization of terrestrial organic matter in the water column and marine sediments. However, lignin is highly recalcitrant due to its size and structure (Zoghlami and Paës 2019). Therefore, lignin degradation was long attributed only to aerobic soil fungi and bacteria, containing an extensive repertoire of ligninolytic enzymes (Crawford and Crawford 1980, Sánchez 2009, Bugg et al. 2011, Tian et al. 2014). Until now, only a limited number of studies on microbial lignin degradation in marine sediments are available (Woo and Hazen 2018, Yu et al. 2018, Wang et al. 2021, Li et al. 2022, Ley et al. 2023, Peng et al. 2023).

Multiple of these studies identified Pseudomonadota (Gammaproteobacteria, Alphaproteobacteria), Desulfobacterota, Bacteroidota, Chloroflexota, Campylobacterota and Bathyarchaeia as the most abundant taxa in lignin-rich marine water and sediments (Woo and Hazen 2018, Li et al. 2022, Ley et al. 2023, Peng et al. 2023, Yu et al. 2023b). Genes involved

in the depolymerization of lignin were detected only in the aerobic metabolism of some of these taxa (Woo and Hazen 2018, Peng et al. 2023). Still, these depolymerizing genes were of lower abundance in marine sediments (Li et al. 2022). Instead, pathways for the anaerobic degradation of aromatic compounds derived from lignin were prevalent (Li et al. 2022, Peng et al. 2023). These observations follow the previously observed decrease of only monoaromatic compounds in anoxic sediments of the Amazon River estuary (Dittmar and Lara 2001), suggesting an interplay of different microorganisms in the complete degradation of lignin polymers. Some organisms, preferably aerobic taxa, depolymerize lignin and microorganisms in anoxic sediments utilize remaining monolignols (Peng et al. 2023, Yu et al. 2023b). Monolignols of the heteropolymer lignin contain methoxy (-OCH_3) groups (Vanholme et al. 2010). The only bacterial isolate known so far from anoxic marine sediments involved in the degradation of methoxylated aromatic compounds (MACs) is the Clostridia species *Acetobacterium carbinolicum* (Paarup et al. 2006). More recently, the archaea *Methermicoccus shengliensis*, *Archaeoglobus fulgidus*, and Bathyarchaeia were identified as participating in the degradation of MACs in marine anoxic settings (Kurth et al. 2021, Welte et al. 2021, Lin et al. 2022, Yu et al. 2023a). Yet, beyond these few strains, *in situ* studies and further evaluation of the diversity of MAC-degrading microorganisms in anoxic marine sediments are missing.

Most organisms involved in the degradation of MACs are acetogens (Khomyakova and Slobodkin 2023). Under anoxic conditions, methylotrophic acetogenic organisms demethylate MACs. Methylotrophy is defined as the growth of microorganisms on reduced one-carbon compounds or multi-carbon compounds without carbon-carbon bonds as the sole carbon source (Anthony 1982). Methylotrophic organisms acting on methoxy groups are considered methoxydotrophs (Khomyakova and Slobodkin 2023). During methoxydotrophic growth on MACs, microorganisms utilize an *O*-demethylase system consisting of three different proteins: methyltransferase I (MTI), which transfers a methyl group to a corrinoid protein, and methyltransferase II (MTII), which transfers the methyl group from the corrinoid protein to a methyl group acceptor (Kaufmann et al. 1998, Khomyakova and Slobodkin 2023). Methoxy groups transferred by the methyltransferase are accepted by tetrahydrofolate (THF) (Kaufmann et al. 1998, Engelmann et al. 2001, Drake and Daniel 2004, Studenik et al. 2012, Chen et al. 2016, Kremp et al. 2018). In order to facilitate redox balancing, four methyl groups are required, of which one is oxidized to CO_2 in the methyl branch of the reductive acetyl-CoA pathway and three combined with CO_2 to form acetyl-CoA via the carbonyl branch (Kremp et al. 2018). The oxidation of one methyl group to CO_2 is required to produce electrons needed for the CO_2 fixation in the carbonyl branch. As shown in previous studies, reduced ferredoxin

is formed by an electron bifurcating hydrogenase (*hydABC*) from hydrogen as electron donor coupled to reduction of NAD^+ (Schuchmann and Müller 2012, Kremp et al. 2018). Moreover, the Rnf complex converts NADH to reduced ferredoxin using an electrochemical Na^+ potential. The electrochemical Na^+ potential is generated by sodium dependent ATPase.

In demethoxylation, methyltransferase I acts substrate-specific, so considerable differences between MTI protein sequences were observed (Khomyakova and Slobodkin 2023). So far, six different methyltransferase systems have been characterized (Kaufmann et al. 1998, Engelmann et al. 2001, Naidu and Ragsdale 2001, Studenik et al. 2012, Chen et al. 2016, Kurth et al. 2021, Welte et al. 2021). Yet also, the methanol methyltransferase *mtaABC* from *Methanosarcina barkerii* or *Acetobacterium woodii*, catalyzing the methyl group transfer from methanol, showed structural similarities to *O*-demethylases (Harms and Thauer 1996, Kremp et al. 2018). Well-studied genomes, such as *Acetobacterium woodii* or *Desulfitobacterium hafniense*, contained up to 30 different homologs of the MTI gene (Studenik et al. 2012, Kremp et al. 2018, Lechtenfeld et al. 2018), for most of which the substrate spectra remain elusive. Multiple methoxylated aromatic compounds must be tested to evaluate the substrate spectrum of microorganisms.

In this study, we aimed to identify non-methanogenic methoxydotrophs in anoxic marine sediments. Using RNA stable isotope probing (Whiteley et al. 2007), we identified active microorganisms involved in MAC-derived methoxy group utilization. By using additional enrichments amended with different MACs, we further evaluated the substrate spectrum of identified groups. We analyzed the detected groups' metabolic potential for methoxydotrophy using metagenomic sequencing. By further screening the methyltransferase I of the *O*-demethylase system, we aimed to identify substrate-specific methyltransferase genes involved in the degradation of methoxy groups.

4.1.3 Materials and methods

4.1.3.1 Stable Isotope Probing (SIP) enrichment setup

Sediment was collected from the Helgoland mud area (54°06’11.3904’’N, 7°57’45.864’’E) by gravity coring in 2019 during the RV HEINCKE cruise HE531. Sediments from 0-25 cm depth were selected from gravity core HE531/3-1 for SIP incubations. Anoxic incubations were set up using sediment and sterilized artificial sea water (ASW; composition 26.4 g NaCl, 11.2 g $MgCl_2 \cdot 6\ H_2O$, 1.5 g $CaCl_2 \cdot 2\ H_2O$ and 0.7 g KCl and 4.26 g Na_2SO_4 per liter) at a ratio of 1:4 (w/v). A 50 ml mixed slurry was dispensed in 120-ml serum bottles, sealed with butyl rubber stoppers. Residue air was exchanged with N_2. Triplicates were set up for each treatment,

containing a combination of ^{13}C-labeled and unlabeled MACs with only the methoxy group labeled (vanillin, 2-methoxybenzoic acid, 2-methoxyphenol) and ^{13}C-labeled and unlabeled bicarbonate (i.e., dissolved inorganic carbon - DIC) or only ^{13}C-labeled or unlabeled DIC as control (Table S1). DIC was supplied at a concentration of 10 mM, MACs were supplied at a concentration of 1 mM. One additional treatment for each of the tested MACs was amended with 5 mM sodium 2-bromoethanesulfonate (BES) for the inhibition of methanogenic activity. Samples were incubated at 10°C in the dark.

To determine suitable stopping time points for each of the SIP incubations, δ^{13}C values of CO_2 in the headspace of triplicate treatments were measured. For this 1 ml gas sample was injected into a Thermo Finnigan Trace GC connected via a GC III interface to a DELTA Plus IRMS (Finnigan MAT, Bremen, Germany) with settings as described previously (Ertefai et al. 2010). The δ^{13}C values were monitored as δ^{13}C (‰) relative to the Vienna Pee Dee Belemnite (VPDB) standard (Fig. S1). SIP enrichments were stopped once a δ^{13}C value of 1,000 ‰ was reached. The treatments containing 2-methoxyphenol were stopped after 81 days and treatments with vanillin after 111 days. No change in the δ^{13}C value was observed for the incubations containing 2-methoxybenzoic acid. Therefore, only one replicate was stopped after 188 days and used for nucleic acid SIP. The calculation of the percentage of ^{13}C from δ^{13}C is explained in the supplementary material.

4.1.3.2 Density separation

Nucleic acid extraction, removal of DNA, quantification, isopycnic centrifugation and gradient fractionation were performed according to Yin et al. (2019). To retrieve sufficient RNA from the incubations, a total of 3 g sediment was used from each replicate. For the 2-methoxybenzoic acid treatments, only one of the replicates was used for extraction. RNA was extracted following the phenol-chloroform extraction protocol by Lueders et al. (2004), with the following modifications: after the addition of polyethylene glycol 6000 for nucleic acid precipitation, samples were centrifuged at 20,817 *g* at 4°C for 90 min. DNA was removed using the RQ1 DNase kit (Promega, Madison, Wisconsin, USA) at 37°C for 45 min. Following DNase digestion, RNA was purified using phenol-chloroform-isoamyl and chloroform-isoamyl alcohol, and precipitation was performed using polyethylene glycol 6000. Samples were centrifuged at 20,817 *g* at 4°C for 90 min. RNA extracts of all triplicates were pooled to retrieve a sufficient amount of RNA for density gradient separation. RNA was quantified with Quanti-iT RiboGreen (Quanti-iT RiboGreen™ RNA-Kit, Invitrogen™, Thermo Fisher Scientific) and 0.2-1 µg was used for density separation by ultracentrifugation (Table S2-S5).

The CsTFA solution was prepared from $CsOH \cdot 2\,H_2O$ (99.5% purity, Sigma-Aldrich, Germany) and trifluoracetic acid (TFA, 99.9% purity, Carl Roth, Germany), autoclaved and 0.2 µm filtered for sterilization. The pH of the solution was adjusted to 7.0 and density to 2.0 g/ml (detailed protocol in supplementary material). From each treatment, RNA was separated into 14 fractions, with fraction 1 having the highest density and fraction 14 having the lowest (Fig. S2). From each of the treatments, fractions 3 (ultra-heavy, 1.829-1.835 g/ml), 5 (heavy, 1.818-1.825 g/ml), 7 (midpoint, 1.801-1.812 g/ml), 9 (light, 1.791-1.798 g/ml), and 11 (ultra-light, 1.778-1.784 g/ml) were used for cDNA synthesis and subsequently sequenced as in methods section 5. An overview of sequenced fractions per sample with their according densities and RNA amounts is shown in Tables S2-S5.

4.1.3.3 MAC enrichment setup

Sediment was collected from the Helgoland mud area (54°05'15.5" N, 7°58'05.5" E) by gravity coring in 2017 during the RV HEINCKE cruise HE483. Sediments from a 45-70 cm depth were selected from gravity core HE483/2-2 for slurry incubations. Anoxic slurry was prepared, similar to the SIP incubations. Two replicates were set up for each treatment, containing either 12 mM of one of the selected 15 methoxylated aromatic compounds (MACs) or no additional carbon substrate as control (Table S6). Due to low solubility of some MACs, not all substrate was dissolved. Duplicates were additionally amended with 5 mM 2-bromoethanesulfonate (BES) to inhibit methanogenesis (Liu et al. 2011). Samples were incubated for 458 days at 20°C.

4.1.3.4 Nucleic acid extraction

For each treatment of the MAC enrichments, 1.5 ml of the slurry was transferred anaerobically at days 92, 200, 367 and 458 for nucleic acid extraction. Nucleic acids were extracted following the phenol-chloroform extraction protocol by Lueders et al. (2004), with the following modifications: during precipitation with polyethylene glycol 6000, samples were incubated for 30 min at room temperature and centrifuged at 20,817 *g* at 4°C for 60 min. DNA pellets were washed twice with 1 ml cold 70% ethanol and eluted in 100 µl diethylpyrocarbonate (DEPC) treated water (Carl Roth, Germany). The quality of nucleic acids was checked with a NanoDrop 1000 spectrophotometer (Peqlab Biotechnologie, Erlangen, Germany).

4.1.3.5 16S rRNA gene amplicon sequencing and analysis

Illumina amplicon sequencing libraries of the MAC enrichments and SIP incubations were prepared for bacteria and archaea. Primers targeting the V4 region of the bacterial 16S rRNA

gene were Bac515F (5′-GTGYCAGCMGCCGCGGTAA-3′) (Parada et al. 2016) and Bac805R (5′-GACTACHVGGGTATCTAATCC-3′) (Herlemann et al. 2011), primers used for the V4 region of archaeal 16S rRNA gene were Arc519F (5'-CAGCMGCCGCGGTAA-3') (Ovreås et al. 1997) and Arc806R (5'-GGACTACVSGGGTATCTAAT-3') (Takai and Horikoshi 2000). The library preparation was performed according to Wunder et al. (2024). Thermal cycling conditions used for the sequencing PCR were as follows: initial denaturation at 95°C for 3 min, 30 cycles of denaturation at 95°C for 20 s, annealing at 60°C for 15 s, elongation at 72°C for 15 s, final elongation at 72 °C for 1 min. Sequencing of multiplexed libraries was performed at Novogene (Cambridge, UK) on the NovaSeq 6000 platform (2 x 250 bp, Illumina) in mixed orientation by ligation, therefore resulting in forward and reverse amplicon orientation in both forward (R1) and reverse reads (R2) for both, archaea and bacteria separately. Reads were demultiplexed and primers clipped using cutadapt v3.1. (Martin 2011) and further processed using the package dada2 v1.28.0 (Callahan et al. 2016) in R v4.3.1 (R Core Team 2023). Forward reads (R1) and reverse reads (R2) were trimmed with a maximum error rate of 2 to a summed length of 290 bp based on their quality profile, separately per sequence lane (Table S7 for more details). Subsequently, error rates were learned and samples were dereplicated and denoised independently for each library orientation, by pooling the data from all samples, using a modified LOESS function adapted for libraries with binned quality scores (Salazar 2020). Error-corrected R1 and R2 reads were merged into amplicon sequence variants (ASVs) and sequence tables for forward and reverse orientations were combined by reorientation of the reverse-forward ASVs. Chimeras, ASVs of unexpected lengths (< 249 bp and > 254 bp for bacteria; < 251 bp and > 255 bp for archaea) and singletons were removed. A bootstrap cutoff of 70 (MAC enrichments Bacteria) or 80 (MAC enrichments Archaea, SIP incubations Bacteria and Archaea) was used to perform taxonomic classification with the assignTaxonomy function of dada2 with a newly formatted GTDB r214 reference database (Parks et al. 2018) containing the complete 16S rRNA gene set. For further processing of ASVs, archaea were removed from the bacterial data set, all bacterial taxa were removed from the archaeal data set. Sequence community composition was checked with rarefaction curves (iNEXT package version 3.0.0, Fig. S3-S6) (Hsieh et al. 2016).

As the class Clostridia was our target group in this study, ASVs of the class Clostridia were screened against the 16S ribosomal RNA sequence database of the online NCBI BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=GeoBlast&PAGE_TYPE=BlastSearch, accessed 29.08.2024). The search algorithm was adjusted to blastn, searching somewhat similar sequences with BLAST+2.16.0 (Altschul et al. 1990). The

pairwise identity of ASVs of the class Clostridia was computed using the blastn function of BLAST+2.16.0 with an e-value threshold of 1 x 10^{-10} (Altschul et al. 1990); results are displayed in Fig. S7.

4.1.3.6 qPCR

16S rRNA gene copies of Archaea, Bacteria and Clostridia were quantified in all MAC enrichments for time points day 0, 92, 200, 367 and 458. Reaction mixtures contained 1x Takyon Master Mix (Eurogentec, Seraing, Belgium), 4 µg bovine serum albumin (Roche, Mannheim, Germany), 300 nM primers, 1 ng of DNA template or 2 µl of standard in a total reaction volume of 20 µl. The primer pairs used were 806F (5'-GACTACHVGGGTATCTAATCC-3') and 912R (5'-GTGCTCCCCCGCCAATTC CTTTA-3', annealing temperature 58°C) for archaea (Yu et al. 2005), 8Fmod (5'-AGAGTTTGATYMTGGCTCAG-3') (Satokari et al. 2001) and 338Rmod (5'-GCWGCCWCCCGTAGGWGT-3', annealing temperature 58°C) (Daims et al. 1999) for bacteria and newly designed 579F/747R (5'-AAGCCMCGGCTAACTACGTG-3' / 5'-CGCTACACTAGGAATTCC RCYT-3', annealing temperature 60°C) for Clostridia targeting Clostridia clones of different orders retrieved from the MAC enrichments (further details in supplementary material). Thermal cycling conditions were as follows: initial denaturation at 95°C for 5 min, 40 cycles of denaturation at 95°C for 30 sec, annealing at given temperatures for 30 sec and elongation at 72°C for 40 sec.

4.1.3.7 LC-MS/MS measurements of MACs

To track the decrease of targeted MACs, 0.5 ml of the slurry was sampled at days 0, 92, 200, 367 and 458. Samples were centrifuged at 20,817 *g* for 10 min for separating sediment from supernatant. The supernatant was centrifuged twice at 20,817 *g* for 10 min to remove the remaining particles. Sediment-bound MACs were extracted from the remaining sediment using the same volume of MeOH as the previously removed supernatant. Samples were sonicated for 1 hour (Elma Transsonic Digital T 830/H, Germany) and further centrifuged at 20,817 *g* until the remaining particles were removed. Supernatant and sediment extract were diluted in ASW to yield a dilution of 1:900. Standards in ASW and MeOH for each of the targeted MACs were set up in concentrations of 10 mM, 8 mM, 6 mM, 4 mM, 2 mM, and 1 mM and diluted as above in ASW or MeOH. MACs were quantified using an LC-MS/MS with an UHPLC coupled to a Q-Exactive Plus mass spectrometer (both Thermo Fisher Scientific). For further information see supplementary material.

4.1.3.8 Metagenomic analysis and annotation

Five samples containing high relative abundances of different Clostridia orders and families therein were used for metagenomic sequencing to reconstruct their metabolic potential. An aliquot of ~ 1 µg extracted DNA of MAC amended samples of day 458 and one SIP enrichment (3,4,5-trimethoxyphenol, 3,4,5-trimethoxybenzoic acid, 3,4,5-trimethoxycinnamic acid, 2-methoxyhydroquinone and ^{13}C vanillin + ^{13}C-DIC + BES) were sequenced on the Illumina HiSeq 4000 platform (2 x 150 bp) at Novogene (Cambridge, UK). Metagenomic reads were adapter and quality trimmed, and dereplicated using bbduk from bbmap version 38.86 (Bushnell 2014). *De novo* assemblies were generated for each sample using SPAdes v.3.15.5 with the flag *meta* (Prjibelski et al. 2020) and megahit v1.2.9 with the preset *meta-sensitive* (Li et al. 2015). Additionally, a co-assembly was generated from all five samples using megahit v1.2.9 with the preset *meta-sensitive* (Li et al. 2015). For read mapping, fasta headers of each of the assemblies were simplified with anvio-7.1 (Eren et al. 2021); contigs < 1000 bp were removed and subsequently the quality trimmed reads of each sample were mapped back to each of the single sample assemblies and the co-assembly using bowtie2 v2.3.5.1 (Langmead and Salzberg 2012). For each assembly, binning was performed with metabat2 v2.12.1 (Kang et al. 2019), CONCOCT v1.0.0 (Alneberg et al. 2014) and vamb v4.1.3 (Nissen et al. 2021), followed by bin refinement using the bin refinement module of metaWRAP v1.3.1 (Uritskiy et al. 2018). Bin refinements of both assemblies (SPAdes and megahit) for each of the single sample assemblies and bin refinement of the co-assembly were combined through bin dereplication by dRep v3.0.0 using the ANImf algorithm with a primary ANI of 0.9 (mash v2.3, (Ondov et al. 2016)) and a secondary ANI of 0.98 (fastANI v1.32 (Jain et al. 2018)) with a minimum aligned fraction of 0.5 (Olm et al. 2017). Dereplicated bins were reassembled using a modified bin reassembly based on the reassembly module of metaWRAP. Along with strict and permissive reassemblies, which are based on reads from all samples, reads from non-redundant clusters derived through dereplication were used for the reassembly of single bins with strict and permissive settings. Clostridia genomes were manually refined using anvio-7.1 (Eren et al. 2021). The quality of obtained MAGs was calculated with checkM2 v1.0.2 (Chklovski et al. 2022) and taxonomic classification was assigned through gtdbtk v2.1.0 (Chaumeil et al. 2022), using the GTDB database v214 (Parks et al. 2018). Additionally, average nucleotide identity (ANI) was calculated using fastANI v1.32 (Jain et al. 2018) and amino acid identity (AAI) was computed using fastAAI v0.1.18 (Konstantinidis et al. 2022).

For annotation of all medium to high quality (> 80% completeness and < 5% contamination) Clostridia MAGs, genes were predicted using prodigal v2.6.3 (Hyatt et al. 2010). For full

annotations, predicted genes were annotated using NCBI NR release 07.02.2024 (accessed 13.03.2024) (Sayers et al. 2022) and KEGG release 109 (accessed 31.12.2023) (Kanehisa et al. 2015) using diamond blastp v2.0.15 (Buchfink et al. 2021). Signal peptides were predicted using signalp v6.0 (Teufel et al. 2022). Hydrogenases were additionally searched using HMM profiles of the electron-bifurcating iron hydrogenase *hydABCD*, which was also observed in other acetogenic bacteria (Wang et al. 2013, Katsyv et al. 2023). The protein sequences of each hydrogenase subunit of the acetogenic model organism *Acetobacterium woodii* were used for the HMM profiles. HMM profiles were built for each subunit using hmmbuild, implemented in HMMER v3.3.2 (Eddy 2009). All Clostridia MAGs were screened for the iron hydrogenase subunits using hmmsearch. Genes with an e-value < 0.001 were selected and manually checked for the iron hydrogenase subunits. Only HMM hits with two or more subunits located consecutively on the same contig were considered. In addition, all other MAGs retrieved through this study were annotated using NCBI NR and KEGG, as described above.

4.1.3.9 Marker gene tree

For the calculation of a marker gene tree, additional Clostridia genomes were acquired, selecting species representatives of Clostridia from GTDB with completeness of > 90% and contamination of < 5%. All species representatives from orders, in which we detected Clostridia MAGs (methods section 8, Acetivibrionales, Ch29, Eubacteriales, Oscillospirales, Peptostreptococcales, Tissierellales) were selected as input for the tree and downloaded from RefSeq and GenBank. Additionally, 15 randomly chosen genomes from each other order within the Clostridia were selected to improve taxonomic placements of MAGs without order assignment, along with 20 genomes of the phylum Bacillota_B as an outgroup. Using the *de novo* workflow of gtdbtk v2.1.0 a multiple sequence alignment was computed using 120 marker genes. ModelFinder (Kalyaanamoorthy et al. 2017) as implemented within iqtree2 v2.2.2.7 (Minh et al. 2020) was used to identify a suitable model (LG+F+I+R10) for tree calculation. The tree was calculated using fasttree v2.1.11 with 1,500 bootstraps (Price et al. 2010) and manually grouped using iTol v6.9.1 (Letunic and Bork 2007), following the GTDB v214 taxonomy (Parks et al. 2018).

4.1.3.10 16S rRNA gene phylogenetic tree

For a 16S rRNA gene phylogenetic tree, 16S rRNA gene sequences were extracted from Clostridia MAGs derived from own metagenomic analyses, as well as the acquired GTDB Clostridia species representatives of the orders Acetivibrionales, Eubacteriales, Oscillospirales, Tissierellales and Peptostreptococcales (methods section 9) using barrnap v0.9 (Seemann

2018). Additional 16S rRNA gene sequences of the order Clostridiales were selected as outgroup. All 16S rRNA gene sequences, including the ASV sequences, were aligned using SINA v1.7.2 (Pruesse et al. 2012) against the SILVA nr99 v138.1 ARB database (Quast et al. 2012). Sequences were trimmed to fixed start and end positions of the full-length 16S rRNA gene (*E. coli* position 1,043 – 6,884) using seqtk v1.3-r106 (https://github.com/lh3/seqtk) and all gaps caused by one sequence were removed with clipkit v2.2.4 (Steenwyk et al. 2020). For tree construction, sequences with at least 1000 bp were chosen. Redundancy in sequences was removed by seqkit v2.3.1 (Shen et al. 2016). For all non-redundant sequences above 1,000 bp in length, a GTR+I+G4 model for the 16S rRNA gene phylogenetic tree was determined using modeltest-ng v0.1.7 (Darriba et al. 2020), which was further used for calculation of the tree with raxml-ng v1.1.0 (Kozlov et al. 2019). In total 50 starting trees were inferred; bootstrap convergence at a cutoff of 0.02 was reached after 1,200 trees. Shorter 16S rRNA gene sequences found in Clostridia genomes, along with ASVs from the SIP and slurry enrichment were added after tree calculation. Short sequences were placed into the existing tree using epa-ng v0.3.8 (Barbera et al. 2019) and gappa v0.7.1 (Czech et al. 2020). The tree was manually rooted and grouped in iTOL v6.9.1 (Letunic and Bork 2007).

4.1.3.11 Detection of methyltransferases (MTI) in Clostridia MAGs and species representatives

As the annotation did only yield one single gene assigned to the *mtvB* methyltransferase of the *O*-demethylase system (EC:2.1.1.382 in the MAG L3_d458_spades_bin_6_orig), a protein family model was constructed to search all reconstructed Clostridia MAGs. For this, *mtvB, omdB* and *mtoB* genes of *Methermicoccus shengliensis*, *Archaeoglobus fulgidus*, *Moorella thermoacetica*, *Acetobacterium dehalogenans* and our MAG were aligned using muscle v5.1 (Edgar 2022). From the alignment an HMM profile was built using hmmbuild implemented in HMMER v3.3.2 (Eddy 2009). All MAGs were searched for methyltransferase (MTI) genes using hmmsearch, validated against closely related genes and significant hits were selected (further details see Supplementary material). Using the same technique, we searched MTI genes and removed false positives in all Clostridia species representative genomes (methods section 9).

Additionally, orthofinder v2.5.5 (Emms and Kelly 2019) was used to search all orthologous genes in the MAGs. From the orthogroups we subset those containing MTI sequences. Constructed orthogroup trees were used to evaluate the evolution of MTI genes within the MAGs. For evaluation of the diversity of MTI genes in found Clostridia MAGs and species

representatives, MTI genes were clustered based on 70% identity and a 90% alignment coverage of the target with mmseq2 (Steinegger and Söding 2017). To establish if MTI genes were specific for used substrates, the mapped MTI genes in the five different metagenomes were counted using featureCounts v2.0.6 (Liao et al. 2013). Read counts were normalized using the gene length and sequencing depth.

4.1.3.12 Statistical analyses

Samples were analyzed based on their enrichment in Clostridia using the data from amplicon sequencing. By calculating centered log-ratio transformations using the library SpiecEasi v1.1.1 in R (Kurtz et al. 2015), samples were classified as enriched, if the centered log-ratio was > 8.75 (> 30% relative abundance) or not enriched (< 30% relative abundance). Coherence between Clostridia enrichment and the number of methoxy groups of the MAC substrate analyzed was tested using a Chi-Square test implemented in the package vegan v2.6-6.1 (Oksanen et al. 2024).

4.1.3.13 Data availability

All code used to analyze the data and create figures is available under https://github.com/mmaeke/MAC_degradation. Supplementary excel tables are also supplied in the repository. Clone sequences will be submitted to GenBank upon manuscript submission. Amplicon and metagenomic raw reads will be submitted to ENA upon submission. Until then, this data is available upon request.

4.1.4 Results

4.1.4.1 Methoxy group utilization by Clostridia

RNA stable isotope probing (RNA-SIP) enrichments with three ^{13}C-labeled methoxylated aromatic compounds (MACs): vanillin, 2-methoxyphenol, and 2-methoxybenzoic acid were set up to identify microorganisms involved in methoxydotrophic growth. MACs were only labeled at the carbon of the methoxy group to ensure labeling only by methoxydotrophic metabolism. Additional unlabeled or ^{13}C-labeled DIC was used as second carbon substrate, as methoxydotrophic growth requires the presence of both a MAC and DIC (Ragsdale and Pierce 2008).

In the vanillin (V) and 2-methoxyphenol (MP) amended SIP enrichments, we observed an increase of the δ^{13}C (‰) in headspace CO2 throughout the incubation, reaching up to 6566‰ (7.8% ^{13}C) in the vanillin amended sample after 111 days and 4013‰ (5.3% ^{13}C) in the 2-methoxyphenol after 81 days (Figure S1). No change of the δ^{13}C was observed in samples amended with 2-methoxybenzoic acid. In the vanillin and 2-methoxyphenol amended SIP enrichments, we observed high relative abundances of Clostridia in all treatments, reaching between 11 and 43% in ultra-heavy fractions when one or both substrates were labeled (Fig. S8), while Clostridia were not detected in 2-methoxybenzoic acid and DIC-only controls (Fig. S8-S9). Upon density separation of RNA, we detected the highest label in the middle fractions (light, midpoint, heavy) when only one of the amended substrates (MAC or DIC) contained the ^{13}C label (Fig. 1).

With ^{13}C-MAC and DIC, the highest relative abundance of Clostridia was detected in the midpoint fraction (V: 61%; MP: 26%). With both substrates labeled, the highest relative abundances were observed in only the heavier fractions (midpoint, heavy, ultra-heavy), with relative abundances of 21% (V) and 25% (MP) in the heavy and 17% (V) and 13% (MP) in the ultra-heavy fraction. Upon adding BES, the relative abundances increased slightly to 25% (V) and 38% (MP) in the heavy and 22% (V) and 43% (MP) in the ultra-heavy fraction. Two main ASVs were detected (sq1 and sq5), which had a pairwise sequence identity of 98.8% (Fig. S7). Two further ASVs (sq36, sq78) with pairwise sequence identities to sq1 and sq5 of 98.4-99.2% could be enriched upon BES addition in the 2-methoxyphenol treatments. All ASVs detected in the RNA-SIP treatments of vanillin and 2-methoxyphenol were assigned to Alkalibacteraceae unclassified below family level. Archaea were not labeled by adding any MAC (Fig. S10).

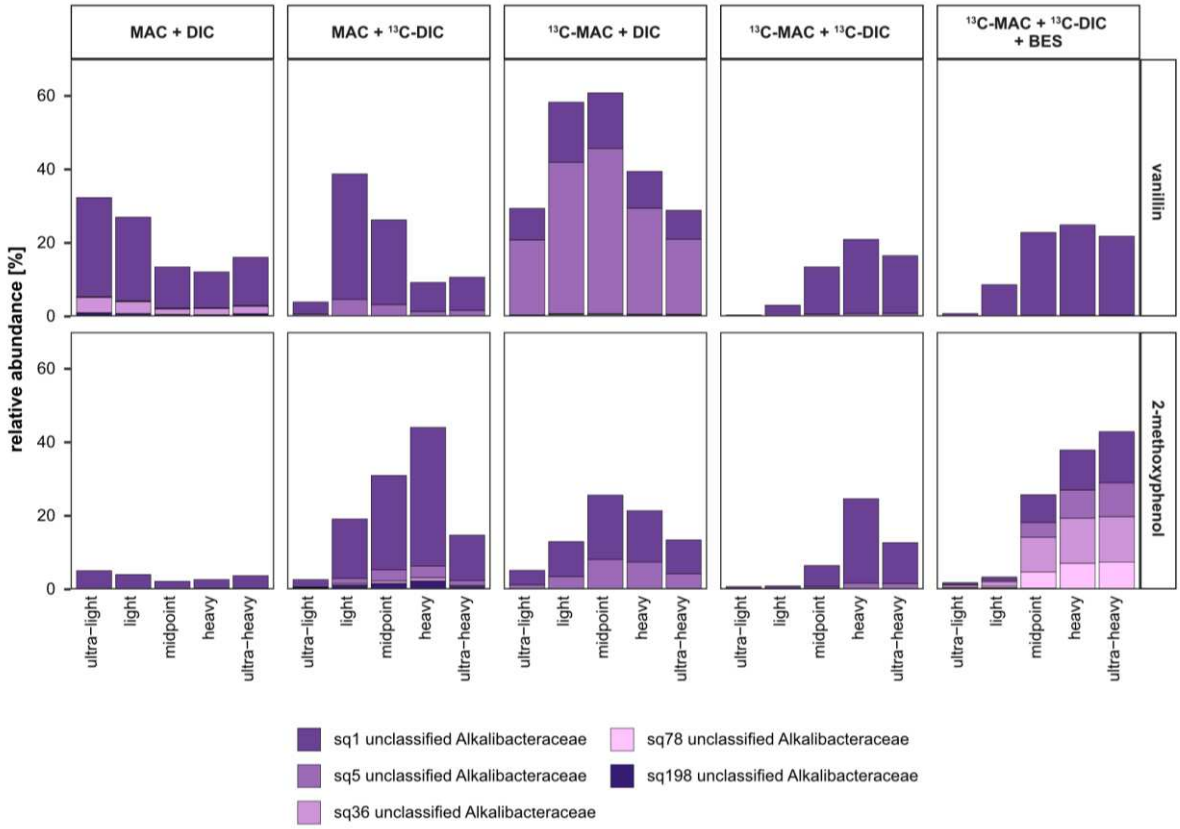

Figure 1 Relative 16S rRNA gene abundance of ASVs within the class Clostridia in density-resolved RNA fractions (ultra-heavy, heavy, midpoint, light, ultra-light). Different treatments were amended with vanillin or 2-methoxyphenol as methoxylated aromatic compound (MAC) and dissolved inorganic carbon (DIC). Labeled substrates are indicated by ^{13}C, for MACs only the methoxy group was labeled. One of the treatments was additionally amended with sodium 2-bromoethanesulfonate (BES), an inhibitor for methanogenesis.

4.1.4.2 Enrichment of Clostridia with MACs in marine sediment slurry incubations

To further identify the substrate spectrum of MACs used by the class of Clostridia, enrichments utilizing a variety of MACs were set up (Table S6). Here we investigated the process of non-methanogenic methoxydotrophy by adding sodium 2-bromoethanesulfonate (BES) to inhibit methanogenesis in enrichments during an incubation period of 458 days. In 12 treatments, we observed an enrichment of the class Clostridia (up to 95% in bacteria), represented by five orders: Acetivibrionales, Eubacteriales, Oscillospirales, Peptostreptococcales and Tissierellales (Fig. 2, Fig. S11-12). Within the control samples, no enrichment was observed (0.2-0.5%). The reported relative abundances reflect the maximum relative abundance during the incubation period of 458 days in the treatments.

Acetivibrionales were only enriched in the treatment with 3,4,5-trimethoxybenzoic acid, reaching 89% relative abundance. Eubacteriales were detected in all treatments with highest relative abundances in treatments with vanillin (78%), 2-methoxyphenol (60%), 3,4,5-trimethoxyphenol (83%), 3,4-dimethoxycinnamic acid (62%), 3,4,5-trimethoxycinnamic

acid (61%), 2-methoxyhydroquinone (40%), 1,2,4-trimethoxybenzene (86%) and 3,4,5-trimethoxybenzaldehyde (69%). Specifically, one ASV sequence (sq1), assigned to unclassified Alkalibacteraceae, could be detected in nine treatments and was highly abundant in seven (Table S8, Fig. S12). Oscillospirales were abundant in the treatments with 3,4-dimethoxycinnamic acid (20%), 3,4,5-trimethoxycinnamic acid (19%) and 3,4,5-trimethoxybenzaldehyde (48%). Peptostreptococcales were most abundant in the treatment with 3,4,5-trimethoxybenzaldehyde (24%). Tissierellales were most abundant at day 200 in the treatment with 2-methoxyhydroquinone (26%). Further details on the abundance of ASVs is given in the supplementary material.

In addition to Clostridia, we observed an enrichment of members of the phylum Bacillota_B in treatments with 2-methoxybenzoic acid (46%), 2-methoxyhydroquinone (42%), and 3,4-dimethoxycinnamic acid (44%).

Using a chi-square test, we tested for coherence between Clostridia enrichment and the number of methoxy groups of MACs. We identified that more samples were observed to be enriched (two methoxy groups: 8 samples; three methoxy groups: 27 samples) in Clostridia (relative abundance of Clostridia > 30%) in the MAC treatments containing two or three methoxy groups than would be expected (two methoxy groups: 6.5 samples; three methoxy groups: 19.6 samples). At the same time, more samples without Clostridia enrichment (relative abundance of Clostridia < 30%) were observed for MACs with one methoxy group (28 samples) than would have been expected (19.1 samples) (Table S9). The significance of these results was confirmed by a *p*-value of 0.0007.

In addition to the community changes based on 16S rRNA gene relative abundances, we quantified 16S rRNA gene copies of Clostridia within our enrichments over 458 days (Fig. S13). We detected higher gene copies of Clostridia in all treatments compared to the unamended control (up to 2.6×10^6). The highest increases in gene copy numbers were observed in treatments with vanillin (4.58×10^7, 17-fold), 3,4,5-trimethoxyphenol (3.27×10^8, 126-fold), 3,4,5-trimethoxybenzoic acid (2.18×10^8, 83-fold), 2-methoxyhydroquinone (2.06×10^8, 79-fold), 3,4-dimethoxycinnamic acid (6.86×10^7, 26-fold) and 3,4,5-trimethoxycinnamic acid (1.32×10^8, 50-fold).

Using mass spectrometry, we tracked the decrease of targeted MACs during the incubation of 458 days. In all treatments, the concentration of MACs decreased throughout the incubation time of 458 days with increasing relative abundance of Clostridia in single enrichments (Fig. S14). 2-methoxyphenol and 2-methoxyhydroquinone could not be detected using this method, likely due to low solubility of the substrates.

No Clostridia enrichment was observed within a slurry-only control and further tested treatments (Fig. S11, Fig. S15). To exclude archaea as participants in methoxydotrophy, we analyzed archaeal 16S rRNA gene amplicon sequences and quantified archaeal gene copies in all enrichments. Compared to the control (Fig. S16-17), we did not observe enrichment of archaeal taxa, nor did archaeal 16S rRNA gene copy numbers increase during our incubation time of 458 days (Fig. S17).

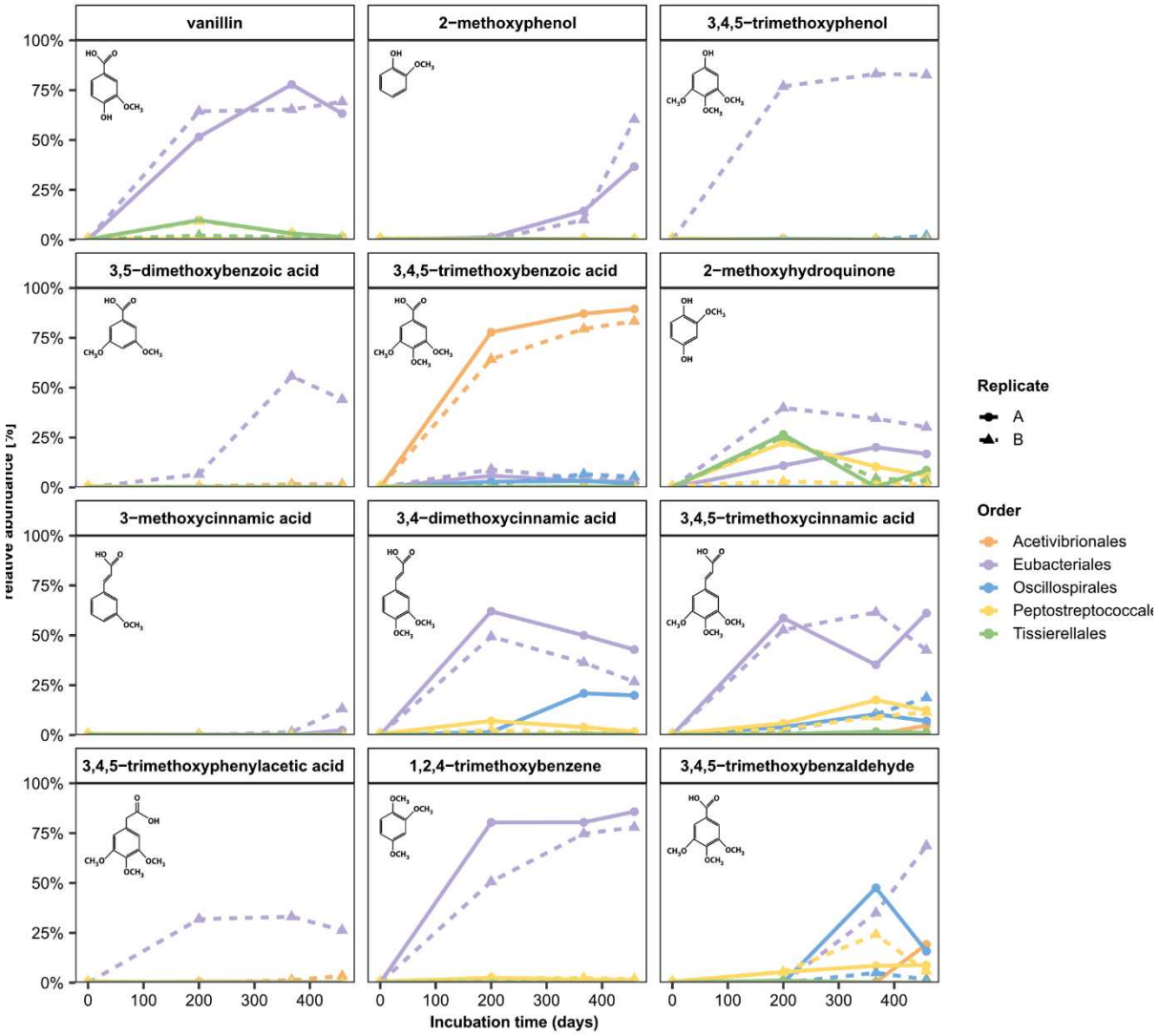

Figure 2 Relative 16S rRNA gene abundance of Clostridia within MAC-amended samples (amended with indicated MAC and BES) during an incubation period of 458 days. Colors indicate different orders within the class of Clostridia. Line types indicate the two replicates. MAC structures of the added MAC are displayed in the upper left corner of each graph. For this plot, only Clostridia orders with a relative abundance of at least 5% in any of the samples were selected.

4.1.4.3 Phylogenetic novelty among detected Clostridia

Abundant Clostridia ASVs with at least 5% relative abundance in any of the MAC or SIP enrichments were placed in a 16S rRNA gene phylogenetic tree (Fig. 3). None of the ASVs could be assigned to a type strain on species level (Table S10-S11). Yet, we observed that most of the identified ASVs grouped closely within the five identified orders. Unclassified Alkalibacteraceae ASVs found in the MAC and the SIP enrichments showed high 16S rRNA gene sequence identity (95.2-100%, Fig. S7) and showed the highest sequence identities to the type strain of *Alkalibaculum sporogenes* (94.02-98.01%; M08DMB species).

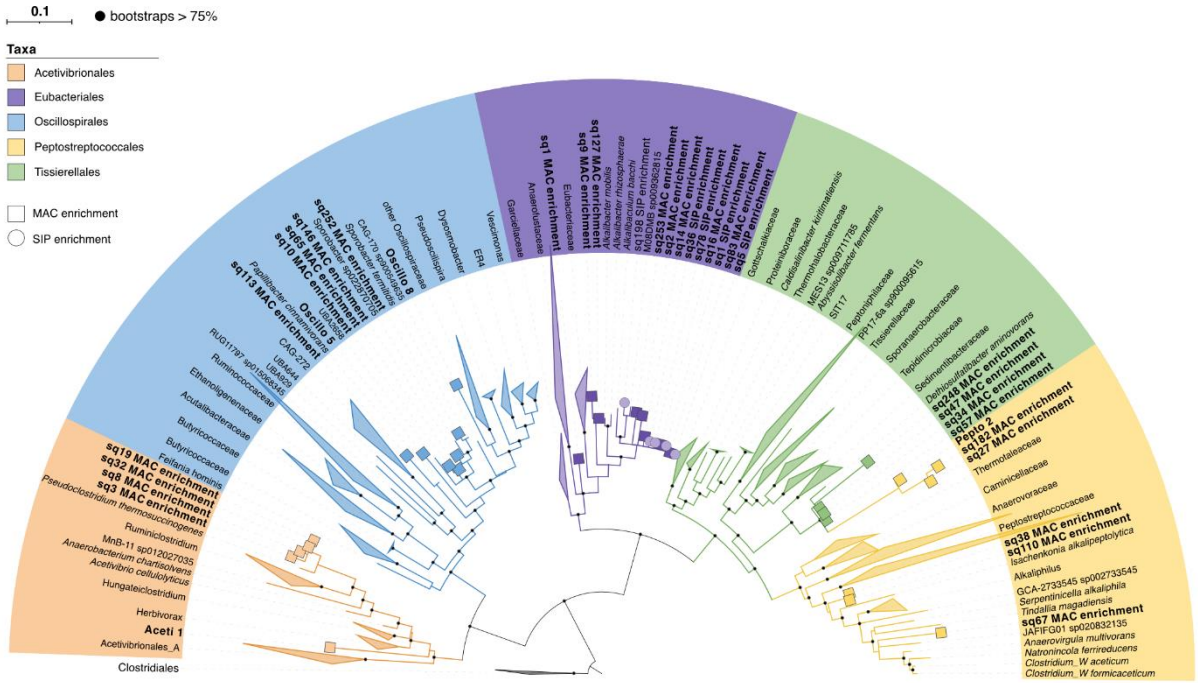

Figure 3 16S rRNA gene phylogenetic tree of the class Clostridia. 16S rRNA genes for tree reconstruction were retrieved from Clostridia species representatives present on the GTDB database (completeness > 90%, contamination <5%) of the five orders: Acetivibrionales, Eubacteriales, Oscillospirales, Peptostreptococcales and Tissierellales. Outgroup sequences were derived from Clostridiales species representatives (completeness > 90%, contamination <5%). The tree was calculated with raxml, bootstrap convergence of 0.02 was reached after 700 trees. ASVs and other shorter sequences were added after tree calculation. Single orders are indicated by color, ASVs from different enrichments are indicated by shape.

By metagenomic sequencing, we aimed to gain further insights into the metabolic potential of the identified Clostridia. For this, we performed metagenomic sequencing of multiple samples. In total, we reconstructed 26 medium- to high-quality metagenome-assembled genomes (MAGs) (completeness > 80%, contamination ≤ 5%) assigned to the class Clostridia (Table S12). The completeness of single MAGs ranged between 83.77-99.94%, and contamination ranged between 0-5% (Table S13). Though we could not detect 16S rRNA genes in most of the MAGs, MAGs were assigned to the five previously found orders using the GTDB

r214 (Parks et al. 2018) (Acetivibrionales, Eubacteriales, Oscillospirales, Peptostreptococcales, and Tissierellales) and follow a similar taxonomic classification as was initially detected for the ASVs (Table S12). One additional MAG was assigned to the order Ch29; another MAG could not be assigned to any known Clostridia order and was regarded as novel on order level based on relative evolutionary divergence (RED) computed during the GTDB classification. None of the four MAG-derived 16S rRNA genes could be assigned to any ASV on species level (Table S14). Using 120 bacterial marker genes, we computed a phylogenomic tree (Fig. 4).

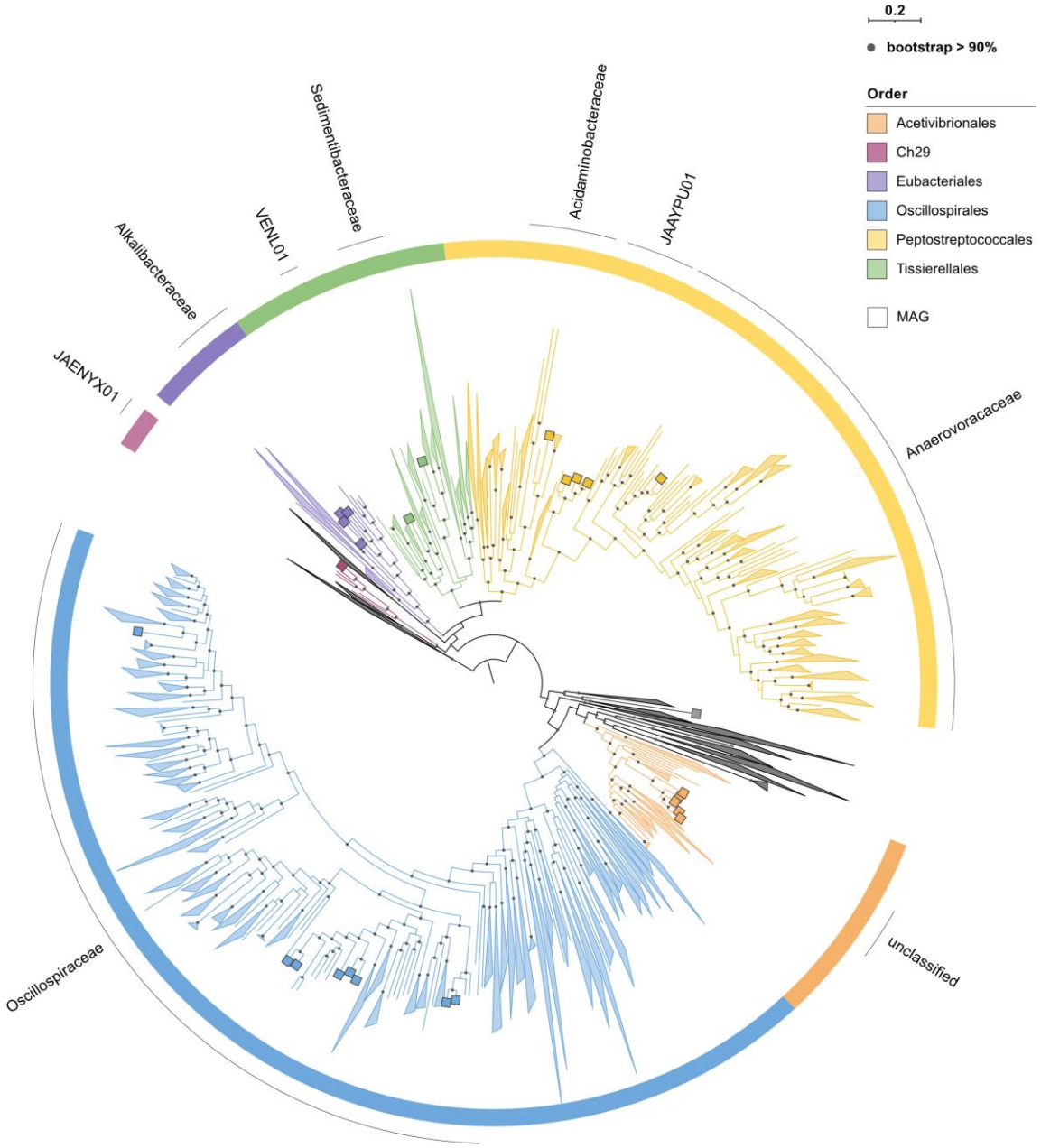

Figure 4 Marker gene tree of 120 bacterial maker genes for 3170 Clostridia species representatives (completeness > 90%, contamination < 5%) and 26 MAGs of Clostridia reconstructed by metagenomic sequencing. The tree was computed using fasttree v2.1.11 with 1500 bootstraps. Orders MAGs were found in, are indicated by color. Labels indicate the families these MAGS were assigned to.

Based on the GTDB classification as well as computed average nucleotide identity (ANI) (Fig. S18) and amino acid identity (AAI) (Fig. S19) we detected a novel family within the Acetivibrionales, a novel family and genus within the Eubacteriales and novel genera within the Oscillospirales (Table S12). Further details on the phylogenomic classification are given in the supplementary material.

4.1.4.4 Methoxydotrophy via the reductive acetyl-CoA pathway as the primary strategy to degrade MACs

We analyzed all MAGs for the metabolic potential of methoxy group utilization (Table S15). The methyltransferase subunit MTI facilitates the initial step of demethoxylation (Khomyakova and Slobodkin 2023). Therefore, we used the presence of the MTI gene as an indicator for the presence of the *O*-demethylase system. The MTI genes were present in all MAGs of the orders Acetivibrionales and Eubacteriales, all except one Oscillospirales MAGs, two Peptostreptococcales MAGs and the Sedimentibacteraceae (Tissierellales) MAG (Fig. 5). To test if MTI genes potentially act on methoxy groups of polyaromatic compounds too large to be imported into the cell, all MTI genes were screened for signal peptides, however none of the MTI genes contained signal peptide signatures (Table S16). Both orders, Acetivibrionales and Eubacteriales did contain all genes required for the methoxy group utilization, including genes of the methyl branch of the reductive acetyl-CoA pathway: formate dehydrogenase (*fdh*), formate-tetrahydrofolate ligase (*fhs*), methylenetetrahydrofolate dehydrogenase (*folD*), methenyltetrahydrofolate cyclohydrolase (*fchA*), methylenetetrahydrofolate reductase (metF), and the carbonyl branch: anaerobic carbon-monoxide dehydrogenase (*acsA*), acetyl-CoA decarbonylase/synthase CODH/ACS complex subunits beta, gamma and delta (*acsBCD*) and the 5-methyltetrahydrofolate corrinoid/iron-sulfur protein (*acsE*). Acetate formation from acetyl-CoA could be facilitated by phosphate acetyltransferase (*pta*) and acetate kinase (*ack*). MTI genes containing MAGs of the order Oscillospirales did contain genes for the reductive acetyl-CoA pathway however only two of the MAGs (Oscillo_2, Oscillo_3) did encode genes for the formate dehydrogenase. The genes for the carbon-monoxide dehydrogenase and CODH/ACS complex subunits beta, gamma and delta were absent in the *Sporobacter* Oscillo_5 MAG despite being detected in a second *Sporobacter* MAG (Oscillo_4). The Ch29 MAG did lack genes for the CODH/ACS complex (*acsBCDE*). In the order Peptostreptococcales, all MAGs lacked genes for the formate dehydrogenase and CODH/ACS complex subunits *acsBCDE*. The Tissierellales MAG Tissi_1 contained all genes of the Wood Ljungdahl pathway, only lacking the *acsA* gene. The Tissierellales MAG Tissi_2 and

unclassified Clostridia MAG Clostridia_1 lacked most genes involved in the reductive acetyl-CoA pathway.

To facilitate redox balancing, all MAGs except Pepto_2 contained genes for at least five subunits of the Rnf complex. Additionally, V/A-type and F-type ATP synthase genes were present in all MAGs, though they partially lacked a few subunits of either. Last, all MAGs had genes for the Fe-Fe hydrogenase *hydABC*.

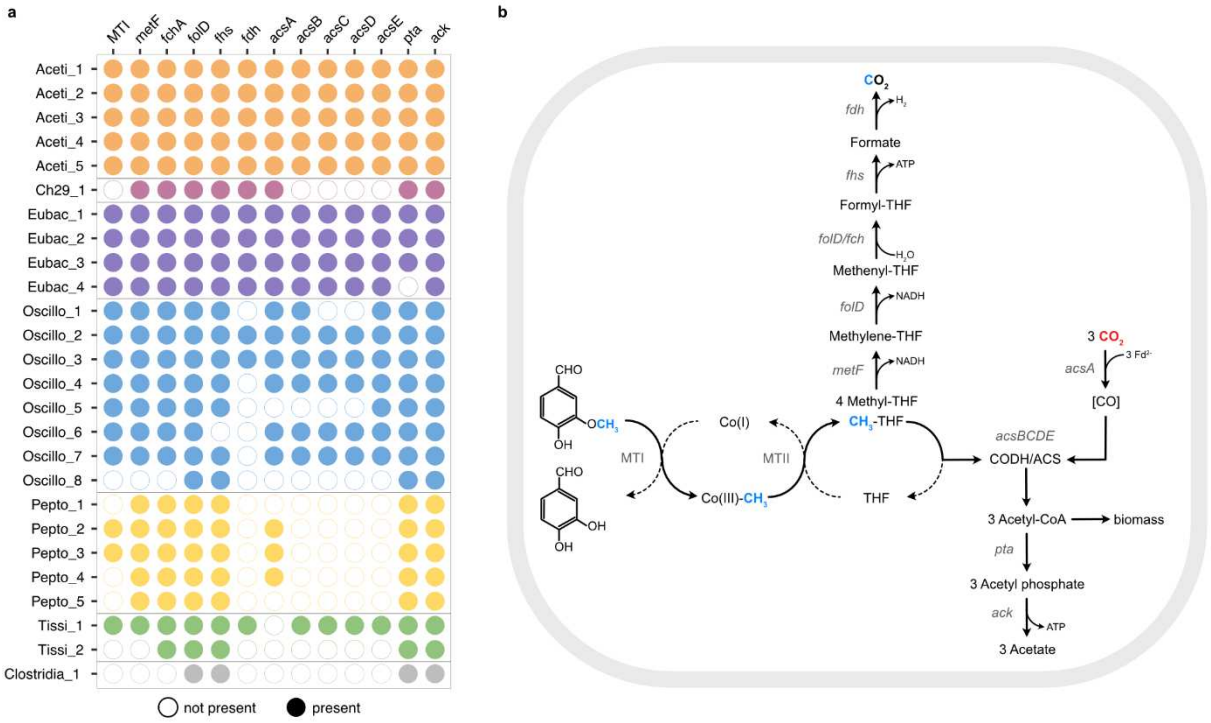

Figure 5 Metabolic potential of Clostridia **(a)** Presence and absence of genes involved in methoxydotrophy using the reductive acetyl-CoA pathway for 26 Clostridia MAGs. Colors indicate different orders within the Clostridia (orange - Acetivibrionales, pink - Ch29, purple - Eubacteriales, blue - Oscillospirales, yellow - Peptostreptococcales, green - Tissierellales, grey - unclassified). **(b)** Metabolic pathway reconstruction of methoxydotrophy using the reductive acetyl-CoA pathway. The methyl group derived from the MAC is colored in blue, and the CO_2 fixed in this process is colored in red. Gene abbreviations: MTI - methyltransferase I, MTII - methyltransferase II, *metF* - methylenetetrahydrofolate reductase, *folD* - methylenetetrahydrofolate dehydrogenase, *fch* - methenyltetrahydrofolate cyclohydrolase, *fhs* - formate-tetrahydrofolate ligase, *fdh* - formate dehydrogenase, *acsA* - anaerobic carbon-monoxide dehydrogenase, *acsBCDE* - acetyl-CoA decarbonylase/synthase, CODH/ACS complex, *pta* - phosphate acetyltransferase, *ack* - acetate kinase.

4.1.4.5 High MTI gene abundance in Clostridia MAGs

In order to facilitate demethoxylation, microorganisms require an *O*-demethylase system consisting of a substrate-specific methyltransferase I (MTI), a cobalamin-binding corrinoid protein and methyltransferase II (MTII), which transfers the methyl group to a terminal acceptor (Khomyakova and Slobodkin 2023). The subunit MTI catalyzes the initial methyl transfer from the methoxy group of a MAC to the corrinoid protein. Therefore, we screened all Clostridia MAGs for the presence of these genes to establish if methoxydotrophy was feasible.

The highest number of MTI genes was detected in MAGs of the orders Eubacteriales (up to 65 genes) and Oscillospirales (up to 52 genes) (Fig. 6, Table S17). Besides, MAGs of the order Acetivibrionales had up to 22 genes. Most MAGs of the order Peptostreptococcales did not contain any MTI genes. Only one MAG of the family Anaerovoracaceae was found to have two genes, one of three MAGs of the family JAAYPU01 contained one gene. Within the order Ch29, the unclassified Clostridia and MAGs of the genus JAEWRY01 (Oscillospiraceae), JAFGAC01 (Acidaminobacteraceae) and the unclassified VENL01 family (Tissierellales), no MTI genes were detected.

To further investigate the evolutionary history of MTI genes, we computed orthogroups for all found Clostridia MAGs and gene trees for the most abundant orthogroups. Orthogroups are defined as genes descending from a single last common ancestor (Fitch 1970). Identified MTI genes were sorted into 14 orthogroups, with most genes assigned to five orthogroups (OG 25, 39, 48, 179, 257) (Fig. S20). Orthogroup 25, containing 128 MTI genes in total was dominated by MTI genes found in the order Eubacteriales (75% of MTI genes) and Oscillospirales (20% of MTI genes). MTI genes of both orders formed two distinct clusters in the orthogroup tree (Fig. S21). Orthogroups 39 (94 MTI genes), 48 (87 MTI genes) and 257 (36 MTI genes) mainly consisted of MTI genes found in Oscillospirales MAGs (82%, 59% and 83% of MTI genes). For multiple genes within the trees of the orthogroups 39 (Fig. S21) and 257 (Fig. S22), various branches contained both MTI genes derived from Oscillospirales and Eubacteriales. The tree of orthogroup 48 formed two distinct clusters, one dominated by MTI genes of the order Oscillospirales, while the other consisted of MTI genes derived from the order Eubacteriales. Most MTI genes of the order Acetivibrionales were detected in orthogroup 179 (44 MTI genes, 82% MTI genes of Acetivibrionales). Within the tree, a few genes of the orders Eubacteriales and Oscillospirales were detected; however, they clustered with greater distance to genes of the Acetivibrionales (Fig. S22). Single MTI genes of the orders Peptostreptococcales and Tissierellales could be identified in orthogroups 25, 48, 179, 3021 and 4337, clustering closely with MTI genes of the orders Eubacteriales, Oscillospirales, or Acetivibrionales (Fig. S20-S22).

4.1.4.6 High MTI gene diversity in Clostridia MAGs

Additionally, we aimed to evaluate the MTI gene diversity within Clostridia. For this, we clustered MTI genes of all found MAGs based on 70% identity to retrieve functional similar genes. In total, we identified 260 unique MTI gene clusters, of which 79 (30%) gene clusters were found in multiple MAGs (Fig. 6). The remaining 181 (70%) MTI gene clusters were

detected as singletons. The highest numbers of singletons were detected in the unclassified Alkalibacteraceae MAG Eubac_1 (47 singletons, 72% of genes), M08DMB MAG Eubac_3 (25 singletons, 54% of genes) and *Sporobacter* MAG Oscillo_6 (23 singletons, 62% of genes). High proportions of singletons were also observed for Aceti_5 (55%) and Eubac_4 (100%). Of the non-singleton gene clusters, 22 could be identified in up to three gene copies in the different MAGs (further details in supplementary material).

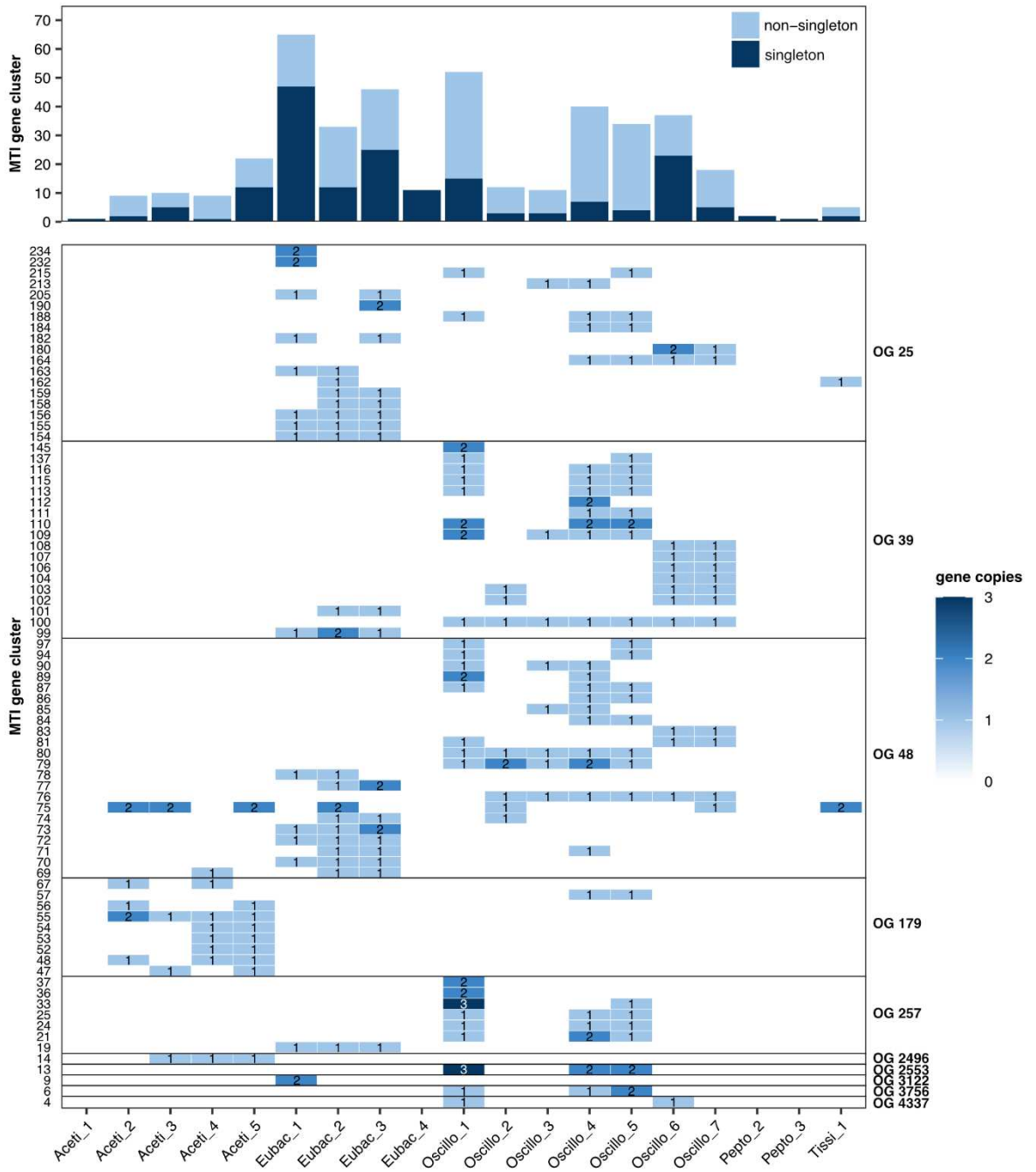

Figure 6 Heatmap of non-singleton MTI genes in Clostridia MAGs, sorted by orthogroup. Each row represents a unique MTI gene cluster. The according gene clusters are listed in Table S18. Gene clusters present as singletons only were not included in the heatmap.

4.1.4.7 High MTI gene abundance and diversity in Clostridia species representatives

Our findings are consistent with MTI gene abundance and diversity in all searched Clostridia species representatives of the GTDB r214. The highest numbers of homologous MTI genes were detected in the orders Eubacteriales (up to 70 MTI genes) and Acetivibrionales (up to 40 MTI genes), yet also high numbers of homologous genes could be observed in some genomes of the order Oscillospirales (up to 35 MTI genes) (Fig. 7a-b, Fig. S23). Notably, when comparing the MTI gene abundance and diversity in MAGs and Clostridia species representatives, the abundance and diversity of MTI gene copies in Oscillospirales MAGs were found to be outliers. Overall, the average number of genes was low in most families of the order Oscillospirales (Fig. S23), though outliers in the Oscillospirales families Acutalibacteraceae, Ruminococcaceae, Oscillospiraceae, and CAG-272 indicated a high abundance of MTI genes in some families. The order Ch29, from which we reconstructed one MAG, contained no genomes with MTI genes. We screened further orders of the Clostridia and could observe that beyond these five orders found in this study, more orders in the Clostridia exhibited high gene numbers of the MTI gene, e.g., Caldicoprobacterales, DUPQ01, Lachnospirales or Mahellales (Fig. S24). Furthermore, we observed that among all Clostridia genomes of the five orders Acetivibrionales, Eubacteriales, Oscillospirales, Peptostreptococcales and Tissierellales genomes with high MTI gene abundance (> 30 MTI genes) contained increasingly more copies of MTI genes, resulting in a non-proportional increase of MTI gene diversity (Fig. 7c).

4.1.4.8 Link of the specific MTI gene clusters to MACs

To establish which MTI genes were specific for substrates used in the enrichments, MTI genes were mapped to the five different metagenomes, and mapped reads were counted and normalized. Of 247 detected homologous MTI gene clusters, 140 (57%) were unique in single treatments (Fig. 8).

Four unique MTI gene clusters affiliated with the Oscillo_2, Aceti_1 and Pepto_3 MAG were identified in the 3,4,5-trimethoxyphenol sample. Though the Oscillo_2 (2%), Aceti_1 (1.5%) and Pepto_3 (0.1%) MAGs were only of low abundance in the 3,4,5-trimethoxyphenol sample, compared to the Eubac_4 MAG (30%, Fig. S25), no sample-specific genes were identified for Eubacteriales. Forty unique MTI gene clusters of the Eubac_1 MAG were detected in the sample containing 3,4,5-trimethoxybenzoic acid. Nine additional gene clusters could be assigned to Acetivibrionales MAGs. Especially the Aceti_3 MAG showed a high abundance of 49% in this sample (Fig. S25). MTI gene clusters found only in the 3,4,5-

trimethoxycinnamic acid (35 unique MTI gene clusters) and 2-methoxyhydroquinone (23 unique MTI gene clusters) treatments were derived from Oscillospirales. MTI gene clusters of the 3,4,5-trimethoxycinnamic acid treatment were found in multiple different MAGs of the order Oscillospirales (Oscillo_1, Oscillo_3, Oscillo_4, Oscillo_5 and Oscillo_7). Within the 2-methoxyhydroquinone treatment, all specific MTI gene clusters were affiliated with the MAG Oscillo_6. Twenty-seven unique MTI gene clusters of the Eubac_3 MAG were found in the vanillin sample of the vanillin + DIC + BES RNA-SIP treatment. Only one cluster was additionally found in the Eubac_1 MAG. Most gene clusters found in multiple treatments were only detected in single orders (Fig. S26). Only a few gene clusters were shared among multiple orders (69,71,75,162).

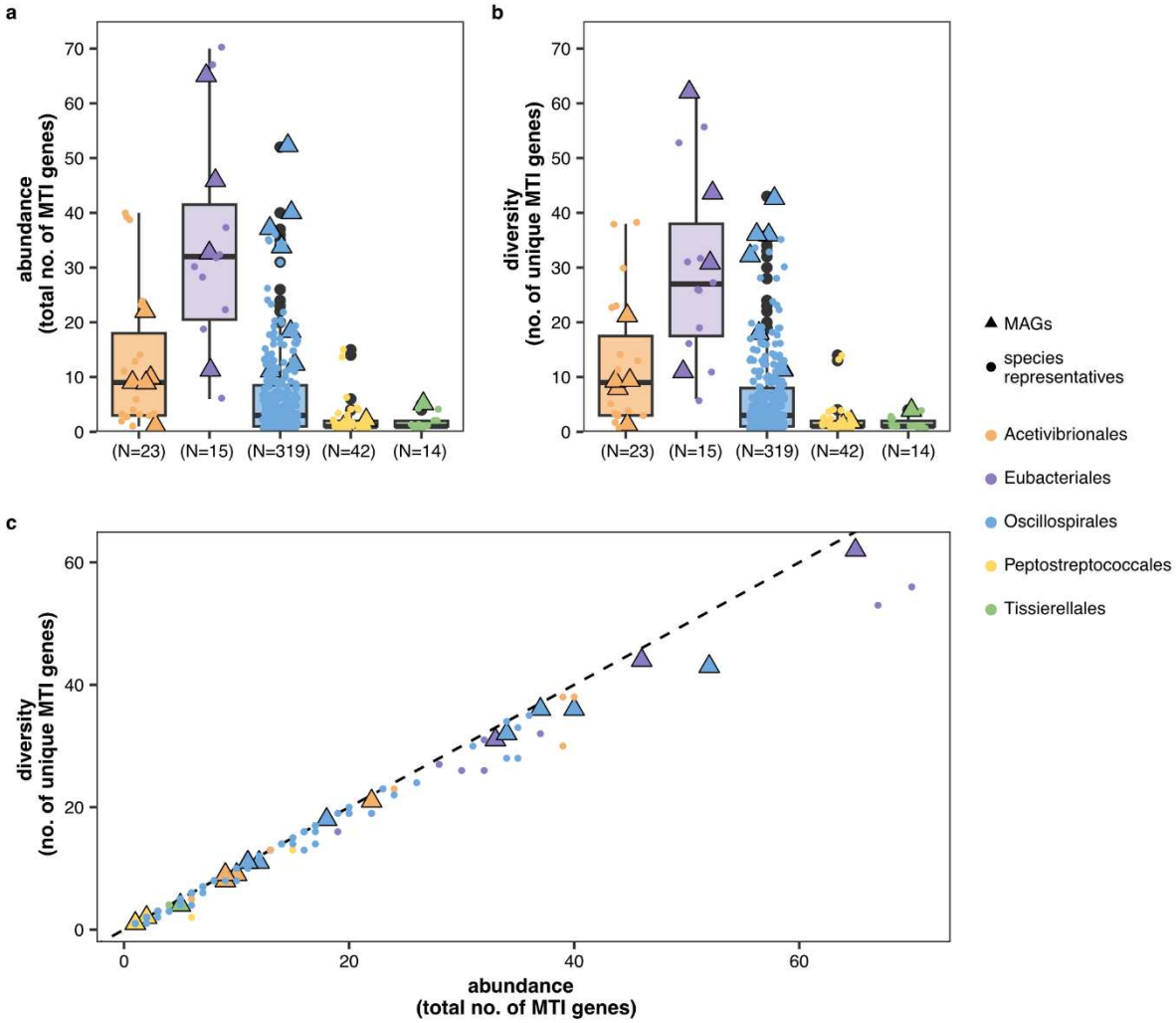

Figure 7 MTI gene abundance and diversity in Clostridia. **(a)** Abundance of MTI genes in MAGs and all Clostridia species representatives (completeness > 90%, contamination < 5%) on GTDB. **(b)** Diversity of MTI genes in MAGs and all Clostridia species representatives. **(c)** Diversity vs. abundance of MTI genes in MAGs and Clostridia species representatives.

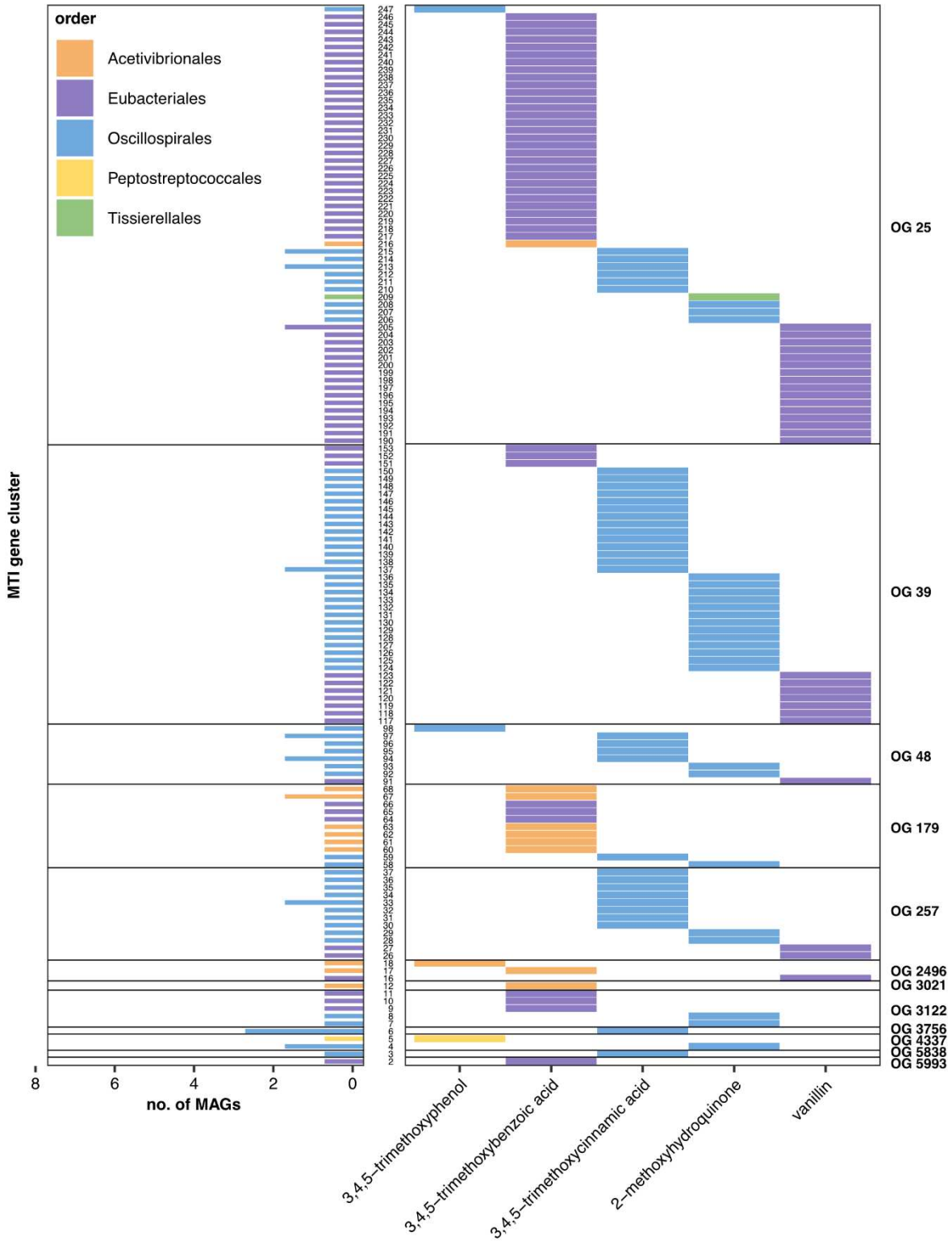

Figure 8 Presence of MTI gene cluster in different treatments. Numbers between the plots indicate the gene cluster (Table S16). The barplot shows the number of unique MAGs per order, present per MTI gene cluster; horizontal lines in the plot separate orthogroups (OG). Substrate-specific gene clusters are indicated by color for the corresponding order.

4.1.4.9 Aromatic ring degradation by other community members

The remaining hydroxylated aromatic ring structures could be further degraded through ring cleavage. We could detect multiple genes of subunits of the gallate decarboxylase (*lpdABC*) which catalyzes the reaction of 3,4,5-trihydroxybenzoate to 1,2,3-trihydroxybenzene (pyrogallol) and CO_2 (Grant and Patel 1969) and pyrogallol hydroxytransferase catalyzing the reaction of 1,2,3-trihydroxybenzene to 1,3,5-trihydroxybenzene (phloroglucinol) using 1,2,3,5-tetrahydroxybenzene as co-substrate (Brune and Schink 1990) in some MAGs of the orders Acetivibrionales and Oscillospirales (Table S15).

In order to investigate if other community members were involved in aromatic ring cleavage, additional bacterial (125 MAGs) and archaeal MAGs (28 MAGs) derived through the metagenomic analysis were screened for genes involved (MAG overview in Table S19). Genes of the benzoyl-CoA reductase *bcrABCD* or *bamBC* could be detected in MAGs affiliated with the orders Thermoanaerobaculales (Acidobacteriota), Tissierellales (Bacillota_A), E26-bin7, UBA7937 (Anaerolinea), Desulfatiglandales, Desulfobacterales (Desulfobacterota), Desulfuromonadales, G020346125 (Desulfobacterota_F) and Rhizobiales (Pseudomonadota) (Table S20). Yet, only in two of the Anaerolinea MAGs, two Desulfobacterales and one Desulfuromonadia MAG, we additionally identified all other genes required for the upper benzoyl-CoA pathway, including cyclohexa-1,5-dienecarbonyl-CoA hydratase (*dch*), 6-hydroxycyclohex-1-ene-1-carbonyl-CoA dehydrogenase (*had*) and 6-oxocyclohex-1-ene-carbonyl-CoA hydrolase (*oah*).

Overall, all MAGs affiliated with the cleavage of aromatic ring structures were low in abundance in the analyzed samples, reaching a maximum relative abundance of 0.8% (Fig. S27).

4.1.5 Discussion

Terrestrial organic matter deposited in marine coastal sediments introduces large amounts of highly recalcitrant organic carbon into the marine biosphere (Burdige 2005, Bianchi 2011). While first studies investigated the potential of depolymerization of lignin in the marine water column and sediments, the fate of remaining lignin-derived monomers remains mostly unexplored in fully anoxic marine sediments (Woo and Hazen 2018, Wang et al. 2021, Ley et al. 2023, Peng et al. 2023, Yu et al. 2023b). Here, we investigated the fate of such remaining monolignols, which are available as methoxylated aromatic compounds (MACs).

4.1.5.1 Clostridia are the primary degraders of MACs in anoxic marine sediments of the Helgoland mud area

We identified Clostridia as the main degraders of vanillin- and 2-methoxyphenol-derived methoxy groups by RNA stable isotope probing. Among Clostridia, the family Alkalibacteraceae within the order Eubacteriales was the sole family identified to be involved in methoxy group utilization of these two susbtrates. The closest type strain of identified Alkalibacteraceae ASVs was *Alkalibaculum sporogenes* (97.21-89.01% sequence identity), a Clostridia species found in salsa lakes of the terrestrial mud volcano Karabetova Gora in Russia (Khomyakova et al. 2020). Studies showed the growth of *Alkalibaculum sporogenes* on the MACs 3,4-dimethoxybenzoic acid and 2-methoxyphenol. Though we did not use 3,4-dimethoxybenzoic acid as a substrate in the SIP enrichment, we tested 15 different MACs in slurry enrichments to investigate the substrate spectrum of Clostridia. Multiple high identical ASVs detected by SIP and the MAC enrichments (SIP: sq1, 5, 36, 78, 198; slurry: sq2, 14, 16, 83, 253) could be identified in vanillin, 3,4-dimethoxycinnamic acid, 3,4,5-trimethoxycinnamic acid and 2-methoxyhydroquinone MAC enrichments, suggesting that more than the MACs analyzed in the RNA-SIP experiment can be degraded by members of this genus. However, *Alkalibaculum* ASVs were not detected in the 3,5-dimethoxybenzoic acid enrichment, emphasizing differences to the terrestrial *Alkalibaculum sporogenes*.

In 11 of 15 analyzed MAC-amended enrichments, Clostridia were the most abundant bacterial class. Especially treatments containing vanillin, 3,4,5-trimethoxyphenol, 3,4,5-trimethoxybenzoic acid, 2-methoxyhydroquinone, 3,4-dimethoxycinnamic acid and 3,4,5-trimethoxycinnamic acid stimulated the growth of Clostridia shown in the high increase of 16S rRNA gene copies. Moreover, we could observe a decrease in the concentration of MACs throughout the incubation time in those enrichments with a high abundance of Clostridia, confirming the consumption of these substrates. Overall, our enrichment results indicate that MACs with multiple methoxy groups promoted the growth of Clostridia, which agrees with previously reported bond dissociation energies of methoxy groups in MACs. The bond dissociation energy of single methoxy groups is higher (246,86 kJ/mol) than for MACs containing two (232.21 kJ/mol) or three (217.99 kJ/mol) methoxy groups, making those MACs with multiple groups more energetically favorable (Li et al. 2021). Additional substituents to the ring can increase the bond dissociation energy required to cleave methoxy groups (Li et al. 2021), thus some MACs can become more difficult to degrade, as we observed for MACs, such as 3-methoxycinnamic acid, 2-methoxybenzoic acid, 3-methoxyphenylacetic acid or 2-methoxy-pyridinylboranic acid.

The family Alkalibacteraceae was detected in most enrichments (10 enrichments), with single ASVs being abundant in up to nine different treatments, indicating Alkalibacteraceae as the main degraders of methoxylated aromatic compounds and being adapted for multiple different MACs. A high relative abundance of ASVs of the order Acetivibrionales was detected in the enrichment amended with 3,4,5-trimethoxybenzoic acid, suggesting substrate specificity of Acetivibrionales for this substrate. Notably, the ASVs could not be classified below order level, similar to reconstructed Acetivibrionales MAGs and therefore must be regarded as a novel family. Further, Oscillospirales, Peptrostreptococcales, and Tissierellales were identified. Yet, they were of lower abundance in all screened treatments, hinting either towards competition for the MACs, with Eubacteriales and Acetivibrionales being advantageous or overtaking different roles.

Clostridia are common in anoxic soils (Janssen 2006), yet they could also be identified in some marine sediments (Davies 1969, Zinger et al. 2011). Their primary role in the marine biosphere was so far limited to the degradation of cellulose, proteins and amino acids (Davies 1969, Pelikan et al. 2021, Yu et al. 2023b). Here, we identified Clostridia as the major degraders of MACs, playing an essential role in the degradation of terrestrial-derived organic matter in anoxic marine sediments. These findings of Clostridia as primary degraders of MACs were previously reported for soil systems (Kato et al. 2015). Though more than 30 isolated Clostridia species could so far be affiliated with the growth on MACs (Khomyakova and Slobodkin 2023), only one of these, *Acetobacterium carbinolicum*, was isolated from marine anoxic sediments (Paarup et al. 2006). The high abundance of different Clostridia groups detected in this study indicates that Clostridia not only play an essential role in the degradation of terrestrial-derived organic matter in soil but also coastal anoxic sediments.

4.1.5.2 Intracellular methyltransferases facilitate methoxy group cleavage

By analyzing the genomic potential, we identified multiple novel families (Aceti_1-Aceti_5, Eubac_4) and genera (Eubac_1-Eubac_2, Oscillo_1-Oscillo_3, Tissi_1), in addition to the undescribed genus DUPK01, containing genes involved in methoxydotrophy. To facilitate methoxydotrophy, these taxa require an *O*-demethylase system, including the methyltransferase MTI. The annotation of the MTI was based on identifying genes through the annotation of NCBI NR (Sayers et al. 2022) and KEGG (Kanehisa et al. 2015). However, using these tools, the methyltransferase systems could not be determined. Related genes of the methyltransferase MTI were annotated as uroporphyrinogen decarboxylase instead of the known methoxytransferase genes *odmB* or *vdmB* found in *Acetobacterium dehalogenans*

(Kaufmann et al. 1998, Engelmann et al. 2001), *mtvB* found in *Moorella thermoacetica* (Naidu and Ragsdale 2001) or *mtoB* found in *Archaeoglobus fulgidus* and *Methermicoccus shengliensis* (Kurth et al. 2021, Welte et al. 2021), indicating lower sequence identities of found MTI genes to already characterized genes involved in methoxy group utilization. Even genes of the methyltransferases *vdmB* and *odmB* detected in the same genome of *Acetobacterium dehalogenans* only showed amino acid identities of 63% (Schilhabel et al. 2009). The high number of unique genes discovered in Clostridia MAGs (up to 62 unique MTI genes), especially of the orders Eubacteriales, Oscillospirales and Acetivibrionales, support this finding. Genomes with a high number of MTI genes (abundance > 40) did not show a proportional increase in diversity, implying that several copies of the same gene cluster are maintained in these genomes, potentially serving redundant roles in metabolic pathways. The presence of such high numbers of MTI genes in single genomes agrees with our findings of utilization of up to nine different MACs by Alkalibacteraceae and other groups within Clostridia. Furthermore, the high abundance of methyltransferase MTI indicates a general preference for C_1-compounds in members of the detected Clostridia beyond methoxy group utilization. Previous studies on MTI genes showed that *vdmB* isolated from *Acetobacterium dehalogenans* additionally converts methyl chloride and a methyltransferase identified in *Acetobacterium woodii* demethylates glycine betaine, coming to a similar conclusion (Schilhabel et al. 2009, Lechtenfeld et al. 2018). None of the detected MTI genes contained signal peptides (Owji et al. 2018), indicating no extracellular activity of MTI. Therefore, MACs need to be taken up by the organisms to utilize the methoxy group and microorganisms cannot act on methoxy groups of larger lignin polymers.

We identified functionally similar gene clusters by clustering the MTI genes with 70% identity. We could detect 107 gene clusters in multiple MAC treatments. Therefore, these genes likely do not act specific for a single MAC. But it also needs to be noted that previous studies identified MTI genes capable of acting on multiple MACs (Engelmann et al. 2001), so despite identifying substrate-specific methyltransferases, other methyltransferases detected in multiple samples could potentially be involved in the methoxy group cleavage of similar substrates.

Multiple MTI gene clusters could be detected in single samples, indicating possible unique or substrate-specific MTI genes. Acetivibrionales were highly abundant in the enrichment containing 3,4,5-trimethoxybenzoic acid; specifically, the MAG Aceti_3 showed a high relative abundance of 49%. We identified four gene clusters (60, 61, 62, 63) specific to this MAG. It is most likely that the identified sample-specific MTI genes act on methoxy groups of 3,4,5-trimethoxybenzoic acid. Furthermore, three unique gene clusters for the Eubac_1

MAG and two additional gene clusters for other found Acetivibrionales MAGs Aceti_2 and Aceti_4 were detected in the same orthogroup 179. Due to these genes deriving from a single last common ancestor (LCA), it is likely that the Eubac_1 MAG also contains MTI genes specific for 3,4,5-trimethoxybenzoic acid, which evolved orthologously to those of the Acetivibrionales. We identified further sample-specific MTI genes of the Eubac_1 MAG in the orthogroups 25 and 39, suggesting that this MAG might contain more MTI genes, specialized for 3,4,5-trimethoxybenzoic acid or other substrates unique for the genome Eubac_1.
In the same orthogroups, sample-specific MTI genes of the order Eubac_3 MAG for the MAC vanillin and sample-specific genes of the order Oscillospirales for the MACs 3,4,5-trimethoxycinnamic acid (Oscillo_1, Oscillo_3, Oscillo_4, Oscillo_5 and Oscillo_7) and 2-methoxyhydroquinone (Oscillo_6) were identified, indicating evolution of MTI genes towards different substrate-specificity. These findings are further supported by the clustering of MTI genes in the orthogroup trees 25 and 39, with Oscillospirales-derived MTI genes forming two diverging clusters. Nevertheless, metagenomic sequencing was only conducted on five of the fifteen treatments and thus, the results on substrate-specific methyltransferases can only be based on observations made from these enrichments. To prove the substrate-specificity actively involved methyltransferases need to be identified. As this study did not perform transcriptomic analyses, we can only hypothesize that some of the identified methyltransferases act specifically on the methoxy groups of the used substrates.

4.1.5.3 Methoxydotrophy via the reductive acetyl-CoA pathway as the primary strategy of multiple novel groups within the Clostridia to degrade methoxy groups of MACs

After the transfer of the methyl group, derived from the methoxy group of a MAC via the methyltransferase MTI onto a corrinoid protein and further onto the methyl group acceptor tetrahydrofolate (THF), the resulting methyl-THF would be further degraded to acetate (Kremp et al. 2018). Hereby, methyl-THF is combined with CO_2 derived from the carbonyl branch of the reductive acetyl-CoA pathway to form acetate (Kremp et al. 2018). In this study, we did not measure the acetate concentrations in the enrichments to prove acetogenic activity. However, results from the RNA-SIP enrichments showed that the Clostridia were only found in SIP treatments containing a combination of a MAC and DIC and were absent from DIC-only controls. Furthermore, these results indicate that the detected Alkalibacteraceae are heterotrophic and require both the methoxy group of a MAC and DIC for growth, most likely utilizing a reductive acetyl-CoA pathway, leading to acetate formation. Thus, these results could explain why the relative abundance of Alkalibacteraceae was shifted towards the

midpoint fractions in treatments in which only the MAC or DIC was labeled (mixed labeling) but showed the highest abundances in heavy fractions with both substrates labeled.

Apart from the MAG retrieved from the SIP study, we detected that also all other MAGs of the order Eubacteriales contained the entire gene repertoire to facilitate the incorporation of the methoxy groups via an acetogenic methoxydotrophic pathway and likely act similarly to the Alkalibacteraceae detected in the RNA-SIP study. In previous studies, multiple members of the Eubacteriales were proven to utilize MACs by coupling the degradation of methoxy groups to the electron acceptor CO_2 (Khomyakova and Slobodkin 2023). Among known Eubacteriales are *Acetobacterium carbinolicum*, the only isolate from anoxic marine sediments that was initially found in freshwater sediments (Eichler and Schink 1984, Paarup et al. 2006), *Acetobacterium dehalogenans* (Kaufmann et al. 1998), *Acetobacterium malicum* (Tanaka and Pfennig 1988), *Acetobacterium woodii* (Bache and Pfennig 1981), *Alkalibacter mobilis* (Khomyakova et al. 2021), *Alkalibaculum bacchi* (Allen et al. 2010), *Alkalibaculum sporogenes* (Khomyakova et al. 2020), *Eubacterium limosum* (DeWeerd et al. 1988) and *Eubacterium callanderi* (Mountfort et al. 1988). We present the first marine genus within the family of Alkalibacteraceae (Eubac_1 and Eubac_2) and a second novel family, represented by the MAG Eubac_4, engaged in methoxy group utilization.

Acetivibrionales showed equal potential for acetogenic methoxydotrophy. ASVs of the order Acetivibrionales were closest related to the sugar-degrading species *Clostridium thermosuccinogenes* (Koendjbiharie et al. 2018). Other genera of Acetivibrionales, such as the *Acetivibrio* (Patel et al. 1980, Shiratori et al. 2009), *Herbivorax* (Zhilina et al. 2005, Koeck et al. 2016), *Hungateiclostridium* (Kato et al. 2004, Rettenmaier et al. 2019a, Ha-Tran et al. 2021) and *Ruminiclostridium* (Hungate 1944, Madden et al. 1982, Sukhumavasi et al. 1988, Hethener et al. 1992, Monserrate et al. 2001, Desvaux 2005, Nishiyama et al. 2009, Rettenmaier et al. 2019b) were shown to be involved in the degradation of cellulose but no other components of lignocellulose. Therefore, the detected novel family represents the first Acetivibrionales family with shown metabolic potential for the degradation of MACs. Yet, high gene copy numbers of the MTI genes (up to 40 gene copies) could also be detected in other genomes of the screened Acetivibrionales species representatives, indicating that further members of the order Acetivibrionales could be growing on MACs.

Among Oscillospirales, only MAGs assigned to the yet unclassified Oscillospiraceae, *Sporobacter*, and DUPK01 had genes involved in methoxy group utilization, which is consistent with our findings that within Oscillospirales species representatives only selected families within the Oscillospirales contained MTI genes. In the MAC enrichments, we mostly

detected unclassified Oscillospiraceae ASVs as having high abundance (19-48%). The MAGs Oscillo_2 and Oscillo_3, affiliated with yet unclassified Oscillospiraceae, contained all genes for methoxydotrophic acetogenesis. At the same time, other Oscillospirales lacked at least the formate dehydrogenase gene, suggesting that mainly the genomes of Oscillo_2 and Oscillo_3 were involved in the degradation of MACs via the reductive acetyl-CoA pathway. Our annotation also detected that only one of two *Sporobacter* MAGs contained all but one gene of the reductive-acetyl-CoA pathway. A second *Sporobacter* MAG lacked most genes despite both MAGs having high gene copy numbers of the MTI gene. The closest relative S*porobacter termitidis* isolated from the digestive tract of the wood-feeding termite *Nasutitermes lujae* was found to grow exclusively on MACs, such as 3,4,5-trimethoxycinnamic acid, 3,4-dimethoxycinnamic acid, 3,4,5-trimethoxybenzoic acid, sinapate, vanillate, ferulate and syringate (Grech-Mora et al. 1996). However, *Sporobacter termitidis* requires sulfide or cysteine as a methyl acceptor to utilize methoxy groups. Therefore, missing genes in *Sporobacter* MAGs for the reductive acetyl-CoA pathway could result from this genus utilizing a different methoxy group utilization strategy, rather than missing genes due to incompleteness of the MAGs. A lack or limited amount of required methyl group acceptors could explain the lower relative abundance of *Sporobacter*-assigned ASVs in the enrichments. Accordingly, other Oscillospirales MAGs lacking subunits of the reductive acetyl-CoA pathway might use a different strategy to utilize methoxy groups other than the reductive acetyl-CoA pathway.

In the order Tissierellales, ASVs with the highest abundance were closest related to the genus *Dethiosulfatibacter*. The genus *Dethiosulfatibacter* couples the oxidation of amino and organic acids to the reduction of thiosulfate or sulfur and is likely not involved in the degradation of MACs (Takii et al. 2007). Based on our quality filtering of MAGs, we could not retrieve good-quality MAGs of this genus. However, we could find an additional MAG of low abundance assigned to a novel genus within the family Sedimentibacteraceae. The type strain *Sedimentibacter hydroxybenzoicus,* initially isolated from freshwater ponds, is known for its ability to degrade para-hydroxybenzoates and is therefore involved in the degradation of hydroxylated aromatic compounds (Zhang et al. 1994). Apart from lacking the anaerobic acetyl-CoA synthetase (*acsA*) gene, the Sedimentibacteraceae MAG Tissi_1 had all genes required to utilize methoxy groups and might, therefore, be involved in the process of MAC degradation. As all other genes of the reductive acetyl-CoA pathway are present, the missing gene could result from the incompleteness of the MAG.

Peptostreptococcales, Ch29 and unclassified Clostridia MAGs were not found to be involved in methoxydotrophic acetogenesis based on the lack of MTI genes and the lack of the acetyl-

CoA decarbonylase/synthase, CODH/ACS complex. Supporting this finding, MAGs of these groups could only be identified in relative abundances up to 1.54% and instead play a minor role in the samples. Multiple families of the Peptostreptococcales are known to degrade amino acids (Zhilina et al. 1998, Meijer et al. 1999, Alain et al. 2002) and carbohydrates (Ravot et al. 1999, Alain et al. 2002, Smii et al. 2015, Qiu et al. 2021), consistent with our finding that methoxy groups could not be metabolized. As we could not identify Peptostreptococcales in control samples and they became abundant in only selected MAC enrichments, it is most likely that Peptostreptococcales act on substrates produced by primary fermentation of some of the MACs. Ch29 and unclassified Clostridia were not detected in any of the enrichment experiments. Further analyses of their metabolic potential are required to investigate their potential roles in the environment.

In order to perform redox balancing during methoxydotrophic growth, one methoxy group must be oxidized to CO_2 in the reverse methyl branch of the Wood Ljungdahl pathway (Kremp et al. 2018). We identified the production of $\delta^{13}C$ CO_2 in SIP enrichments with high abundances of Clostridia, proving that CO_2 is formed from the labeled methoxy group and thus suggesting that Clostridia use a similar strategy. Genes for the Rnf complex, V/A type ATPase, F-type ATPase and the Fe-Fe hydrogenase *hydABC* could be identified in most MAGs and might be strongly involved in conserving energy, as previously suggested by Kremp et al. (2018). Though we could not identify all subunits of some of these complexes, most were present, and missing genes are likely due to MAG incompleteness.

4.1.5.4 Clostridia are restricted to the utilization of methoxy groups

The final cleavage of aromatic ring structures proceeds via three intermediates: benzoyl-CoA, resorcinol (1,3-dihydroxybenzene) and phloroglucinol (1,3,5-trihydroxybenzene) (Schink et al. 2000). While we observed multiple orders in the Clostridia as the main abundant bacteria performing methoxydotrophic acetogenesis, none of the genomes contained genes involved in the degradation of the remaining aromatic ring structures via these intermediates based on their genomic potential. However, some Clostridia of the order Oscillospirales (Oscillo_2, Oscillo_4, Oscillo_5) could potentially facilitate the transformation of 3,4,5-trihydroxybenzoate to 1,2,3-trihydroxybenzene and further to 1,3,5-trihydroxybenzene via gallate decarboxylase and pyrogallol hydroxytransferase as preparatory steps for other organisms involved in aromatic ring cleavage of phloroglucinol. Only a few Clostridia have been characterized to be involved in the aromatic ring degradation via the phloroglucinol pathway (Grech-Mora et al. 1996, Mechichi et al. 1999, Lomans et al. 2001). Among these are

Sporobacter termitidis (Grech-Mora et al. 1996), *Sporobacterium olearium* and *Parasporobacterium paucivorans*. Based on these previous findings, the MAGs Oscillo_4 and Oscillo_5 assigned to the genus *Sporobacter* might also be involved in this process. However, genes involved in the anaerobic phloroglucinol degradation have only recently been characterized (Zhou et al. 2023). Likely, the annotation for involved genes is still missing from the used databases, and further screening of genes using HMMs could lead to the identification of the complete phloroglucinol pathway in some MAGs.

Other MAGs potentially involved in aromatic ring degradation were of minor abundance in the samples, only reaching a maximum relative abundance of 0.78%, indicating a minor role of these taxa in the enrichments. The abundance of these taxa based on the 16S rRNA gene relative abundance in all enrichments was equally low, supporting that aromatic ring degradation was of subordinate importance and methoxydotrophy the prevalent activity during the incubation period of 458 days.

4.1.6 Conclusion

In summary, our study identified Clostridia as the primary degraders of methoxylated aromatic compounds (MACs) in marine anoxic sediments. We found that Clostridia, particularly members of the Alkalibacteraceae, were well-adapted to degrade a variety of MACs. A novel family within the Acetivibrionales showed substrate-specificity towards 3,4,5-trimethoxy-benzoic acid. The high abundance and dominance of Clostridia in the process of MAC degradation in coastal anoxic sediments suggest they play a crucial role in the turnover of terrestrial organic matter in these environments and contribute to the carbon cycle in anoxic conditions. Our findings of the presence and diversity of MTI genes in Clostridia indicate a widespread metabolic strategy of methoxydotrophic acetogenesis across this class with some members being capable of utilizing a wide spectrum of MACs and some being more restricted. However, while found Clostridia are capable of methoxy group utilization, they lack the metabolic potential to cleave aromatic ring structures, suggesting an interplay of Clostridia and other microbial taxa to degrade MACs completely.

This study broadens the understanding of terrestrial organic matter degradation in anoxic marine sediments and identifies new bacterial families involved in this process. Further research on the interplay between Clostridia and aromatic-ring cleaving organisms could deepen insights into the final degradation processes of monolignols in the anoxic marine biosphere.

4.1.7 Funding statement

DFG supported this study under Germany's excellence Strategy, no. EXC- 2077- 390741603.

4.1.8 Author contributions

MDM, XY, and MWF conceived this study. MDM and XY performed the slurry enrichments and SIP enrichments. MDM, CR and LB performed gas measurements. MDM, JW and ME performed the GC-IRMS measurements. MDM, CR and LB performed sample preparation for 16S rRNA gene sequencing and measurements of MAC concentration. JT and TH performed LC-MS/MS measurements of MAC concentrations. MDM and LCW analyzed the 16S rRNA gene sequencing data. MDM performed bioinformatic analyses. CH and TRH provided support for bioinformatic and statistical analyses. XY, MWF and CH offered constructive feedback and guided the execution of the project. MWF secured the funding for this project. MDM prepared all figures and tables. MDM drafted the first manuscript version. MDM, XY, and MWF edited the manuscript.

4.1.9 References

Alain, K., Pignet, P., Zbinden, M., Quillevere, M., Duchiron, F., Donval, J. P., Lesongeur, F., Raguenes, G., Crassous, P., Querellou, J., et al. (2002). *Caminicella sporogenes* gen. nov., sp. nov., a novel thermophilic spore-forming bacterium isolated from an East-Pacific Rise hydrothermal vent. Int. J. Syst. Evol. Microbiol. **52**:1621-1628. doi: 10.1099/00207713-52-5-1621.

Allen, T. D., Caldwell, M. E., Lawson, P. A., Huhnke, R. L. and Tanner, R. S. (2010). *Alkalibaculum bacchi* gen. nov., sp. nov., a CO-oxidizing, ethanol-producing acetogen isolated from livestock-impacted soil. Int. J. Syst. Evol. Microbiol. **60**:2483-2489. doi: 10.1099/ijs.0.018507-0.

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F. and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. Nat. Methods **11**:1144-1146. doi: 10.1038/nmeth.3103.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. **215**:403-410. doi: 10.1016/s0022-2836(05)80360-2.

Anthony, C. (1982). Chapter 1. Methylotrophic bacteria. *In*: The biochemistry of methylotrophs. Academic Press, London. p. 1-41.

Bache, R. and Pfennig, N. (1981). Selective isolation of *Acetobacterium woodii* on methoxylated aromatic acids and determination of growth yields. Arch. Microbiol. **130**:255-261. doi: 10.1007/BF00459530.

Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T. and Stamatakis, A. (2019). EPA-ng: massively parallel evolutionary placement of genetic sequences. Syst. Biol **68**:365-369. doi: 10.1093/sysbio/syy054.

Bianchi, T. S. (2011). The role of terrestrially derived organic carbon in the coastal ocean: a changing paradigm and the priming effect. Proc. Natl. Acad. Sci. U. S. A. **108**:19473-19481. doi: 10.1073/pnas.1017982108.

Boerjan, W., Ralph, J. and Baucher, M. (2003). Lignin biosynthesis. Annu. Rev. Plant Biol. **54**:519-546. doi: 10.1146/annurev.arplant.54.031902.134938.

Brune, A. and Schink, B. (1990). Pyrogallol-to-phloroglucinol conversion and other hydroxyl-transfer reactions catalyzed by cell extracts of *Pelobacter acidigallici*. J. Bacteriol. **172**:1070-1076. doi: 10.1128/jb.172.2.1070-1076.1990.

Buchfink, B., Reuter, K. and Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods **18**:366-368. doi: 10.1038/s41592-021-01101-x.

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

Burdige, D. J. (2005). Burial of terrestrial organic matter in marine sediments: a re-assessment. Global Biogeochem. Cycles **19**. doi: 10.1029/2004GB002368.

Bushnell, B. (2014). BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Laboratory LBNL Report #: LBNL-7065E.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods **13**:581-583. doi: 10.1038/nmeth.3869.

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. and Parks, D. H. (2022). GTDB-Tk v2: memory friendly classification with the Genome Taxonomy Database. PREPRINT bioRxiv. doi: 10.1101/2022.07.11.499641.

Chen, J. X., Deng, C. Y., Zhang, Y. T., Liu, Z. M., Wang, P. Z., Liu, S. L., Qian, W. and Yang, D. H. (2016). Cloning, expression, and characterization of a four-component *O*-demethylase from human intestinal bacterium *Eubacterium limosum* ZL-II. Appl. Microbiol. Biotechnol. **100**:9111-9124. doi: 10.1007/s00253-016-7626-1.

Chklovski, A., Parks, D. H., Woodcroft, B. J. and Tyson, G. W. (2022). CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. PRERPRINT bioRxiv. doi: 10.1101/2022.07.11.499243.

Crawford, D. L. and Crawford, R. L. (1980). Microbial degradation of lignin. Enzyme Microb. Technol. **2**:11-22. doi: 10.1016/0141-0229(80)90003-4.

Czech, L., Barbera, P. and Stamatakis, A. (2020). Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. Bioinformatics **36**:3263-3265. doi: 10.1093/bioinformatics/btaa070.

Daims, H., Brühl, A., Amann, R., Schleifer, K. H. and Wagner, M. (1999). 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. doi: 10.1016/s0723-2020(99)80053-8.

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B. and Flouri, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol. Biol. Evol. **37**:291-294. doi: 10.1093/molbev/msz189.

Davies, J. A. (1969). Isolation and identification of Clostridia from North Sea sediments. J. Appl. Bacteriol. **32**:164-169. doi: 10.1111/j.1365-2672.1969.tb00962.x.

Desvaux, M. (2005). *Clostridium cellulolyticum*: model organism of mesophilic cellulolytic Clostridia. FEMS Microbiol. Rev. **29**:741-764. doi: 10.1016/j.femsre.2004.11.003.

DeWeerd, K. A., Saxena, A., Nagle, D. P., Jr. and Suflita, J. M. (1988). Metabolism of the ^{18}O-methoxy substituent of 3-methoxybenzoic acid and other unlabeled methoxybenzoic acids by anaerobic bacteria. Appl. Environ. Microbiol. **54**:1237-1242. doi: 10.1128/aem.54.5.1237-1242.1988.

Dittmar, T. and Lara, R. J. (2001). Molecular evidence for lignin degradation in sulfate-reducing mangrove sediments (Amazônia, Brazil). Geochim. Cosmochim. Acta **65**:1417-1428. doi: 10.1016/S0016-7037(00)00619-0.

Drake, H. L. and Daniel, S. L. (2004). Physiology of the thermophilic acetogen *Moorella thermoacetica*. Res. Microbiol. **155**:869-883. doi: 10.1016/j.resmic.2004.10.002.

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. Genome Inform. **23**:205-211.

Edgar, R. C. (2022). Muscle5: high-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat. Commun. **13**:6968. doi: 10.1038/s41467-022-34630-w.

Eichler, B. and Schink, B. (1984). Oxidation of primary aliphatic alcohols by *Acetobacterium carbinolicum* sp. nov., a homoacetogenic anaerobe. Arch. Microbiol. **140**:147-152. doi: 10.1007/BF00454917.

Emms, D. M. and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**:238. doi: 10.1186/s13059-019-1832-y.

Engelmann, T., Kaufmann, F. and Diekert, G. (2001). Isolation and characterization of a veratrol:corrinoid protein methyl transferase from *Acetobacterium dehalogenans*. Arch. Microbiol. **175**:376-383. doi: 10.1007/s002030100275.

Eren, A. M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., Fink, I., Pan, J. N., Yousef, M., Fogarty, E. C., et al. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. Nat. Microbiol. **6**:3-6. doi: 10.1038/s41564-020-00834-3.

Ertefai, T. F., Heuer, V. B., Prieto-Mollar, X., Vogt, C., Sylva, S. P., Seewald, J. and Hinrichs, K.-U. (2010). The biogeochemistry of sorbed methane in marine sediments. Geochim. Cosmochim. Acta **74**:6033-6048. doi: 10.1016/j.gca.2010.08.006.

Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. Syst. Biol. **19**:99-113. doi: 10.2307/2412448.

Gargulak, J. D., Lebo, S. E. and McNally, T. J. (2015). Lignin. *In*: Kirk-Othmer Encyclopedia of Chemical Technology. John Wiley & Sons, Inc. p. 1-26. doi: 10.1002/0471238961. 12090714120914.a01.pub3.

Grant, D. J. W. and Patel, J. C. (1969). The non-oxidative decarboxylation ofp-hydroxybenzoic acid, gentisic acid, protocatechuic acid and gallic acid by *Klebsiella aerogenes* (*Aerobacter aerogenes*). Antonie van Leeuwenhoek **35**:325-343. doi: 10.1007/BF02219153.

Grech-Mora, I., Fardeau, M.-L., Patel, B. K. C., Ollivier, B., Rimbault, A., Prensier, G., Garcia, J.-L. and Garnier-Sillam, E. (1996). Isolation and characterization of *Sporobacter termitidis* gen. nov., sp. nov., from the digestive tract of the wood-feeding termite *Nasutitermes lujae*. Int. J. Syst. Evol. Microbiol. **46**:512-518. doi: 10.1099/00207713-46-2-512.

Ha-Tran, D. M., Nguyen, T. T. M., Lo, S. C. and Huang, C. C. (2021). Utilization of monosaccharides by *Hungateiclostridium thermocellum* ATCC 27405 through adaptive evolution. Microorganisms **9**. doi: 10.3390/microorganisms9071445.

Harms, U. and Thauer, R. K. (1996). Methylcobalamin:coenzyme M methyltransferase isoenzymes MtaA and MtbA from *Methanosarcina barkeri*. Cloning, sequencing and differential transcription of the encoding genes, and functional overexpression of the *mtaA* gene in *Escherichia coli*. Eur. J. Biochem. **235**:653-659. doi: 10.1111/j.1432-1033.1996.00653.x.

Herlemann, D. P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J. and Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. **5**:1571-1579. doi: 10.1038/ismej.2011.41.

Hethener, P., Brauman, A. and Garcia, J.-L. (1992). *Clostridium termitidis* sp. nov., a cellulolytic bacterium from the gut of the wood-feeding termite, *Nasutitermes lujae*. Syst. Appl. Microbiol. **15**:52-58. doi: 10.1016/S0723-2020(11)80138-4.

Hsieh, T. C., Ma, K. H. and Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol. Evol. **7**:1451-1456. doi: 10.1111/2041-210X.12613.

Hungate, R. E. (1944). Studies on cellulose fermentation: I. The culture and physiology of an anaerobic cellulose-digesting bacterium. J. Bacteriol. **48**:499-513. doi: 10.1128/jb.48.5.499-513.1944.

Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W. and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. **11**:119. doi: 10.1186/1471-2105-11-119.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. **9**:5114. doi: 10.1038/s41467-018-07641-9.

Janssen, P. H. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl. Environ. Microbiol. **72**:1719-1728. doi: 10.1128/AEM.72.3.1719-1728.2006.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods **14**:587-589. doi: 10.1038/nmeth.4285.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. **44**:D457-D462. doi: 10.1093/nar/gkv1070.

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H. and Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ **7**:e7359. doi: 10.7717/peerj.7359.

Kato, S., Chino, K., Kamimura, N., Masai, E., Yumoto, I. and Kamagata, Y. (2015). Methanogenic degradation of lignin-derived monoaromatic compounds by microbial enrichments from rice paddy field soil. Sci. Rep. **5**:14295. doi: 10.1038/srep14295.

Kato, S., Haruta, S., Cui, Z. J., Ishii, M., Yokota, A. and Igarashi, Y. (2004). *Clostridium straminisolvens* sp. nov., a moderately thermophilic, aerotolerant and cellulolytic bacterium isolated from a cellulose-degrading bacterial community. Int. J. Syst. Evol. Microbiol. **54**:2043-2047. doi: 10.1099/ijs.0.63148-0.

Katsyv, A., Kumar, A., Saura, P., Pöverlein, M. C., Freibert, S. A., T. Stripp, S., Jain, S., Gamiz-Hernandez, A. P., Kaila, V. R. I., Müller, V., et al. (2023). Molecular basis of the electron bifurcation mechanism in the [FeFe]-hydrogenase complex HydABC. J. Am. Chem. Soc. **145**:5696-5709. doi: 10.1021/jacs.2c11683.

Kaufmann, F., Wohlfarth, G. and Diekert, G. (1998). *O*-demethylase from *Acetobacterium dehalogenans*--substrate specificity and function of the participating proteins. Eur. J. Biochem. **253**:706-711. doi: 10.1046/j.1432-1327.1998.2530706.x.

Khomyakova, M., Merkel, A., Novikov, A., Klyukina, A. and Slobodkin, A. (2021). *Alkalibacter mobilis* sp. nov., an anaerobic bacterium isolated from a coastal lake. Int. J. Syst. Evol. Microbiol. **71**. doi: 10.1099/ijsem.0.005174.

Khomyakova, M. and Slobodkin, A. I. (2023). Transformation of methoxylated aromatic compounds by anaerobic microorganisms. Microbiol **92**:97-118. doi: 10.1134/S0026261722603128.

Khomyakova, M. A., Merkel, A. Y., Petrova, D. A., Bonch-Osmolovskaya, E. A. and Slobodkin, A. I. (2020). *Alkalibaculum sporogenes* sp. nov., isolated from a terrestrial mud volcano and emended description of the genus *Alkalibaculum*. Int. J. Syst. Evol. Microbiol. **70**:4914-4919. doi: 10.1099/ijsem.0.004361.

Koeck, D. E., Mechelke, M., Zverlov, V. V., Liebl, W. and Schwarz, W. H. (2016). *Herbivorax saccincola* gen. nov., sp. nov., a cellulolytic, anaerobic, thermophilic bacterium isolated via in sacco enrichments from a lab-scale biogas reactor. Int. J. Syst. Evol. Microbiol. **66**:4458-4463. doi: 10.1099/ijsem.0.001374.

Koendjbiharie, J. G., Wiersma, K. and van Kranenburg, R. (2018). Investigating the central metabolism of *Clostridium thermosuccinogenes*. Appl. Environ. Microbiol. **84**. doi: 10.1128/aem.00363-18.

Konstantinidis, K., Ruiz-Perez, C., Gerhardt, K., Rodriguez-R, L., Jain, C., Tiedje, J. and Cole, J. (2022). FastAAI: Efficient estimation of genome average amino acid identity and phylum-level relationships using tetramers of universal proteins. PREPRINT Research Square. doi: 10.21203/rs.3.rs-1459378/v1.

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics **35**:4453-4455. doi: 10.1093/bioinformatics/btz305.

Kremp, F., Poehlein, A., Daniel, R. and Müller, V. (2018). Methanol metabolism in the acetogenic bacterium *Acetobacterium woodii*. Environ. Microbiol. **20**:4369-4384. doi: 10.1111/1462-2920.14356.

Kurth, J. M., Nobu, M. K., Tamaki, H., de Jonge, N., Berger, S., Jetten, M. S. M., Yamamoto, K., Mayumi, D., Sakata, S., Bai, L., et al. (2021). Methanogenic archaea use a bacteria-like methyltransferase system to demethoxylate aromatic compounds. ISME J. **15**:3549-3565. doi: 10.1038/s41396-021-01025-6.

Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J. and Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput. Biol. **11**:e1004226. doi: 10.1371/journal.pcbi.1004226.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**:357-359. doi: 10.1038/nmeth.1923.

Lechtenfeld, M., Heine, J., Sameith, J., Kremp, F. and Müller, V. (2018). Glycine betaine metabolism in the acetogenic bacterium *Acetobacterium woodii*. Environ. Microbiol. **20**:4512-4525. doi: 10.1111/1462-2920.14389.

Letunic, I. and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics **23**:127-128. doi: 10.1093/bioinformatics/btl529.

Ley, Y., Cheng, X.-Y., Ying, Z.-Y., Zhou, N.-Y. and Xu, Y. (2023). Characterization of two marine lignin-degrading consortia and the potential microbial lignin degradation network in nearshore regions. Microbiol. Spectr. **11**:e04424-04422. doi:10.1128/spectrum.04424-22.

Li, D., Liu, C. M., Luo, R., Sadakane, K. and Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics **31**:1674-1676. doi: 10.1093/bioinformatics/btv033.

Li, J.-L., Duan, L., Wu, Y., Ahmad, M., Yin, L.-Z., Luo, X.-Q., Wang, X., Fang, B.-Z., Li, S.-H., Huang, L.-N., et al. (2022). Unraveling microbe-mediated degradation of lignin and lignin-derived aromatic fragments in the Pearl River Estuary sediments. Chemosphere **296**:133995. doi: 10.1016/j.chemosphere.2022.133995.

Li, R., Du, T., Liu, J., Aquino, A. J. A. and Zhang, J. (2021). Theoretical study of O-CH(3) bond dissociation enthalpy in Anisole systems. ACS Omega **6**:21952-21959. doi: 10.1021/acsomega.1c02310.

Liao, Y., Smyth, G. K. and Shi, W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30**:923-930. doi: 10.1093/bioinformatics/btt656.

Lin, D.-D., Liu, Y.-F., Zhou, L., Yang, S.-Z., Gu, J.-D. and Mu, B.-Z. (2022). Stimulation of Bathyarchaeota in enrichment cultures by syringaldehyde, 4-hydroxybenzaldehyde and vanillin under anaerobic conditions. Int. Biodeter. Biodegr. **171**:105409. doi: 10.1016/j.ibiod.2022.105409.

Liu, H., Wang, J., Wang, A. and Chen, J. (2011). Chemical inhibitors of methanogenesis and putative applications. Appl. Microbiol. Biotechnol. **89**:1333-1340. doi: 10.1007/s00253-010-3066-5.

Lomans, B. P., Leijdekkers, P., Wesselink, J. J., Bakkes, P., Pol, A., van der Drift, C. and den Camp, H. J. (2001). Obligate sulfide-dependent degradation of methoxylated aromatic compounds and formation of methanethiol and dimethyl sulfide by a freshwater sediment isolate, *Parasporobacterium paucivorans* gen. nov., sp. nov. Appl. Environ. Microbiol. **67**:4017-4023. doi: 10.1128/aem.67.9.4017-4023.2001.

Lueders, T., Manefield, M. and Friedrich, M. W. (2004). Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. Environ. Microbiol. **6**:73-78. doi: 10.1046/j.1462-2920.2003.00536.x.

Madden, R. H., Bryder, M. J. and Poole, N. J. (1982). Isolation and characterization of an anaerobic, cellulolytic bacterium, *Clostridium papyrosolvens* sp. nov. Int. J. Syst. Evol. Microbiol. **32**:87-91. doi: 10.1099/00207713-32-1-87.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17**:3. doi: 10.14806/ej.17.1.200.

Mechichi, T., Labat, M., Garcia, J. L., Thomas, P. and Patel, B. K. (1999). *Sporobacterium olearium* gen. nov., sp. nov., a new methanethiol-producing bacterium that degrades aromatic compounds, isolated from an olive mill wastewater treatment digester. Int. J. Syst. Bacteriol. **49 Pt 4**:1741-1748. doi: 10.1099/00207713-49-4-1741.

Meijer, W. G., Nienhuis-Kuiper, M. E. and Hansen, T. A. (1999). Fermentative bacteria from estuarine mud: phylogenetic position of *Acidaminobacter hydrogenoformans* and description of a new type of gram-negative, propionigenic bacterium as *Propionibacter pelophilus* gen. nov., sp. nov. Int. J. Syst. Bacteriol. **49 Pt 3**:1039-1044. doi: 10.1099/00207713-49-3-1039.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A. and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. **37**:1530-1534. doi: 10.1093/molbev/msaa015.

Monserrate, E., Leschine, S. B. and Canale-Parola, E. (2001). *Clostridium hungatei* sp. nov., a mesophilic, N_2-fixing cellulolytic bacterium isolated from soil. Int. J. Syst. Evol. Microbiol. **51**:123-132. doi: 10.1099/00207713-51-1-123.

Mountfort, D. O., Grant, W. D., Clarke, R. and Asher, R. A. (1988). *Eubacterium callanderi* sp. nov. that demethoxylates *O*-methoxylated aromatic acids to volatile fatty acids. Int. J. Syst. Evol. Microbiol. **38**:254-258. doi: 10.1099/00207713-38-3-254.

Naidu, D. and Ragsdale, S. W. (2001). Characterization of a three-component vanillate *O*-demethylase from *Moorella thermoacetica*. J. Bacteriol. **183**:3276-3281. doi: 10.1128/jb.183.11.3276-3281.2001.

Nishiyama, T., Ueki, A., Kaku, N. and Ueki, K. (2009). *Clostridium sufflavum* sp. nov., isolated from a methanogenic reactor treating cattle waste. Int. J. Syst. Evol. Microbiol. **59**:981-986. doi: 10.1099/ijs.0.001719-0.

Nissen, J. N., Johansen, J., Allesøe, R. L., Sønderby, C. K., Armenteros, J. J. A., Grønbech, C. H., Jensen, L. J., Nielsen, H. B., Petersen, T. N., Winther, O., et al. (2021). Improved metagenome binning and assembly using deep variational autoencoders. Nat. Biotechnol. **39**:555-560. doi: 10.1038/s41587-020-00777-4.

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P. and Stevens, M. H. H. (2024). vegan: community ecology package. https://vegandevs.github.io/vegan/.

Olm, M. R., Brown, C. T., Brooks, B. and Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. **11**:2864-2868. doi: 10.1038/ismej.2017.126.

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S. and Phillippy, A. M. (2016). Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. **17**:132. doi: 10.1186/s13059-016-0997-x.

Ovreås, L., Forney, L., Daae, F. L. and Torsvik, V. (1997). Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. Appl. Environ. Microbiol. **63**:3367-3373. doi: 10.1128/aem.63.9.3367-3373.1997.

Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A. and Ghasemi, Y. (2018). A comprehensive review of signal peptides: structure, roles, and applications. Eur. J. Cell Biol. **97**:422-441. doi: 10.1016/j.ejcb.2018.06.003.

Paarup, M., Friedrich, M. W., Tindall, B. J. and Finster, K. (2006). Characterization of the psychrotolerant acetogen strain SyrA5 and the emended description of the species *Acetobacterium carbinolicum*. Antonie van Leeuwenhoek **89**:55-69. doi: 10.1007/s10482-005-9009-y.

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

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A. and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. **36**:996-1004. doi: 10.1038/nbt.4229.

Patel, G. B., Khan, A. W., Agnew, B. J. and Colvin, J. R. (1980). Isolation and characterization of an anaerobic, cellulolytic microorganism, *Acetivibrio cellulolyticus* gen. nov., sp. nov.†. Int. J. Syst. Evol. Microbiol. **30**:179-185. doi: 10.1099/00207713-30-1-179.

Pelikan, C., Wasmund, K., Glombitza, C., Hausmann, B., Herbold, C. W., Flieder, M. and Loy, A. (2021). Anaerobic bacterial degradation of protein and lipid macromolecules in subarctic marine sediment. ISME J. **15**:833-847. doi: 10.1038/s41396-020-00817-6.

Peng, Q., Lin, L., Tu, Q., Wang, X., Zhou, Y., Chen, J., Jiao, N. and Zhou, J. (2023). Unraveling the roles of coastal bacterial consortia in degradation of various lignocellulosic substrates. mSystems **8**:e01283-01222. doi: 10.1128/msystems.01283-22.

Price, M. N., Dehal, P. S. and Arkin, A. P. (2010). FastTree 2: approximately maximum-likelihood trees for large alignments. PLoS One **5**:e9490. doi: 10.1371/journal.pone.0009490.

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. and Korobeynikov, A. (2020). Using SPAdes de novo assembler. Curr. Protoc. Bioinformatics **70**:e102. doi: 10.1002/cpbi.102.

Pruesse, E., Peplies, J. and Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics **28**:1823-1829. doi: 10.1093/bioinformatics/bts252.

Qiu, D., Zeng, X., Zeng, L., Li, G. and Shao, Z. (2021). *Fusibacter ferrireducens* sp. nov., an anaerobic, Fe(III)- and sulphur-reducing bacterium isolated from mangrove sediment. Int. J. Syst. Evol. Microbiol. **71**. doi: 10.1099/ijsem.0.004952.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. **41**:D590-D596. doi: 10.1093/nar/gks1219.

R Core Team. (2023). R: a language and environment for statistical computing. https://www.R-project.org/.

Ragsdale, S. W. and Pierce, E. (2008). Acetogenesis and the Wood–Ljungdahl pathway of CO_2 fixation. Biochim. Biophys. Acta, Proteins Proteomics **1784**:1873-1898. doi: 10.1016/j.bbapap.2008.08.012.

Ravot, G., Magot, M., Fardeau, M. L., Patel, B. K., Thomas, P., Garcia, J. L. and Ollivier, B. (1999). *Fusibacter paucivorans* gen. nov., sp. nov., an anaerobic, thiosulfate-reducing bacterium from an oil-producing well. Int. J. Syst. Bacteriol. **49 Pt 3**:1141-1147. doi: 10.1099/00207713-49-3-1141.

Rettenmaier, R., Gerbaulet, M., Liebl, W. and Zverlov, V. V. (2019a). *Hungateiclostridium mesophilum* sp. nov., a mesophilic, cellulolytic and spore-forming bacterium isolated from a biogas fermenter fed with maize silage. Int. J. Syst. Evol. Microbiol. **69**:3567-3573. doi: 10.1099/ijsem.0.003663.

Rettenmaier, R., Kowollik, M. L., Klingl, A., Liebl, W. and Zverlov, V. (2019b). *Ruminiclostridium herbifermentans* sp. nov., a mesophilic and moderately thermophilic cellulolytic and xylanolytic bacterium isolated from a lab-scale biogas fermenter fed with maize silage. Int. J. Syst. Evol. Microbiol. **71**. doi: 10.1099/ijsem.0.004692.

Salazar, G. (2020). https://github.com/benjjneb/dada2/issues/938#issuecomment-589657164. (accessed 15. January 2024).

Sánchez, C. (2009). Lignocellulosic residues: biodegradation and bioconversion by fungi. Biotechnol. Adv. **27**:185-194. doi: 10.1016/j.biotechadv.2008.11.001.

Satokari, R. M., Vaughan, E. E., Akkermans, A. D., Saarela, M. and de Vos, W. M. (2001). Bifidobacterial diversity in human feces detected by genus-specific PCR and denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. **67**:504-513. doi: 10.1128/aem.67.2.504-513.2001.

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., et al. (2022). Database resources of the national center for biotechnology information. Nucleic Acids Res. **50**:d20-d26. doi: 10.1093/nar/gkab1112.

Schilhabel, A., Studenik, S., Vödisch, M., Kreher, S., Schlott, B., Pierik, A. J. and Diekert, G. (2009). The ether-cleaving methyltransferase system of the strict anaerobe *Acetobacterium dehalogenans*: analysis and expression of the encoding genes. J. Bacteriol. **191**:588-599. doi: 10.1128/jb.01104-08.

Schink, B., Philipp, B. and Müller, J. (2000). Anaerobic degradation of phenolic compounds. Naturwissenschaften **87**:12-23. doi: 10.1007/s001140050002.

Schuchmann, K. and Müller, V. (2012). A bacterial electron-bifurcating hydrogenase. J. Biol. Chem. **287**:31165-31171. doi: 10.1074/jbc.M112.395038.

Seemann, T. (2018). Barrnap - BAsic Rapid Ribosomal RNA Predictor. https://github.com/tseemann/barrnap.

Shen, W., Le, S., Li, Y. and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One **11**:e0163962. doi: 10.1371/journal.pone.0163962.

Shiratori, H., Sasaya, K., Ohiwa, H., Ikeno, H., Ayame, S., Kataoka, N., Miya, A., Beppu, T. and Ueda, K. (2009). *Clostridium clariflavum* sp. nov. and *Clostridium caenicola* sp. nov., moderately thermophilic, cellulose-/cellobiose-digesting bacteria isolated from methanogenic sludge. Int. J. Syst. Evol. Microbiol. **59**:1764-1770. doi: 10.1099/ijs.0.003483-0.

Smii, L., Ben Hania, W., Cayol, J. L., Joseph, M., Hamdi, M., Ollivier, B. and Fardeau, M. L. (2015). *Fusibacter bizertensis* sp. nov., isolated from a corroded kerosene storage tank. Int. J. Syst. Evol. Microbiol. **65**:117-121. doi: 10.1099/ijs.0.066183-0.

Steenwyk, J. L., Buida, T. J., 3rd, Li, Y., Shen, X. X. and Rokas, A. (2020). ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. **18**:e3001007. doi: 10.1371/journal.pbio.3001007.

Steinegger, M. and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. **35**:1026-1028. doi: 10.1038/nbt.3988.

Studenik, S., Vogel, M. and Diekert, G. (2012). Characterization of an *O*-demethylase of *Desulfitobacterium hafniense* DCB-2. J. Bacteriol. **194**:3317-3326. doi: 10.1128/jb.00146-12.

Sukhumavasi, J., Ohmiya, K., Shimizu, S. and Ueno, K. (1988). *Clostridium josui* sp. nov., a cellulolytic, moderate thermophilic species from Thai compost. Int. J. Syst. Evol. Microbiol. **38**:179-182. doi: 10.1099/00207713-38-2-179.

Takai, K. and Horikoshi, K. (2000). Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. Appl. Environ. Microbiol. **66**:5066-5072. doi: 10.1128/aem.66.11.5066-5072.2000.

Takii, S., Hanada, S., Tamaki, H., Ueno, Y., Sekiguchi, Y., Ibe, A. and Matsuura, K. (2007). *Dethiosulfatibacter aminovoran*s gen. nov., sp. nov., a novel thiosulfate-reducing bacterium isolated from coastal marine sediment via sulfate-reducing enrichment with Casamino acids. Int. J. Syst. Evol. Microbiol. **57**:2320-2326. doi: 10.1099/ijs.0.64882-0.

Tanaka, K. and Pfennig, N. (1988). Fermentation of 2-methoxyethanol by *Acetobacterium malicum* sp. nov. and *Pelobacter venetianus*. Arch. Microbiol. **149**:181-187. doi: 10.1007/BF00422003.

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G. and Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. **40**:1023-1025. doi: 10.1038/s41587-021-01156-3.

Tian, J.-H., Pourcher, A.-M., Bouchez, T., Gelhaye, E. and Peu, P. (2014). Occurrence of lignin degradation genotypes and phenotypes among prokaryotes. Appl. Microbiol. Biotechnol. **98**:9527-9544. doi: 10.1007/s00253-014-6142-4.

Uritskiy, G. V., DiRuggiero, J. and Taylor, J. (2018). MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome **6**:158. doi: 10.1186/s40168-018-0541-1.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J. and Boerjan, W. (2010). Lignin biosynthesis and structure. Plant Physiol. **153**:895-905. doi: 10.1104/pp.110.155119.

Wang, S., Huang, H., Kahnt, J. and Thauer, R. K. (2013). A reversible electron-bifurcating ferredoxin- and NAD-dependent [FeFe]-hydrogenase (HydABC) in *Moorella thermoacetica*. J. Bacteriol. **195**:1267-1275. doi: 10.1128/jb.02158-12.

Wang, X., Lin, L. and Zhou, J. (2021). Links among extracellular enzymes, lignin degradation and cell growth establish the models to identify marine lignin-utilizing bacteria. Environ. Microbiol. **23**:160-173. doi: 10.1111/1462-2920.15289.

Welte, C. U., de Graaf, R., Dalcin Martins, P., Jansen, R. S., Jetten, M. S. M. and Kurth, J. M. (2021). A novel methoxydotrophic metabolism discovered in the hyperthermophilic archaeon *Archaeoglobus fulgidus*. Environ. Microbiol. **23**:4017-4033. doi: 10.1111/1462-2920.15546.

Whiteley, A. S., Thomson, B., Lueders, T. and Manefield, M. (2007). RNA stable isotope probing. Nature Protocols **2**:838-844. doi: 10.1038/nprot.2007.115.

Woo, H. L. and Hazen, T. C. (2018). Enrichment of bacteria from eastern Mediterranean Sea involved in lignin degradation via the phenylacetyl-CoA pathway. Front. Microbiol. **9**. doi: 10.3389/fmicb.2018.00922.

Wunder, L. C., Breuer, I., Willis-Poratti, G., Aromokeye, D. A., Henkel, S., Richter-Heitmann, T., Yin, X. and Friedrich, M. W. (2024). Manganese reduction and associated microbial communities in Antarctic surface sediments. Front. Microbiol. **15**. doi: 10.3389/fmicb.2024.1398021.

Yin, X., Kulkarni, A. C. and Friedrich, M. W. (2019). DNA and RNA stable isotope probing of methylotrophic methanogenic archaea. *In:* M. Dumont and M. Hernández García, editors. Stable Isotope Probing. Methods in Molecular Biology. Humana, New York, NY. p. 189-206. doi: 10.1007/978-1-4939-9721-3_15.

Yu, T., Hu, H., Zeng, X., Wang, Y., Pan, D., Deng, L., Liang, L., Hou, J. and Wang, F. (2023a). Widespread Bathyarchaeia encode a novel methyltransferase utilizing lignin-derived aromatics. mLife **2**:272-282. doi: 10.1002/mlf2.12082.

Yu, T., Wu, W., Liang, W., Lever, M. A., Hinrichs, K.-U. and Wang, F. (2018). Growth of sedimentary Bathyarchaeota on lignin as an energy source. Proc. Natl. Acad. Sci. U. S. A. **115**:6022-6027. doi: 10.1073/pnas.1718854115.

Yu, T., Wu, W., Liang, W., Wang, Y., Hou, J., Chen, Y., Elvert, M., Hinrichs, K.-U. and Wang, F. (2023b). Anaerobic degradation of organic carbon supports uncultured microbial populations in estuarine sediments. Microbiome **11**:81. doi: 10.1186/s40168-023-01531-z.

Yu, Y., Lee, C., Kim, J. and Hwang, S. (2005). Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. Biotechnol. Bioeng. **89**:670-679. doi: 10.1002/bit.20347.

Zhang, X., Mandelco, L. and Wiegel, J. (1994). *Clostridium hydroxybenzoicum* sp. nov., an amino acid-utilizing, hydroxybenzoate-decarboxylating bacterium isolated from methanogenic freshwater pond sediment. Int. J. Syst. Bacteriol. **44**:214-222. doi: 10.1099/00207713-44-2-214.

Zhilina, T. N., Detkova, E. N., Rainey, F. A., Osipov, G. A., Lysenko, A. M., Kostrikina, N. A. and Zavarzin, G. A. (1998). *Natronoincola histidinovorans* gen. nov., sp. nov., A new alkaliphilic acetogenic anaerobe. Curr. Microbiol. **37**:177-185. doi: 10.1007/s002849900360.

Zhilina, T. N., Kevbrin, V. V., Tourova, T. P., Lysenko, A. M., Kostrikina, N. A. and Zavarzin, G. A. (2005). *Clostridium alkalicellum* sp. nov., an obligately alkaliphilic cellulolytic bacterium from a Soda Lake in the Baikal region. Microbiol **74**:557-566. doi: 10.1007/s11021-005-0103-y.

Zhou, Y., Wei, Y., Jiang, L., Jiao, X. and Zhang, Y. (2023). Anaerobic phloroglucinol degradation by *Clostridium scatologenes*. mBio **14**:e0109923. doi: 10.1128/mbio.01099-23.

Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B. M., Martiny, J. B. H., Sogin, M., Boetius, A. and Ramette, A. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. PloS One **6**:e24570. doi: 10.1371/journal.pone.0024570.

Zoghlami, A. and Paës, G. (2019). Lignocellulosic biomass: understanding recalcitrance and predicting hydrolysis. Front. Chem. **7**. doi: 10.3389/fchem.2019.00874.

4.2 Supplementary

4.2.1 Supplementary Methods

4.2.1.1 Calculation of the ^{13}C percentage in SIP enrichments

To convert $\delta^{13}C$ values to the percentage of ^{13}C in the enrichments, the ratio of ^{13}C to ^{12}C in the sample is calculated using equation (1) with the Vienna Pee Dee Belemnite (VPDB) standard of 0.0112372 for $R_{standard}$. From the R_{sample}, the percentage of ^{13}C ($\%^{13}C$) is calculated using equation (2).

$$R_{sample} = R_{standard} \times \left(\frac{\delta^{13}C_{sample}}{1000} + 1\right) \quad (1)$$

$$\%^{13}C = \frac{R_{sample}}{1 + R_{sample}} \times 100 \quad (2)$$

4.2.1.2 Cesium trifluoroacetate solution

A cesium trifluoroacetate solution was prepared for density separation of RNA during RNA stable isotope probing. All steps of the preparation were performed under a fume hood. For 500 ml CsTFA solution, a 1L-plastic beaker was first rinsed with trifluoracetate (TFA, 99.9% purity, Carl Roth, Germany). 460 g CsOH · 2 H_2O (99.5% purity, Sigma-Aldrich, Germany) were weighed in the beaker, and a maximum of 250 ml pure water (Astacus², membraPure, Germany) was added. The beaker was placed in an ice bath for continuous cooling and stirred during acid addition. 215 ml of TFA was added slowly to the cesium hydroxide solution to prevent a rapid increase in temperature. The solution was stirred for 30 min and the pH was checked using pH paper. The pH was adjusted to 4-5 by adding CsOH or TFA. The density of the solution was determined by weighing 100 µl of the solution and adjusted to 2.2 ± 0.05 g/ml by addition of CsOH or water. The solution was autoclaved, and subsequently, the pH was adjusted to 7.0 and the density to 2.0 g/ml. The solution was sterilized through a 0.2 µm filter by vacuum-filtration into a sterile bottle. A control sample of a mixture of unlabeled and fully ^{13}C-labeled RNA from *Escherichia coli* was run using the newly prepared CsTFA solution to prepare a new density–refractory index standard curve.

4.2.1.3 Cloning

To retrieve long sequences for phylogenetic analyses and quantitative PCR (qPCR) standard preparation, a clone library of bacterial 16S rRNA gene fragments (~800) bp of five of the MAC amended samples (3,4,5-trimethoxyphenol, 3,4,5-trimethoxybenzoic acid,

3,4-dimethoxycinnamic acid, 3,4,5-trimethoxycinnamic acid, 2-methoxyhydroquinone) from the initial enrichments at day 200 (methods section 3) was constructed. The primers 27F (5'-AGAGTTTGATCCTGGCTCAG-3') (Heuer et al. 1997) and 907R (5'-CCGTCAATTCMTTTRAGTTT-3') (Lane et al. 1985) or 1492R (5'-TACGGYTACCTTGTTACGACTT-3') (Heuer et al. 1997) were used for PCR with the ALLin RPH polymerase Kit (highQu, Kraichtal, Germany) and the protocol given by the manufacturer. Thermal cycling conditions included initial denaturation at 95°C for 1 min, followed by 30 cycles with denaturation at 95°C for 15 sec, annealing at 62°C for 15 sec, amplification at 72°C for 30 sec, and final amplification at 72°C for 5 min. PCR products were purified using the Monarch PCR & DNA Cleanup kit (New England Biolabs, Frankfurt am Main, Germany), ligated into the pGEM-t vector (Promega, Mannheim, Germany) and transformed into *Escherichia coli* JM109 competent cells (Promega, Mannheim, Germany) according to the protocol of the manufacturer. DNA of randomly selected white colonies was extracted. Subsequently, PCR amplified with M13 primers (M13F-40, M13b) at the following thermal cycling conditions: initial denaturation at 95°C for 5 min, 28 cycles at 95°C for 30 sec, 55°C for 45 sec, 72°C for 90 sec, followed by final amplification at 72°C for 5 min. The cloning PCR was performed using an AmpliTaq PCR kit (Applied biosystems, Carlsbad, USA). Amplicons of 136 clones were submitted to LGC Genomics (Berlin, Germany) for Sanger sequencing.

4.2.1.4 qPCR Standard

Selected clone sequences affiliated with different families within the class of Clostridia were used for qPCR standard preparation. The 16S rRNA gene of previously extracted colonies (methods section 5) was amplified with the primer set 27F (5'-AGAGTTTGATCCTGGCTCAG-3') (Heuer et al. 1997) and 907R (5'-CCGTCAATTCMTTTRAGTTT-3') (Lane et al. 1985) using the AmpliTaq PCR kit (Applied biosystems, Carlsbad, USA) according to the protocol of the manufacturer. PCR was performed using the following thermal cycling conditions: initial denaturation at 95°C for 5 min, denaturation at 95°C for 30 sec, annealing at 58°C for 45 sec and elongation at 72°C for 60 sec. A total of 30 cycles were run. The resulting ~800 bp PCR product was screened by gel electrophoresis and purified using the Monarch® PCR and DNA purification kit (New England Biolabs, Frankfurt am Main, Germany). Concentrations of standard DNA were quantified using the Quant-iT PicoGreen dsDNA assay kit (Thermo Fisher Scientific, United States). Each qPCR standard was diluted to a concentration of 0.5 ng/µl for further use.

4.2.1.5 qPCR primer design

qPCR primers were designed to quantify the Clostridia present in our slurry enrichments. Clone sequences of the slurry enrichments were aligned using the Silva Incremental Aligner (SINA) v1.2.11 (Pruesse et al. 2012), imported into the ARB software v6.0.2 (Ludwig et al. 2004) and added to the SILVA 16S rRNA gene phylogenetic base tree (database version 138.1 Ref NR 99) using ARB Parsimony. Based on their taxonomic placement, clone sequences of Clostridia were selected. Additionally, clone sequences from other bacterial groups (Desulfobacteria, Planctomycetes, Caldatribacteriota, Izemoplasmatales, Dehalococcoidia, Anaerolinea and uncultured Firmicutes) were chosen as outgroups. For qPCR primer design, Primer Prospector v1.0.1 (Walters et al. 2011) was used to calculate possible primer sequences. Calculated primers were checked in ARB for their theoretical specificity against sequences of other closely related groups and phyla. Resulting primer sequences 579F (5'-AAGCCMCGGCTAACTACGTG-3') and 747R (5'-CGCTACACTAGGAATTCC RCYT-3') were selected. The primer pair was validated with high specificity against the newly prepared standards of the orders Christensenellales, Oscillospirales, Tissierellales and Eubacteriales within the class of Clostridia (positive control) and other standards (negative controls) of Desulfobacteria, Planctomycetes, Caldatribacteriota, Izemoplasmatales, Dehalococcoidia, Anaerolinea and uncultured Firmicutes. Further, a test for the efficiency of the new Clostridia standard at annealing temperatures ranging from 58°C to 62°C was performed, followed by a melting curve stage after PCR, resulting in the best annealing temperature at 60°C with a final primer concentration of 300 nM.

4.2.1.6 LC-MS/MS measurements of MACs

A binary solvent gradient with solvent A = 0.1% formic acid and 4 mM ammonium formate in ultrapure water and solvent B = 0.1% formic acid and 4 mM ammonium formate in methanol was used at a flow rate of 0.45 mL per min on a C18 column (C18 BEH, 100 × 2 mm, 1.7 µm particle size, ACQUITY Waters, equipped with guard-column) at 40°C. The gradient program was as follows: T0 min: B = 5%, T0.8 min B = 5%, T2.5 min: B = 100%, T3.3 min: B = 100%; T3.5 min: B = 5%. The column was equilibrated for 0.5 min between samples. The first 0.7 min of the effluent was diverted to waste. MS spectra were acquired in full scan and data independent (DIA-MS²) mode. Full scans were acquired at RES=70,000 FWHM (m/z 200) with an automatic gain control (AGC) of 3e6, injection time (IT) of 50 ms and scan range of m/z 100 to 1000. DIA-MS² experiments were performed with a full scan at RES=35,000 FWHM (m/z 200) with an automatic gain control (AGC) of 2e5, injection time (IT) of 50 ms,

isolation window of 3.0 m/z and stepped normalized collision energy of 30, 40. The inclusion lists for the DIA-MS² contained only one exact mass for the MAC used in each incubation experiment (Table S21). The heated electrospray ionization source was set to 3.0 kV spray voltage, the aux gas to 430 °C at a rate of 13 and the sheath gas to 51. The capillary temperature was set to 270°C. Positive Ion Calibration Solution (Pierce, Thermo Fisher Scientific) was used for the calibration of the instrument.

4.2.1.7 MTI gene validation

As the gene hits retrieved by the MTI HMM were annotated as uroporphyrinogen decarboxylase (*hemE*) by KEGG and NR, we performed an additional hmmsearch with the protein family model TIGR01464.1 of the *hemE* protein sequence to remove false positives. All genes only annotated by our novel MTI protein family model with an e-value < 0.001 were regarded as hit. Additionally, we removed all hits from our MTI genes, if the *hemE* value and *hemE* score were higher than those computed for the MTI hits. The technique was validated using the type strain of *Acetobacterium woodii*, which, based on Lechtenfeld et al. (2018), contains 23 homolog genes of MTI. Our search detected a total of 23 MTI genes within the genome. Using the annotation of the type strain provided by NCBI, detected MTI genes were manually checked for validation.

4.2.2 Supplementary Results

4.2.2.1 Enrichment of Clostridia with MACs in marine sediment slurry incubations

Acetivibrionales were only enriched in the treatment with 3,4,5-trimethoxybenzoic acid (up to 22%). Further, an ASV, taxonomically assigned to unclassified Clostridia (sq3; up to 83%) was observed in the same treatment. The pairwise identity of ASVs of the order Acetivibrionales and the ASV of the unclassified Clostridia revealed high sequence identity (sq3 vs. sq8, sq19, sq32; 98-98.4% 16S rRNA gene sequence identity; Fig. S7). Therefore, this unclassified Clostridia ASV was eventually assigned to the Acetivibrionales. All ASVs within this order could not be taxonomically classified below the order level.

Eubacteriales were detected in all treatments with highest relative abundances in treatments with vanillin (78%), 2-methoxyphenol (60%), 3,4,5-trimethoxyphenol (83%), 3,4-dimethoxycinnamic acid (62%), 3,4,5-trimethoxycinnamic acid (61%), 2-methoxyhydroquinone (40%), 1,2,4-trimethoxybenzene (86%) and 3,4,5-trimethoxybenzaldehyde (69%). ASVs found in the vanillin, 3,4,5-trimethoxycinnamic acid and 2-methoxyhydroquinone treatments could be assigned to the M08DMB genus. Other ASVs were primarily assigned to the family Alkalibacteraceae, unclassified on genus level.

Specifically, one ASV sequence (sq1), assigned to unclassified Alkalibacteraceae, could be detected in nine treatments and was highly abundant in seven (Table S10). Other ASVs of this group were only present in three or fewer treatments.

Oscillospirales were abundant in the treatments with 3,4-dimethoxycinnamic acid (20%), 3,4,5-trimethoxycinnamic acid (19%) and 3,4,5-trimethoxybenzaldehyde (48%). The most abundant ASVs found in the 3,4-dimethoxycinnamic acid (20%) and 3,4,5-trimethoxybenzaldehyde (47%) treatment were assigned to the family Oscillospiraceae. ASVs in the two different replicates of the 3,4,5-trimethoxycinnamic acid were assigned to unclassified Oscillospirales (12%) or *Sporobacter* (9.7%).

Peptostreptococcales were most abundant in the treatment with 3,4,5-trimethoxybenzaldehyde (24%). The found ASV (sq27) could not be classified below order level. Upon screening of unclassified Clostridia ASVs we additionally detected one ASV (sq182) with a high relative abundance (19%), which showed a sequence identity of 98% to the Peptostreptococcales ASV sq27. Due to the high 16S rRNA gene sequence identity, the unclassified ASV sq182 was sorted into the Peptostreptococcales.

Tissierellales were most abundant on day 200 in the treatment with 2-methoxyhydroquinone (26%) and were assigned to the genus *Dethiosulfatibacter*.

Besides Clostridia, we observed an enrichment of members of the phylum Bacillota_B in treatments with 2-methoxybenzoic acid (46%), 2-methoxyhydroquinone (42%) and 3,4-dimethoxycinnamic acid (44%). Within the 2-methoxybenzoic acid treatment, ASVs were classified in the Desulfotomaculales order (46%). ASVs within the 2-methoxyhydroquinone and 3,4-dimethoxycinnamic acid treatments could be assigned to the family Syntrophomonadaceae (15% and 22%) or yet unclassified Bacillota_B (25% and 13%).

4.2.2.2 Phylogenetic novelty among detected Clostridia

Acetivibrionales MAGs (Aceti_1 – Aceti_5) clustered closely together in the phylogenomic tree with an average nucleotide identity (ANI) of 77-86% (Fig. S18) and an amino acid identity (AAI) of 60-72% (Fig. S19), grouping the MAGs as one family (Table S12). Based on RED Acetivibrionales, MAGs are considered a novel family in which the MAGs Aceti_2-Aceti_5 are of the same genus, following thresholds in Konstantinidis et al. (2017). The Ch29 MAG (Ch29_1) was assigned to the genus JAENYX01 in order Ch29. Of the four MAGs affiliated with Eubacteriales (Eubac_1 – Eubac_4), one MAG derived from the ^{13}C-vanillin + ^{13}C-DIC + BES RNA-SIP sample could be assigned to the genus M08DMB (Eubac_3). Therefore, this MAG represents the found ASVs within the MAC and SIP enrichments assigned to the same

genus. The additional MAGs of the Eubacteriales were sorted into the family of Alkalibacteraceae (AAI of 54-62%) but could not be assigned on genus level. Oscillospirales MAGs (Oscillo_1 – Oscillo_8) were classified as *Sporobacter*, DUPK01 and JAEWRY01 genera. Three of the MAGs could not be classified on genus level. Still, their placement in the phylogenomic tree was close to those MAGs identified as *Sporobacter,* with a higher AAI of 57-59% than other Oscillospirales MAGs (51-55%). MAGs of the Peptostreptococcales (Pepto_1-Pepto_5) could be assigned to the genera JAFGAC01, MT110 and Soudan-22. Tissierellales MAGs were assigned to the families Sedimentibacteraceae and VENL01.

4.2.2.3 High MTI gene diversity in Clostridia MAGs

Acetivibrionales MAGs contained between 1 and 21 unique MTI genes, with four MAGs sharing gene cluster 55. Three Acetivibrionales MAGs further contained gene cluster 75 in duplicates. The Eubacteriales MAGs contained between 11 and 62 unique MTI genes, with Eubac_1-Eubac_3 sharing 8 gene clusters. Oscillospirales MAGs contained between 11 and 43 unique MTI genes. Most Oscillospirales MAGs shared 4 MTI gene clusters (clusters 76, 79, 80, 100). Peptostreptococcales MAGs did contain only singletons. The Tissierellales MAGs could be detected in cluster 75, which was also shared with the order Acetivibrionales and Oscillospirales, and cluster 162, shared with Eubacteriales.

4.2.3 Supplementary Tables

Table S1 Stable Isotope Probing (SIP) enrichment setup. Overview of treatments used for the RNA-SIP. Each treatment was set up in triplicates to retrieve sufficient RNA for subsequent RNA-SIP. The unlabeled and labeled DIC (bicarbonate) controls were set up in two sets. Control (A) was stopped after 111 days of incubation, Control (B) was stopped after 188 days. Enrichments were set up for the three MACs: vanillin, 2-methoxyphenol, and 2-methoxybenzoic acid, in which only the carbon of the methoxy group was labeled. Treatment 5 of each MAC contained sodium 2-bromoethanesulfonate (BES) to inhibit methanogenic activity. DIC was supplied in a concentration of 10 mM, MACs were supplied in a concentration of 1 mM, BES was supplied in a concentration of 5mM.

Substrate	Treatment	DIC	^{13}C-DIC	MAC	^{13}C-MAC	BES
Control	DIC (A)	x				
	^{13}C-DIC (A)		x			
	DIC (B)	x				
	^{13}C-DIC (B)		x			
vanillin	1	x		x		
	2		x	x		
	3	x			x	
	4		x		x	
	5		x		x	x
2-methoxy-phenol	1	x		x		
	2		x	x		
	3	x			x	
	4		x		x	
	5		x		x	x
2-methoxy-benzoic acid	1	x		x		
	2		x	x		
	3	x			x	
	4		x		x	
	5		x		x	x

Table S2 Sequenced fractions of the vanillin amended SIP enrichments. The total RNA amount used for the density separation is indicated for each treatment. All fractions used for amplicon sequencing are listed with their density after density separation and fractionation. The amount of RNA retrieved for each single fraction is listed.

	Treatment	RNA amount used for SIP (ng)	Fraction	Density (g/ml)	RNA amount per fraction (ng)
vanillin	MAC + DIC	1000.0	3	1.832	3.9
			5	1.818	11.9
			7	1.805	23.0
			9	1.795	50.5
			11	1.784	174.3
	MAC + ^{13}C-DIC	799.7	3	1.832	7.2
			5	1.818	20.3
			7	1.805	52.5
			9	1.791	67.5
			11	1.781	46.9
	^{13}C-MAC + DIC	1000.0	3	1.832	9.6
			5	1.818	10.3
			7	1.808	24.9
			9	1.795	114.1
			11	1.781	165.0
	^{13}C-MAC + ^{13}C-DIC	1000.0	3	1.832	9.5
			5	1.818	24.2
			7	1.808	52.5
			9	1.795	61.9
			11	1.781	60.6
	^{13}C-MAC + ^{13}C-DIC + BES	1000.0	3	1.832	11.2
			5	1.818	28.1
			7	1.805	60.0
			9	1.795	43.7
			11	1.784	51.9

Table S3 Sequenced fractions of the 2-methoxyphenol amended SIP enrichments. The total RNA amount used for the density separation is indicated for each treatment. All fractions used for amplicon sequencing are listed with their density after density separation and fractionation. The amount of RNA retrieved for each single fraction is listed.

	Treatment	RNA amount used for SIP (ng)	Fraction	Density (g/ml)	RNA amount per fraction (ng)
2-methoxyphenol	MAC + DIC	801.6	3	1.832	10.8
			5	1.818	16.1
			7	1.808	43.9
			9	1.795	73.8
			11	1.781	58.3
	MAC + ^{13}C-DIC	1000.0	3	1.835	14.6
			5	1.822	36.9
			7	1.808	67.2
			9	1.798	110.0
			11	1.784	79.5
	^{13}C-MAC + DIC	747.7	3	1.835	10.4
			5	1.822	12.9
			7	1.808	29.1
			9	1.795	96.3
			11	1.784	69.0
	^{13}C-MAC + ^{13}C-DIC	766.9	3	1.835	14.7
			5	1.822	33.8
			7	1.808	42.7
			9	1.795	78.9
			11	1.784	62.8
	^{13}C-MAC + ^{13}C-DIC + BES	1000.0	3	1.832	20.6
			5	1.818	60.3
			7	1.805	41.0
			9	1.791	96.6
			11	1.778	42.4

Table S4 Sequenced fractions of the 2-methoxybenzoic acid amended SIP enrichments. The total RNA amount used for the density separation is indicated for each treatment. All fractions used for amplicon sequencing are listed with their density after density separation and fractionation. The amount of RNA retrieved for each single fraction is listed.

	Treatment	RNA amount used for SIP (ng)	Fraction	Density (g/ml)	RNA amount per fraction (ng)
2-methoxybenzoic acid	MAC + DIC	361.6	3	1.835	6.6
			5	1.822	5.0
			7	1.808	17.1
			9	1.798	43.2
			11	1.784	67.5
	MAC + ^{13}C-DIC	336.1	3	1.835	5.3
			5	1.825	13.2
			7	1.812	20.8
			9	1.798	49.8
			11	1.784	41.9
	^{13}C-MAC + DIC	292.0	3	1.835	2.8
			5	1.825	4.9
			7	1.812	12.2
			9	1.798	40.2
			11	1.788	41.6
	^{13}C-MAC + ^{13}C-DIC	301.6	3	1.835	9.4
			5	1.822	13.8
			7	1.808	33.8
			9	1.798	39.2
			11	1.784	36.1
	^{13}C-MAC + ^{13}C-DIC + BES	230.3	3	1.832	7.1
			5	1.818	18.3
			7	1.805	35.0
			9	1.795	50.8
			11	1.781	34.7

Table S5 Sequenced fractions of the DIC control enrichments. For each treatment the total RNA amount used for the density separation is indicated. All fractions used for amplicon sequencing are listed with their according density after density separation and fractionation. The amount of RNA retrieved for each single fraction is listed.

	Treatment	RNA amount used for SIP (ng)	Fraction	Density (g/ml)	RNA amount per fraction (ng)
DIC	Control A	770.3	3	1.832	20.6
			5	1.822	29.6
			7	1.808	64.4
			9	1.795	212.0
			11	1.784	224.0
^{13}C-DIC	Control A	779.1	3	1.829	6.7
			5	1.822	19.8
			7	1.808	43.0
			9	1.795	107.2
			11	1.784	127.1
DIC	Control B	522.3	3	1.832	32.1
			5	1.818	28.5
			7	1.805	59.2
			9	1.791	193.3
			11	1.781	210.1
^{13}C-DIC	Control B	486.1	3	1.829	23.4
			5	1.818	29.0
			7	1.801	65.6
			9	1.791	162.7
			11	1.778	56.7

Table S6 MAC enrichment setup. Overview of used methoxylated aromatic compounds (MACs) and the enrichment setup for initial enrichments. For each substrate, two replicates were set up. MACs were used in a concentration of 12 mM. The control contained no additional carbon substrate and was used as slurry-only control. Additionally, sodium 2-bromoethanesulfonate (BES), was added to all enrichments in a final concentration of 5mM to inhibit methanogenic activity.

Sample	Substrate	Abbreviation	Replicate 1	Replicate 2
A	Control	Control	x	x
B	3,5-dimethoxybenzoic acid	2-MBA	x	x
C	Vanillin	Vanillin	x	x
D	2-methoxyphenol	MP	x	x
E	3,4,5-trimethoxyphenol	3-MP	x	x
F	2-methoxybenzoic acid	MBA	x	x
G	3,4,5-trimethoxybenzoic acid	3-MBA	x	x
H	3-methoxyphenylacetic acid	MPAA	x	x
I	3,4,5-trimethoxyphenylacetic acid	3-MPAA	x	x
J	3-methoxycinnamic acid	MCA	x	x
K	3,4-dimethoxycinnamic acid	2-MCA	x	x
L	3,4,5-trimethoxycinnamic acid	3-MCA	x	x
M	2-methoxyhydroquinone	MHQ	x	x
N	1,2,4-trimethoxybenzene	TMB	x	x
O	3,4,5-trimethoxybenzaldehyde	TMBA	x	x
P	2-methoxy-3-pyridinylboranic acid	2-M 3-PBA	x	x

Table S7 Trimming parameters for single sequencing libraries. Sequencing analysis of single libraries of the MAC and SIP enrichments was conducted using the given trimming parameters (R1 length, R2 length). If sequencing libraries were processed on multiple flowcell lanes to retrieve sufficient reads, parameters for all flowcell lanes are noted.

	Lib ID	Target	Flowcell lane ID	R1 length (bp)	R2 length (bp)	Sequence lengths kept (bp)
MAC enrichments	Lib73	Bacteria	HWCCVDRX2_L2	140	160	249-254
	Lib74		HWCCVDRX2_L2	140	160	249-254
	Lib75		HWCCVDRX2_L2	140	160	249-254
	Lib77		HVW27DRX2_L1	125	175	249-254
	Lib78		HWCCVDRX2_L2	140	160	249-254
	Lib70	Archaea	HWCGMDRX2_L2	110	190	251-255
	Lib71		HWCCVDRX2_L2	125	175	251-255
	Lib72		HVW27DRX2_L1	135	165	251-255
	Lib76		HVW27DRX2_L1	135	165	251-255
	Lib77		HVW27DRX2_L1	135	165	251-255
SIP enrichments	Lib67	Bacteria	HWCCVDRX2_L1	140	150	249-254
	Lib69		HWCCVDRX2_L2	140	150	249-254
	Lib84		HKFKTDRX3_L1	130	170	249-254
	Lib84		HKY7TDRX3_L1	130	170	249-254
	Lib84		HMCHCDRX3_L1	140	160	249-254
	Lib85		HKFKTDRX3_L1	130	170	249-254
	Lib66	Archaea	HWCCVDRX2_L2	145	145	251-255
	Lib68		HVW27DRX2_L2	120	170	251-255
	Lib83		HKFKTDRX3_L2	90	200	251-255
	Lib83		HKY7TDRX3_L1	90	200	251-255
	Lib85		HKFKTDRX3_L1	90	200	251-255

Table S8 Number of different MAC treatments each ASV was present in.

ASV	Order	Number of MAC treatments
sq1	Eubacteriales	9
sq2	Eubacteriales	3
sq3	Acetivibrionales	2
sq8	Acetivibrionales	1
sq9	Eubacteriales	3
sq10	Oscillospirales	3
sq14	Eubacteriales	1
sq16	Eubacteriales	1
sq19	Acetivibrionales	1
sq27	Peptostreptococcales	2
sq32	Acetivibrionales	4
sq34	Tissierellales	2
sq38	Peptostreptococcales	4
sq47	Tissierellales	2
sq57	Tissierellales	1
sq65	Oscillospirales	2
sq67	Peptostreptococcales	1
sq83	Eubacteriales	1
sq110	Peptostreptococcales	1
sq113	Oscillospirales	1
sq127	Eubacteriales	1
sq137	other Clostridia	1
sq146	Oscillospirales	1
sq182	Peptostreptococcales	1
sq219	unclassified sq219	1
sq248	Tissierellales	2
sq252	Oscillospirales	1
sq253	Eubacteriales	1
sq261	Peptostreptococcales	1
sq315	other Clostridia	1
sq329	Tissierellales	1
sq376	Oscillospirales	2
sq392	Peptostreptococcales	1
sq481	Peptostreptococcales	1
sq637	Oscillospirales	1
sq1140	Oscillospirales	1

Table S9 Chi-Square test results (*p*-value: 0.0007). Enrichment of Clostridia is defined as TRUE for samples with a relative abundance of Clostridia > 30%.

		No. of methoxygroups		
	enrichment of Clostridia	**1**	**2**	**3**
observed	FALSE	28	4	9
	TRUE	14	8	27
expected	FALSE	19.133	5.467	16.400
	TRUE	22.867	6.533	19.600

Table S10 Nucleotide BLAST hits for Clostridia ASVs within SIP incubations.

Query ASV	Description	Scientific name	Max Score	Total Score	Query Cover	E-value	Identity (%)	Acc. Len	Accession no.	Publication of species
sq1	*Alkalibaculum sporogenes* strain M08 DMB 16S ribosomal RNA, partial sequence	*Alkalibaculum sporogenes*	426	426	100%	3E-119	97.61%	1481	NR_174299.1	(Khomyakova et al. 2020)
sq5	*Alkalibaculum sporogenes* strain M08 DMB 16S ribosomal RNA, partial sequence	*Alkalibaculum sporogenes*	431	431	100%	2E-120	98.01%	1481	NR_174299.1	(Khomyakova et al. 2020)
sq36	*Alkalibaculum sporogenes* strain M08 DMB 16S ribosomal RNA, partial sequence	*Alkalibaculum sporogenes*	426	426	100%	3E-119	97.61%	1481	NR_174299.1	(Khomyakova et al. 2020)
sq78	*Alkalibaculum sporogenes* strain M08 DMB 16S ribosomal RNA, partial sequence	*Alkalibaculum sporogenes*	422	422	100%	1E-117	97.21%	1481	NR_174299.1	(Khomyakova et al. 2020)
sq198	*Alkalibaculum sporogenes* strain M08 DMB 16S ribosomal RNA, partial sequence	*Alkalibaculum sporogenes*	426	426	100%	3E-119	97.61%	1481	NR_174299.1	(Khomyakova et al. 2020)

Table S11 Nucleotide BLAST hits for Clostridia ASVs within MAC enrichments.

Query ASV	Description	Scientific name	Max Score	Total Score	Query Cover	E-value	Identity (%)	Acc. Len	Accession no.	Publication of species
Alkalibacter saccharofermentans										
sq1	*Alkalibacter saccharofermentans* strain Z-79820 16S ribosomal RNA, partial sequence	*Alkalibacter saccharofermentans*	377	377	100%	4E-104	93.23%	1465	NR_042834.1	(Garnova et al. 2004)
sq9	*Alkalibacter saccharofermentans* strain Z-79820 16S ribosomal RNA, partial sequence	*Alkalibacter saccharofermentans*	381	381	100%	1E-105	93.63%	1465	NR_042834.1	(Garnova et al. 2004)
sq127	*Alkalibacter saccharofermentans* strain Z-79820 16S ribosomal RNA, partial sequence	*Alkalibacter saccharofermentans*	386	386	100%	8E-107	94.02%	1465	NR_042834.1	(Garnova et al. 2004)
Clostridium thermosuccinogenes										
sq3	*Clostridium thermosuccinogenes* strain DSM 5807 16S ribosomal RNA, partial sequence	*Clostridium thermosuccinogenes*	404	404	100%	3E-112	95.62%	1640	NR_119284.1	(Drent et al. 1991)
sq8	*Clostridium thermosuccinogenes* strain DSM 5807 16S ribosomal RNA, partial sequence	*Clostridium thermosuccinogenes*	413	413	100%	6E-115	96.41%	1640	NR_119284.1	(Drent et al. 1991)
sq19	*Clostridium thermosuccinogenes* strain DSM 5807 16S ribosomal RNA, partial sequence	*Clostridium thermosuccinogenes*	417	417	100%	1E-116	96.81%	1640	NR_119284.1	(Drent et al. 1991)
sq32	*Clostridium thermosuccinogenes* strain DSM 5807 16S ribosomal RNA, partial sequence	*Clostridium thermosuccinogenes*	417	417	100%	1E-116	96.81%	1640	NR_119284.1	(Drent et al. 1991)
Papillibacter cinnamivorans										
sq10	*Papillibacter cinnamivorans* strain CIN1 16S ribosomal RNA, partial sequence	*Papillibacter cinnamivorans*	396	396	100%	5E-110	94.82%	1491	NR_025025.1	(Defnoun et al. 2000)
sq113	Papillibacter cinnamivorans strain CIN1 16S ribosomal RNA, partial sequence	Papillibacter cinnamivorans	400	400	100%	4E-111	95.22%	1491	NR_025025.1	(Defnoun et al. 2000)

Query ASV	Description	Scientific name	Max Score	Total Score	Query Cover	E-value	Identity (%)	Acc. Len	Accession no.	Publication of species
Alkalibaculum sporogenes										
sq2	*Alkalibaculum sporogenes* strain M08 DMB 16S ribosomal RNA, partial sequence	*Alkalibaculum sporogenes*	399	399	100%	4E-110	95.22%	1481	NR_174299.1	(Khomyakova et al. 2020)
sq14	*Alkalibaculum sporogenes* strain M08 DMB 16S ribosomal RNA, partial sequence	*Alkalibaculum sporogenes*	422	422	100%	1E-117	97.21%	1481	NR_174299.1	(Khomyakova et al. 2020)
sq16	*Alkalibaculum sporogenes* strain M08 DMB 16S ribosomal RNA, partial sequence	*Alkalibaculum sporogenes*	426	426	100%	3E-119	97.61%	1481	NR_174299.1	(Khomyakova et al. 2020)
sq83	*Alkalibaculum sporogenes* strain M08 DMB 16S ribosomal RNA, partial sequence	*Alkalibaculum sporogenes*	431	431	100%	2E-120	98.01%	1481	NR_174299.1	(Khomyakova et al. 2020)
sq253	*Alkalibaculum sporogenes* strain M08 DMB 16S ribosomal RNA, partial sequence	*Alkalibaculum sporogenes*	386	386	100%	8E-107	94.02%	1481	NR_174299.1	(Khomyakova et al. 2020)
Dethiosulfatibacter aminovorans										
sq34	*Dethiosulfatibacter aminovorans* strain C/G2 16S ribosomal RNA, partial sequence	*Dethiosulfatibacter aminovorans*	425	425	100%	9E-119	97.60%	1479	NR_041309.1	(Takii et al. 2007)
sq47	*Dethiosulfatibacter aminovorans* strain C/G2 16S ribosomal RNA, partial sequence	*Dethiosulfatibacter aminovorans*	420	420	100%	4E-117	97.20%	1479	NR_041309.1	(Takii et al. 2007)
sq57	*Dethiosulfatibacter aminovorans* strain C/G2 16S ribosomal RNA, partial sequence	*Dethiosulfatibacter aminovorans*	434	434	100%	2E-121	98.40%	1479	NR_041309.1	(Takii et al. 2007)
sq248	*Dethiosulfatibacter aminovorans* strain C/G2 16S ribosomal RNA, partial sequence	*Dethiosulfatibacter aminovorans*	398	398	100%	1E-110	95.20%	1479	NR_041309.1	(Takii et al. 2007)

Query ASV	Description	Scientific name	Max Score	Total Score	Query Cover	E-value	Identity (%)	Acc. Len	Accession no.	Publication of species
Intestinimonas timonensis										
sq27	*Intestinimonas timonensis* strain GD4 16S ribosomal RNA, partial sequence	*Intestinimonas timonensis*	363	363	100%	3E-100	92.03%	1495	NR_179423.1	(Durand et al. 2017)
Colidextribacter massiliensis										
sq182	*Colidextribacter massiliensis* strain Marseille-P3083 16S ribosomal RNA, partial sequence	*Colidextribacter massiliensis*	372	372	100%	5E-102	92.83%	1487	NR_147375.1	Ricaboni et al 2017)
Alkaliphilus hydrothermalis										
sq38	*Alkaliphilus hydrothermalis* strain FatMR1 16S ribosomal RNA, partial sequence	*Alkaliphilus hydrothermalis*	402	402	100%	1E-111	95.60%	1454	NR_152624.1	(Ben Aissa et al. 2015)
Sporobacter termitidis										
sq65	*Sporobacter termitidis* strain SYR 16S ribosomal RNA, partial sequence	*Sporobacter termitidis*	399	399	100%	4E-111	95.22%	1506	NR_044972.1	(Grech-Mora et al. 1996)
sq146	*Sporobacter termitidis* strain SYR 16S ribosomal RNA, partial sequence	*Sporobacter termitidis*	426	426	100%	3E-119	97.61%	1506	NR_044972.1	(Grech-Mora et al. 1996)
sq252	*Sporobacter termitidis* strain SYR 16S ribosomal RNA, partial sequence	*Sporobacter termitidis*	435	435	100%	5E-122	98.41%	1506	NR_044972.1	(Grech-Mora et al. 1996)
Sporosalibacterium tautonense										
sq67	*Sporosalibacterium tautonense* strain MRo-4 16S ribosomal RNA, partial sequence	*Sporosalibacterium tautonense*	393	393	100%	6E-109	94.80%	1520	NR_156896.1	(Podosokorskaya et al. 2017)
Serpentinicella alkaliphila										
sq110	*Serpentinicella alkaliphila* strain 3b 16S ribosomal RNA, partial sequence	*Serpentinicella alkaliphila*	411	411	100%	2E-114	96.40%	1417	NR_152685.1	(Mei et al. 2016)

Table S12 Taxonomic classification of Clostridia MAGs by the GTDB database r214. All MAGs were assigned to the domain bacteria and phylum Bacillota_A.

MAG	Name	Order	Family	Genus
E4_d458_spades_bin_53_orig-contigs	Aceti_1	Acetivibrionales	unclassified	unclassified
G3_d458_megahit_bin_18_orig-contigs	Aceti_2	Acetivibrionales	unclassified	unclassified
G3_d458_spades_bin_18_indstr_refined-contigs	Aceti_3	Acetivibrionales	unclassified	unclassified
G3_d458_spades_bin_33_strict_refined-contigs	Aceti_4	Acetivibrionales	unclassified	unclassified
L3_d458_spades_bin_6_orig_refined-contigs	Aceti_5	Acetivibrionales	unclassified	unclassified
E4_d458_spades_bin_64_permissive_refined-contigs	Ch29_1	Ch29	Ch29	JAENYX01
G3_d458_megahit_bin_21_strict_refined-contigs	Eubac_1	Eubacteriales	Alkalibacteraceae	unclassified
L3_d458_spades_bin_34_indper_refined-contigs	Eubac_2	Eubacteriales	Alkalibacteraceae	unclassified
Van_BES_megahit_bin_42_indstr_refined-contigs	Eubac_3	Eubacteriales	Alkalibacteraceae	M08DMB
G3_d458_spades_bin_19_indper_refined-contigs	Eubac_4	Eubacteriales	unclassified	unclassified
Co_bin_102_permissive_refined-contigs	Oscillo_1	Oscillospirales	Oscillospiraceae	unclassified
E4_d458_spades_bin_8_permissive_refined-contigs	Oscillo_2	Oscillospirales	Oscillospiraceae	unclassified
L3_d458_megahit_bin_13_permissive_refined-contigs	Oscillo_3	Oscillospirales	Oscillospiraceae	unclassified
G3_d458_megahit_bin_13_indstr_refined-contigs	Oscillo_4	Oscillospirales	Oscillospiraceae	Sporobacter
L3_d458_spades_bin_8_indstr_refined-contigs	Oscillo_5	Oscillospirales	Oscillospiraceae	Sporobacter
M4_d458_spades_bin_15_permissive_refined-contigs	Oscillo_6	Oscillospirales	Oscillospiraceae	DUPK01
L3_d458_spades_bin_21_orig_refined-contigs	Oscillo_7	Oscillospirales	Oscillospiraceae	DUPK01
M4_d458_megahit_bin_49_permissive_refined-contigs	Oscillo_8	Oscillospirales	Oscillospiraceae	JAEWRY01
Co_bin_88_orig_refined-contigs	Pepto_1	Peptostreptococcales	Acidaminobacteraceae	JAFGAC01
L3_d458_megahit_bin_58_indper-contigs	Pepto_2	Peptostreptococcales	Anaerovoracaceae	MT110
E4_d458_megahit_bin_7_strict_refined-contigs	Pepto_3	Peptostreptococcales	JAAYPU01	Soudan-22
M4_d458_megahit_bin_14_orig_refined-contigs	Pepto_4	Peptostreptococcales	JAAYPU01	Soudan-22
M4_d458_spades_bin_28_indstr_refined-contigs	Pepto_5	Peptostreptococcales	JAAYPU01	Soudan-22
Co_bin_49_orig_refined-contigs	Tissi_1	Tissierellales	Sedimentibacteraceae	unclassified
L3_d458_spades_bin_31_orig_refined-contigs	Tissi_2	Tissierellales	VENL01	unclassified
L3_d458_spades_bin_64_permissive_refined-contigs	Clostridia_1	unclassified	unclassified	unclassified

Table S13 Quality statistics of Clostridia MAGs computed by checkM2. The genome size was adjusted using the MAG completeness.

Name	Completeness	Contamination	Adjusted genome size	Total contigs	Contig N50	Max. contig length	Coding density	Total coding sequences
Aceti_1	99.94	0	3270887.289	33	142653	286963	0.854	2997
Aceti_2	99.6	0.02	2702478.672	32	148690	350469	0.882	2580
Aceti_3	95.51	1.22	3710409.729	65	131374	293911	0.881	3669
Aceti_4	97.21	0.33	3272512.436	65	86355	518720	0.882	3232
Aceti_5	99.14	0.54	3472145.521	26	334540	722750	0.886	3206
Ch29_1	97.57	1.19	2602962.703	149	24355	91394	0.886	2709
Eubac_1	98.58	0.17	3176970.191	37	293019	550352	0.868	3120
Eubac_2	99.69	0	2792367.742	38	108910	393395	0.865	2701
Eubac_3	99.23	0.04	2899864.774	104	46335	139086	0.832	2745
Eubac_4	92.85	0.26	2145683.649	170	22963	85605	0.851	2222
Oscillo_1	95.54	1.77	4426413.104	97	66716	272170	0.874	4186
Oscillo_2	94.68	0.68	3281450.684	110	65962	224742	0.878	3224
Oscillo_3	96.85	0	3257588.49	89	100514	226578	0.877	3246
Oscillo_4	96.47	2.24	3936166.046	75	90364	196957	0.865	3796
Oscillo_5	97.13	0.33	2983959.869	62	71276	258316	0.861	2727
Oscillo_6	97.61	0.15	3429523.446	61	117377	264118	0.868	3327
Oscillo_7	96.84	0	2711489.98	12	460022	555907	0.877	2652
Oscillo_8	95.12	1.26	2363595.027	315	9957	40297	0.897	2616
Pepto_1	93.85	4.11	2430554.517	161	23960	74581	0.902	2650
Pepto_2	95.83	1.09	4154905.143	83	151421	298731	0.886	3950
Pepto_3	84.7	0.08	2223960.277	347	9906	32063	0.878	2661
Pepto_4	96.68	2.31	2719879.104	166	25010	72350	0.872	2679
Pepto_5	99.7	0.99	2926541.956	75	68847	188635	0.863	2774
Tissi_1	90.53	5.08	3669591.906	140	50536	257770	0.872	3701
Tissi_2	97.36	0.7	3276105.584	45	161113	562230	0.851	3161
Clostridia_1	83.77	0.39	2352863.069	374	8968	37070	0.876	2949

Table S14 Top three hits for sequence identities of ASV sequences blasted against 16S rRNA genes derived from Clostridia MAGs using the nucleotide blast.

Query ASV	Closest MAG	Identity (%)	Query Cover	E-value
sq8	Aceti_1	96.813	100	7.89E-118
sq19	Aceti_1	96.016	100	4.09E-115
sq3	Aceti_1	95.219	100	2.12E-112
sq32	Aceti_1	95.219	100	2.12E-112
sq113	Oscillo_5	98.406	100	2.94E-123
sq10	Oscillo_5	96.414	100	3.35E-116
sq65	Oscillo_5	95.618	100	1.74E-113
sq65	Oscillo_8	94.422	100	1.10E-109
sq10	Oscillo_8	94.024	100	4.66E-108
sq146	Oscillo_8	92.032	100	1.52E-101
sq27	Pepto_2	98.008	100	1.25E-121
sq182	Pepto_2	96.016	100	4.09E-115
sq65	Pepto_2	89.243	100	9.01E-92

Tables **S15** and **S16** can be found as separate supplementary excel tables.

Table S17 MTI gene counts per Clostridia MAG.

MAG	Name	MTI counts
E4_d458_spades_bin_53_orig-contigs	Aceti_1	1
G3_d458_megahit_bin_18_orig-contigs	Aceti_2	9
G3_d458_spades_bin_18_indstr_refined-contigs	Aceti_3	10
G3_d458_spades_bin_33_strict_refined-contigs	Aceti_4	9
L3_d458_spades_bin_6_orig_refined-contigs	Aceti_5	22
E4_d458_spades_bin_64_permissive_refined-contigs	Ch29_1	0
G3_d458_megahit_bin_21_strict_refined-contigs	Eubac_1	65
L3_d458_spades_bin_34_indper_refined-contigs	Eubac_2	33
Van_BES_megahit_bin_42_indstr_refined-contigs	Eubac_3	46
G3_d458_spades_bin_19_indper_refined-contigs	Eubac_4	11
Co_bin_102_permissive_refined-contigs	Oscillo_1	52
E4_d458_spades_bin_8_permissive_refined-contigs	Oscillo_2	12
L3_d458_megahit_bin_13_permissive_refined-contigs	Oscillo_3	11
G3_d458_megahit_bin_13_indstr_refined-contigs	Oscillo_4	40
L3_d458_spades_bin_8_indstr_refined-contigs	Oscillo_5	34
M4_d458_spades_bin_15_permissive_refined-contigs	Oscillo_6	37
L3_d458_spades_bin_21_orig_refined-contigs	Oscillo_7	18
M4_d458_megahit_bin_49_permissive_refined-contigs	Oscillo_8	0
Co_bin_88_orig_refined-contigs	Pepto_1	0
L3_d458_megahit_bin_58_indper-contigs	Pepto_2	2
E4_d458_megahit_bin_7_strict_refined-contigs	Pepto_3	1
M4_d458_megahit_bin_14_orig_refined-contigs	Pepto_4	0
M4_d458_spades_bin_28_indstr_refined-contigs	Pepto_5	0
Co_bin_49_orig_refined-contigs	Tissi_1	5
L3_d458_spades_bin_31_orig_refined-contigs	Tissi_2	0
L3_d458_spades_bin_64_permissive_refined-contigs	Clostridia_1	0

Tables **S18**, **S19** and **S20** can be found as separate supplementary excel tables.

Table S21 Exact masses used for the MAC inclusion lists for the DIA-MS². The exact mass was computed using the molar mass plus the mass of a proton (1.0073 Da). Due to single protonation the charge (z) is 1.

Substrate	M [g/mol]	m/z (M+H)
3,5-dimethoxybenzoic acid	182.17	183.18
Vanillin	152.15	153.16
2-methoxyphenol	124.14	125.15
3,4,5-trimethoxyphenol	184.19	185.20
2-methoxybenzoic acid	152.15	153.16
3,4,5-trimethoxybenzoic acid	212.07	213.08
3-methoxyphenylacetic acid	166.18	167.19
3,4,5-trimethoxyphenylacetic acid	226.23	227.24
3-methoxycinnamic acid	178.19	179.20
3,4-dimethoxycinnamic acid	208.21	209.22
3,4,5-trimethoxycinnamic acid	238.24	239.25
2-methoxyhydroquinone	140.14	141.15
1,2,4-trimethoxybenzene	168.19	169.20
3,4,5-trimethoxybenzaldehyde	196.20	197.21
2-methoxy-3-pyridinylboranic acid	152.94	153.95

4.2.4 Supplementary Figures

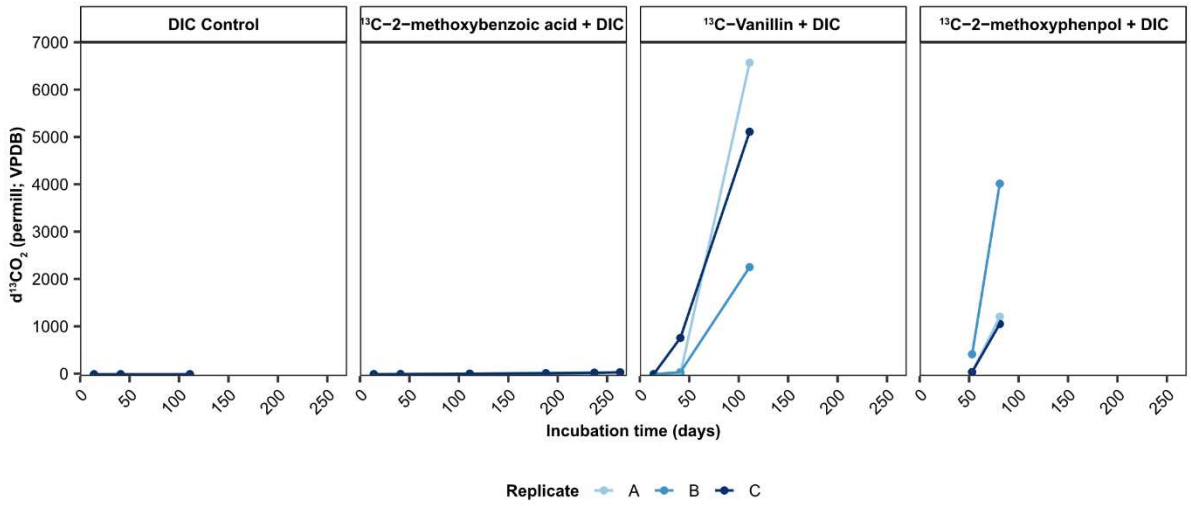

Figure S1 $\delta^{13}C$ values of CO_2 in the headspace of triplicate RNA-SIP enrichments of the unlabeled DIC control, ^{13}C-2-methoxybenzoic acid + unlabeled DIC, ^{13}C-Vanillin + unlabeled DIC and ^{13}C-2-methoxyphenol + unlabeled DIC samples over 263 days. The $\delta^{13}C$ values were monitored as $\delta^{13}C$ (‰) relative to the Vienna Pee Dee Belemnite (VPDB) standard. Triplicates were set up for each treatment, containing a combination of ^{13}C-labeled MACs (Vanillin, 2-methoxybenzoic acid, 2-methoxyphenol) and unlabeled bicarbonate (i.e., dissolved inorganic carbon - DIC). DIC was supplied in a concentration of 10 mM, MACs were supplied in a concentration of 1 mM.

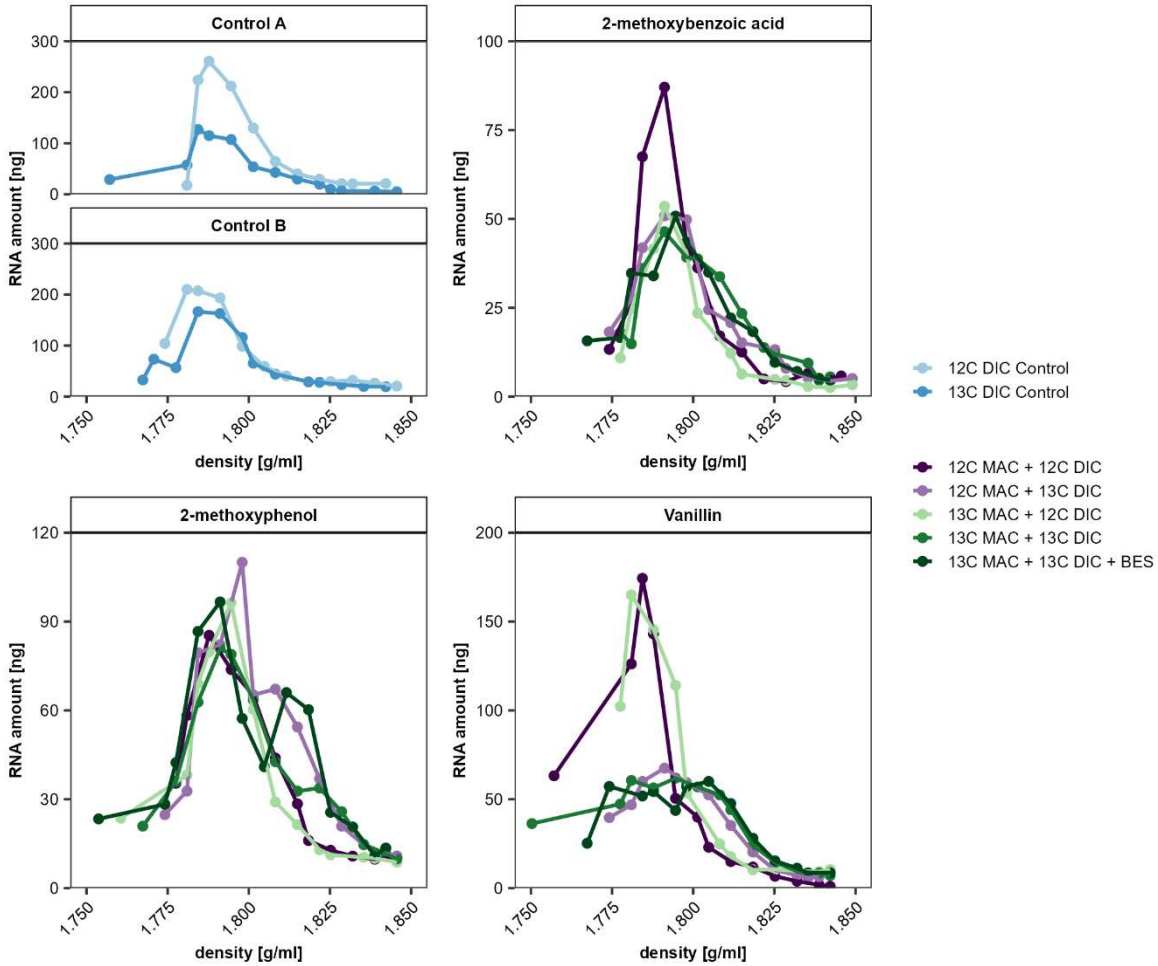

Figure S2 Density separation profiles of MAC amended and control RNA-SIP enrichments. 12C corresponds to the unlabeled substrates, 13C to the labeled substrates. Control A was stopped after 111 days, and control B after 188 days.

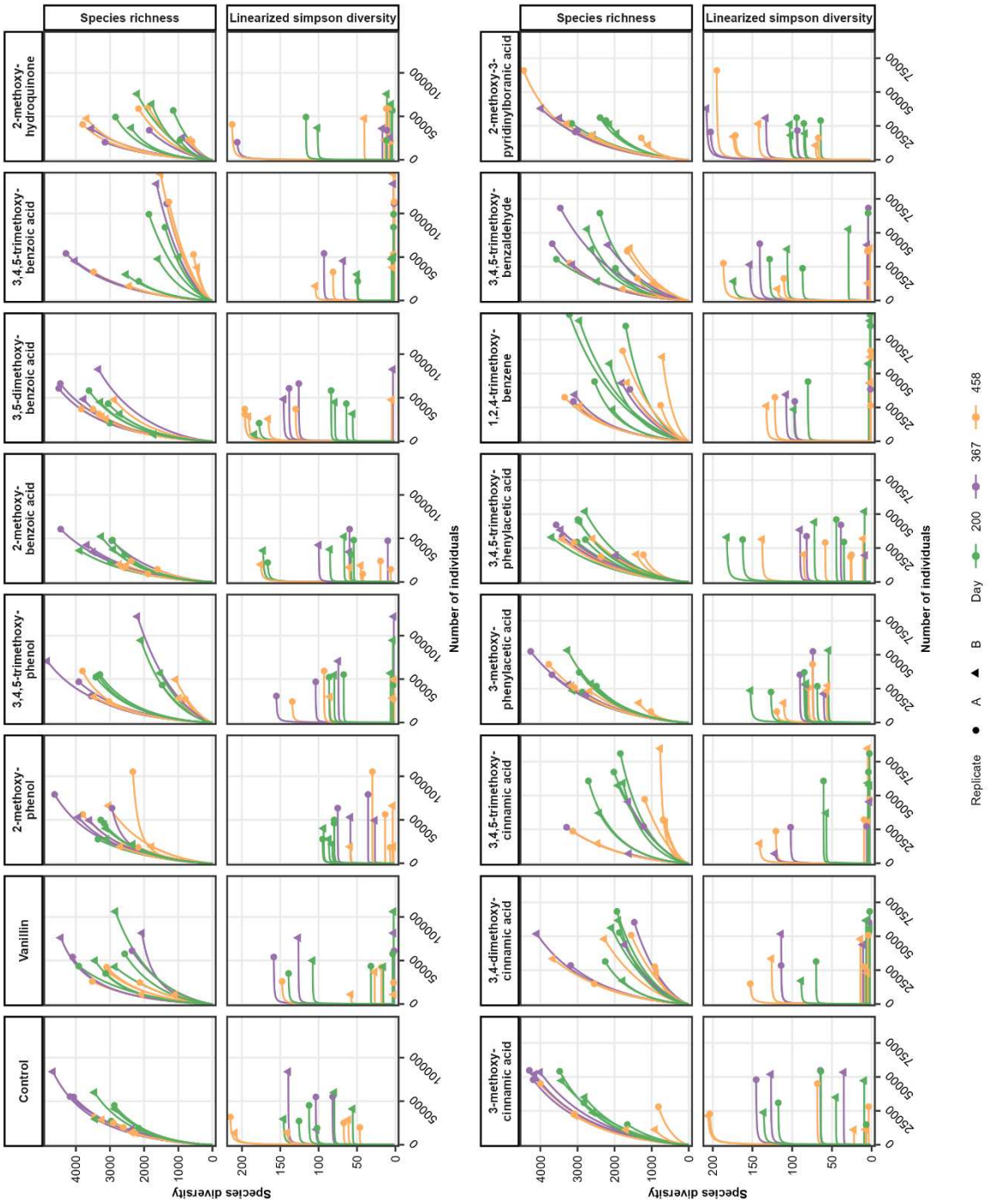

Figure S3 Species richness and linearized Simpson diversity of bacterial samples of the MAC enrichments for single targeted MACs on days 200, 367 and 458 for two biological replicates A and B. Enrichments contain either 12 mM of one of the selected methoxylated aromatic compounds (MACs) or no additional carbon substrate as control. Replicates were additionally amended with 5 mM sodium 2-bromoethanesulfonate (BES) to inhibit methanogenesis.

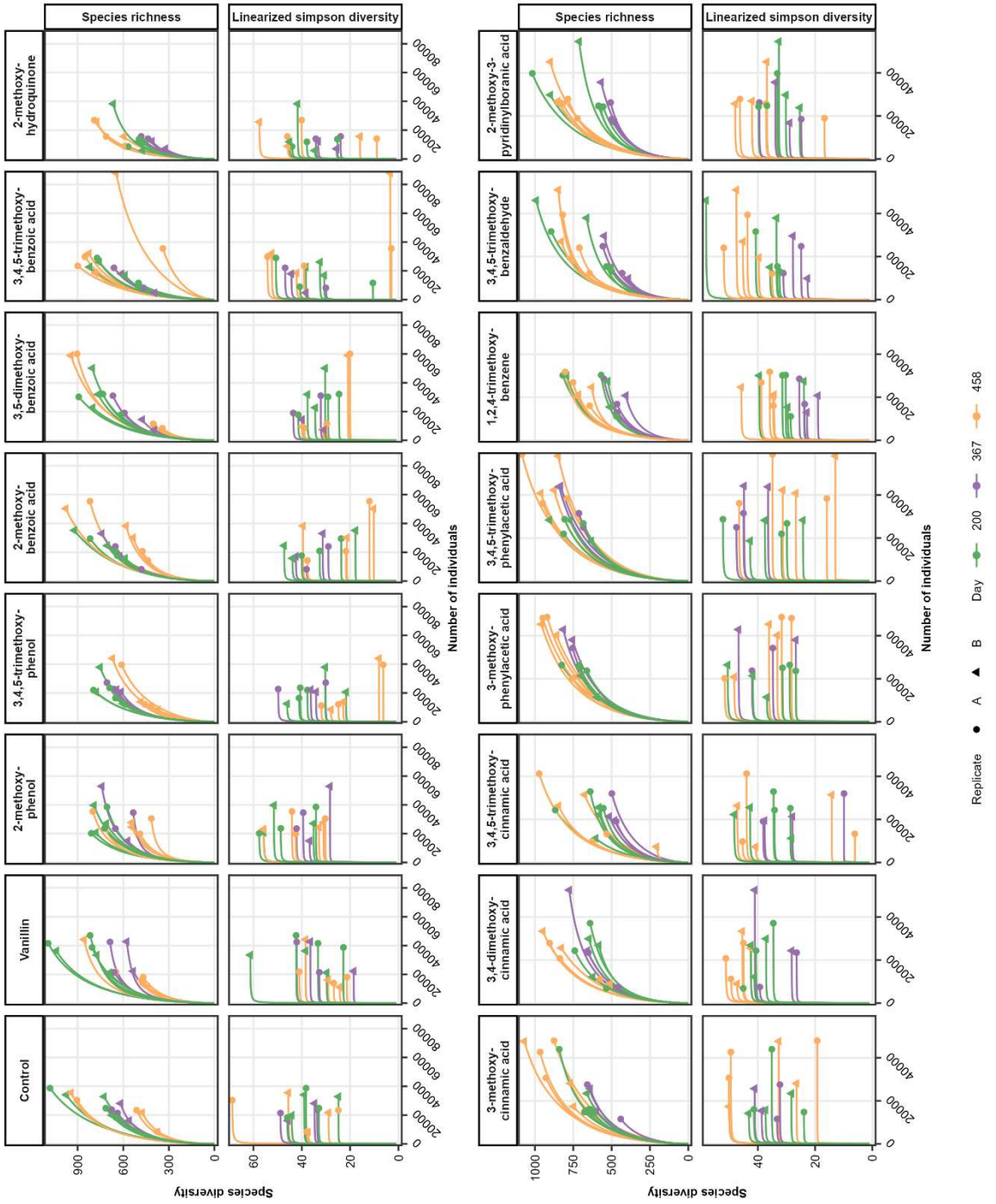

Figure S4 Species richness and linearized Simpson diversity of archaeal samples of the MAC enrichments for single targeted MACs on days 200, 367 and 458 for two biological replicates A and B. Enrichments contain either 12 mM of one of the selected methoxylated aromatic compounds (MACs) or no additional carbon substrate as control. Replicates were additionally amended with 5 mM sodium 2-bromoethanesulfonate (BES) to inhibit methanogenesis.

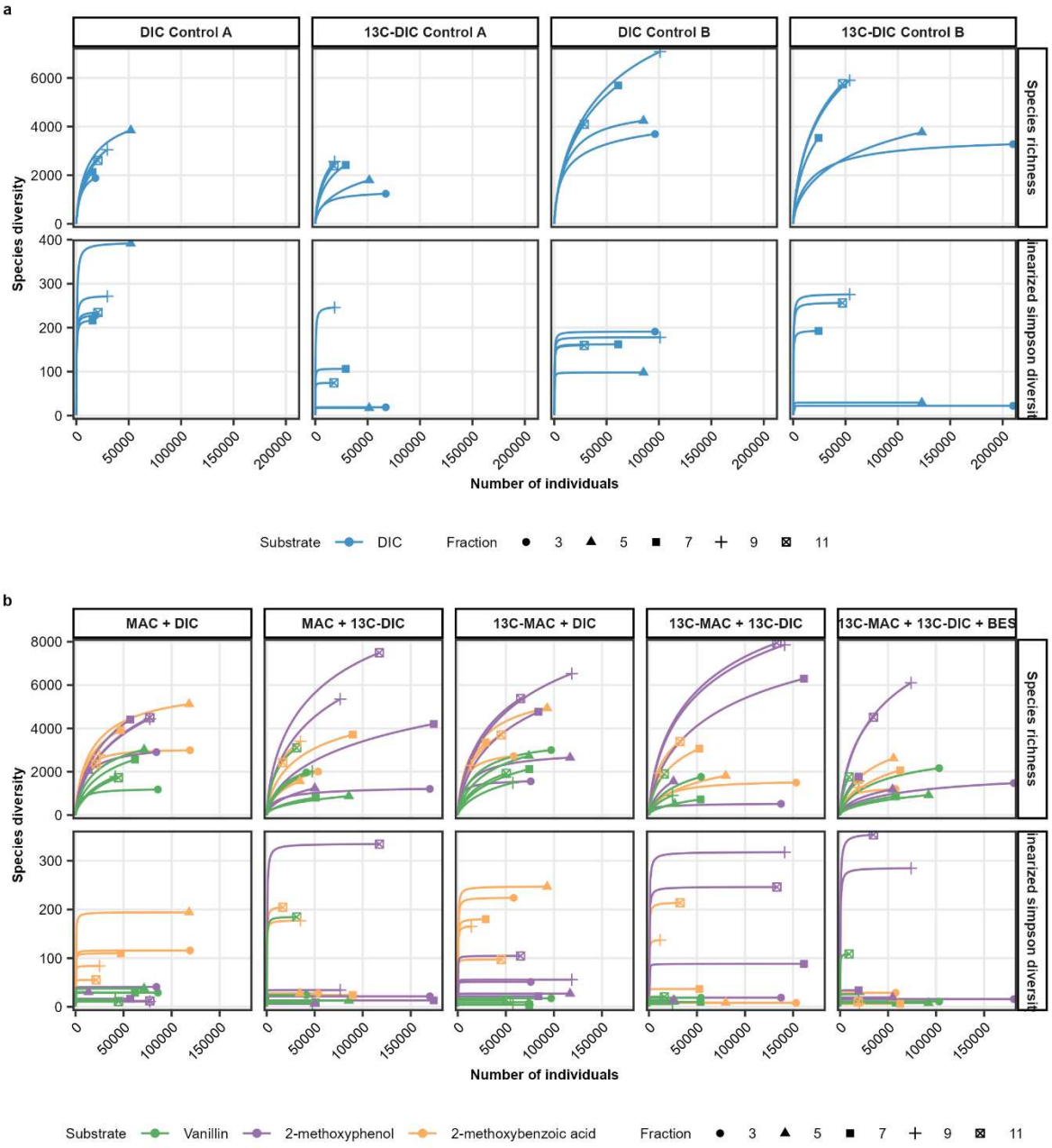

Figure S5 Species richness and linearized Simpson diversity of bacterial samples of the RNA-SIP enrichments of single targeted treatments for fractions 3 (ultra-heavy), 5 (heavy), 7 (midpoint), 9 (light) and 11 (ultra-light) for (**a**) DIC controls and (**b**) SIP incubations amended with methoxylated aromatic compounds (MACs). MAC refers to any of the substrates Vanillin, 2-methoxyphenol or 2-methoxybenzoic acid, indicated by color. Triplicates were set up for each treatment, containing a combination of ^{13}C-labeled and unlabeled MACs (Vanillin, 2-methoxybenzoic acid, 2-methoxyphenol) and bicarbonate (i.e., dissolved inorganic carbon - DIC) or only ^{13}C-labeled or unlabeled DIC as control (Table S2). DIC was supplied in a concentration of 10 mM, MACs were supplied in a concentration of 1 mM. One treatment for each of the tested MACs was additionally amended with 5 mM sodium 2-bromoethanesulfonate (BES) for the inhibition of methanogenic activity.

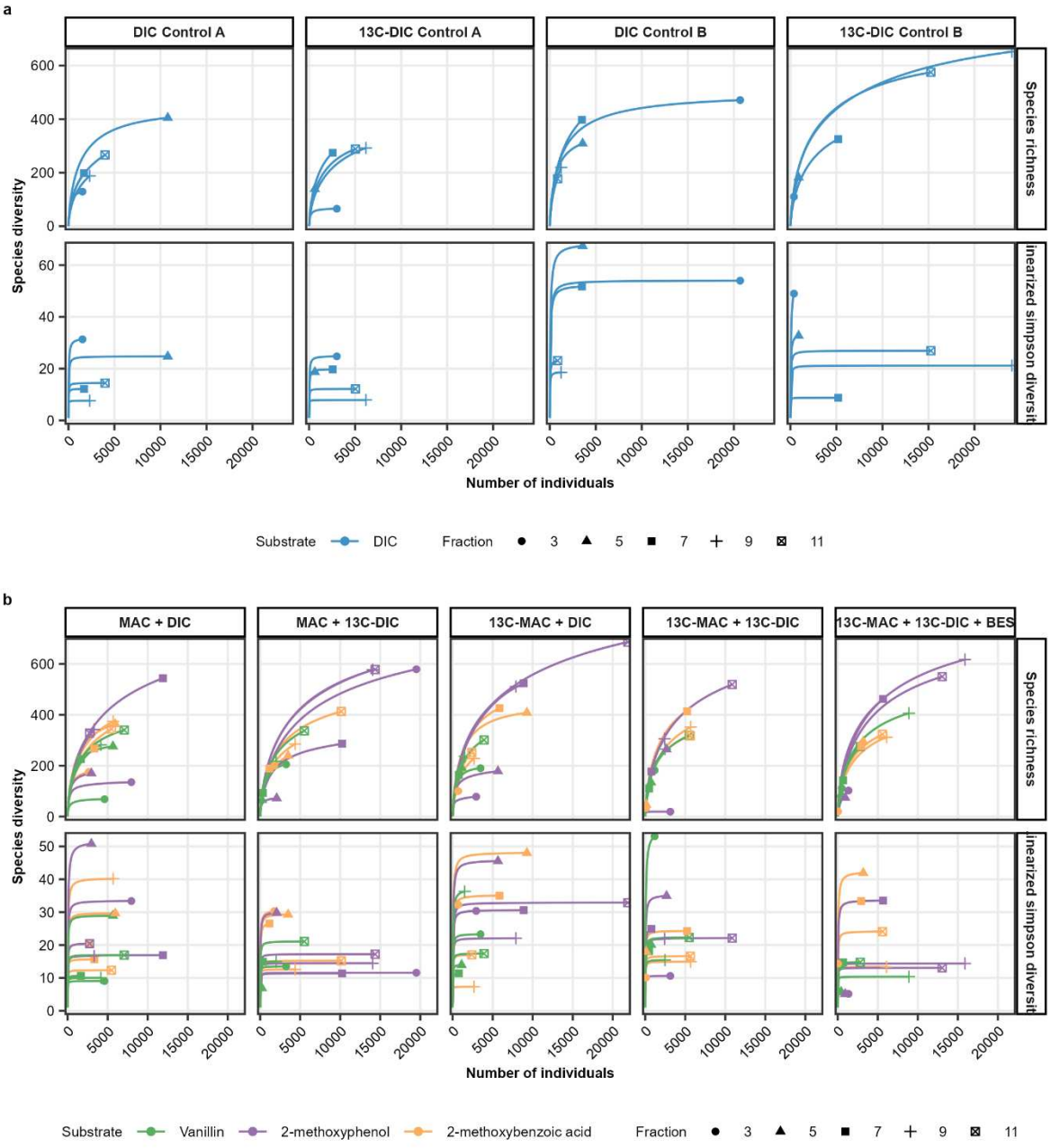

Figure S6 Species richness and linearized Simpson diversity of archaeal samples of the RNA-SIP enrichments of single targeted treatments for fractions 3 (ultra-heavy), 5 (heavy), 7 (midpoint), 9 (light) and 11 (ultra-light) for (**a**) DIC controls and (**b**) SIP incubations amended with methoxylated aromatic compounds (MACs). MAC refers to any of the substrates Vanillin, 2-methoxyphenol or 2-methoxybenzoic acid, indicated by color. Triplicates were set up for each treatment, containing a combination of ^{13}C-labeled and unlabeled MACs (Vanillin, 2-methoxybenzoic acid, 2-methoxyphenol) and bicarbonate (i.e., dissolved inorganic carbon - DIC) or only ^{13}C-labeled or unlabeled DIC as control (Table S2). DIC was supplied in a concentration of 10 mM, MACs were supplied in a concentration of 1 mM. One treatment for each of the tested MACs was additionally amended with 5 mM sodium 2-bromoethanesulfonate (BES) for the inhibition of methanogenic activity.

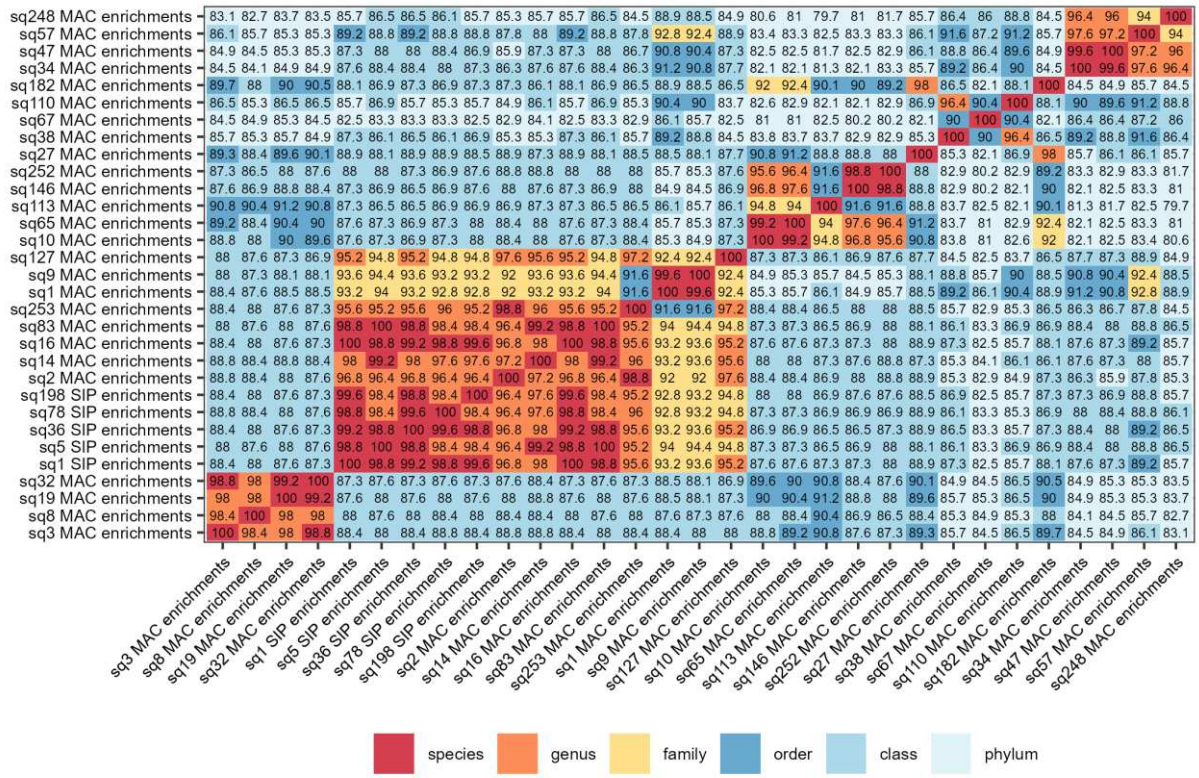

Figure S7 Comparison of 16S rRNA gene sequence identity of ASVs from the SIP incubations and MAC enrichments. Taxonomic level thresholds are indicated by color after the taxon level distinction for 16S rRNA genes by Konstantinidis et al. (2017).

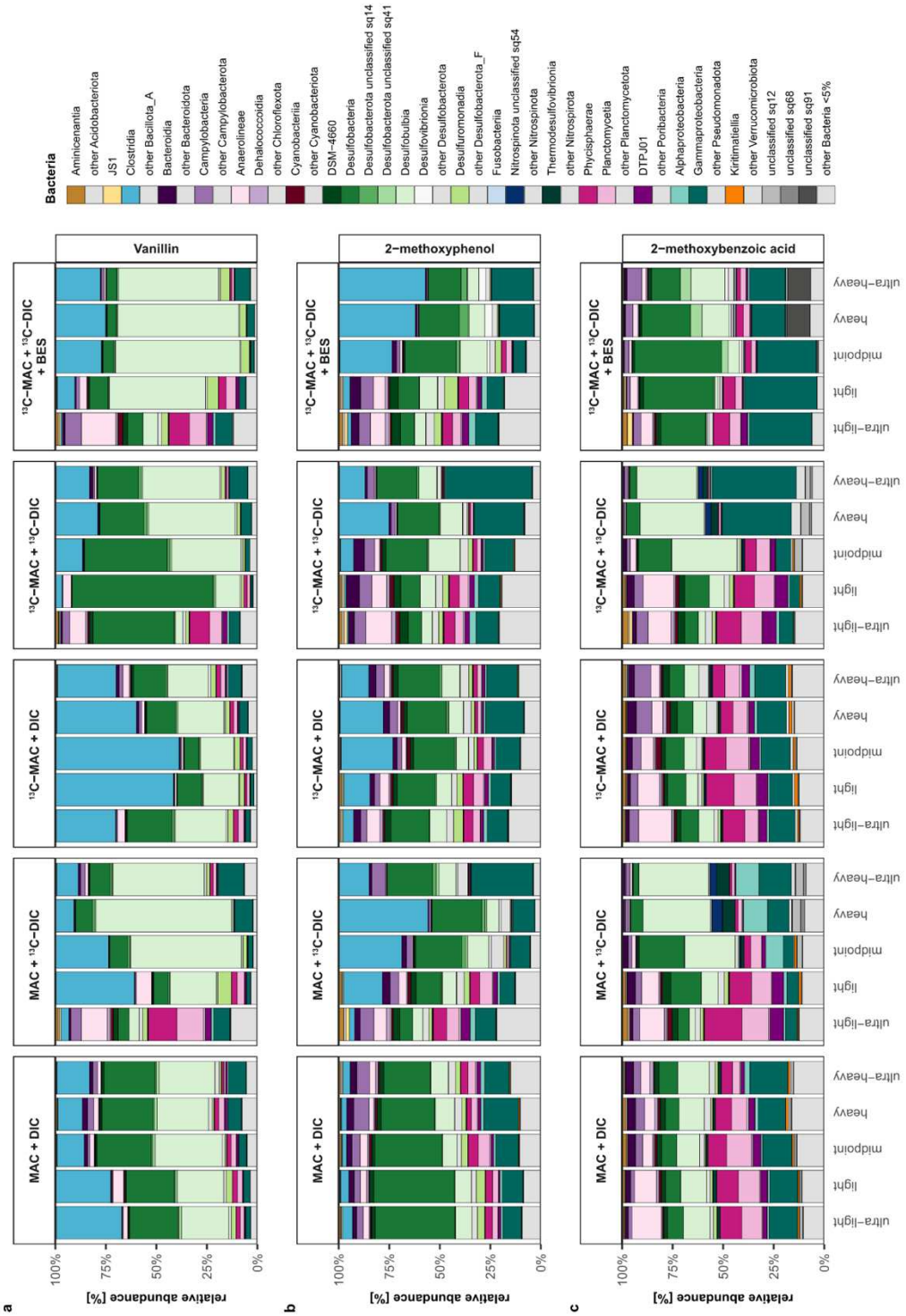

Figure S8 Relative 16S rRNA gene abundance of bacteria in RNA-SIP enrichments of (**a**) Vanillin, (**b**) 2-methoxyphenol and (**c**) 2-methoxybenzoic acid amended samples. MAC refers to the used methoxylated aromatic compound. Triplicates were set up for each treatment, containing a combination of ^{13}C-labeled and unlabeled MACs (Vanillin, 2-methoxybenzoic acid, 2-methoxyphenol) and bicarbonate (i.e., dissolved inorganic carbon - DIC). DIC was supplied in a concentration of 10 mM, MACs were supplied in a concentration of 1 mM. One treatment for each of the tested MACs was additionally amended with 5 mM sodium 2-bromoethanesulfonate (BES) to inhibit methanogenic activity. Triplicates were pooled before density separation. For each treatment the RNA was separated into the fractions ultra-light, light, midpoint, heavy and ultra-heavy.

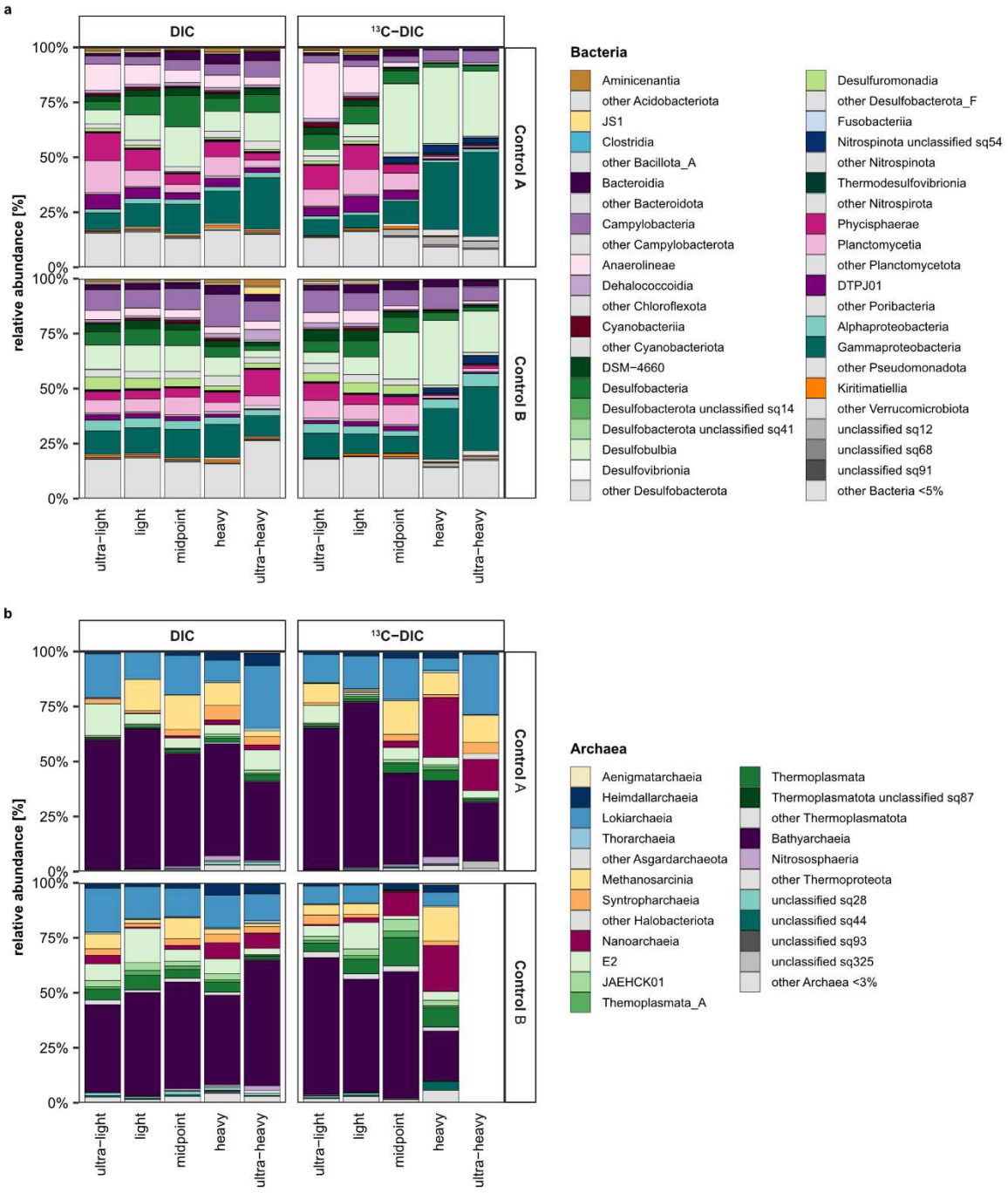

Figure S9 Relative 16S rRNA gene abundance of (**a**) bacteria and (**b**) archaea in RNA-SIP enrichments of DIC only amended control samples. Samples were amended with ^{13}C-labeled or unlabeled DIC. Control A was stopped after 111 days, and control B after 188 days. Triplicates were pooled before density separation. For each treatment the RNA was separated into the fractions ultra-light, light, midpoint, heavy and ultra-heavy. Missing samples were removed due to low community coverage.

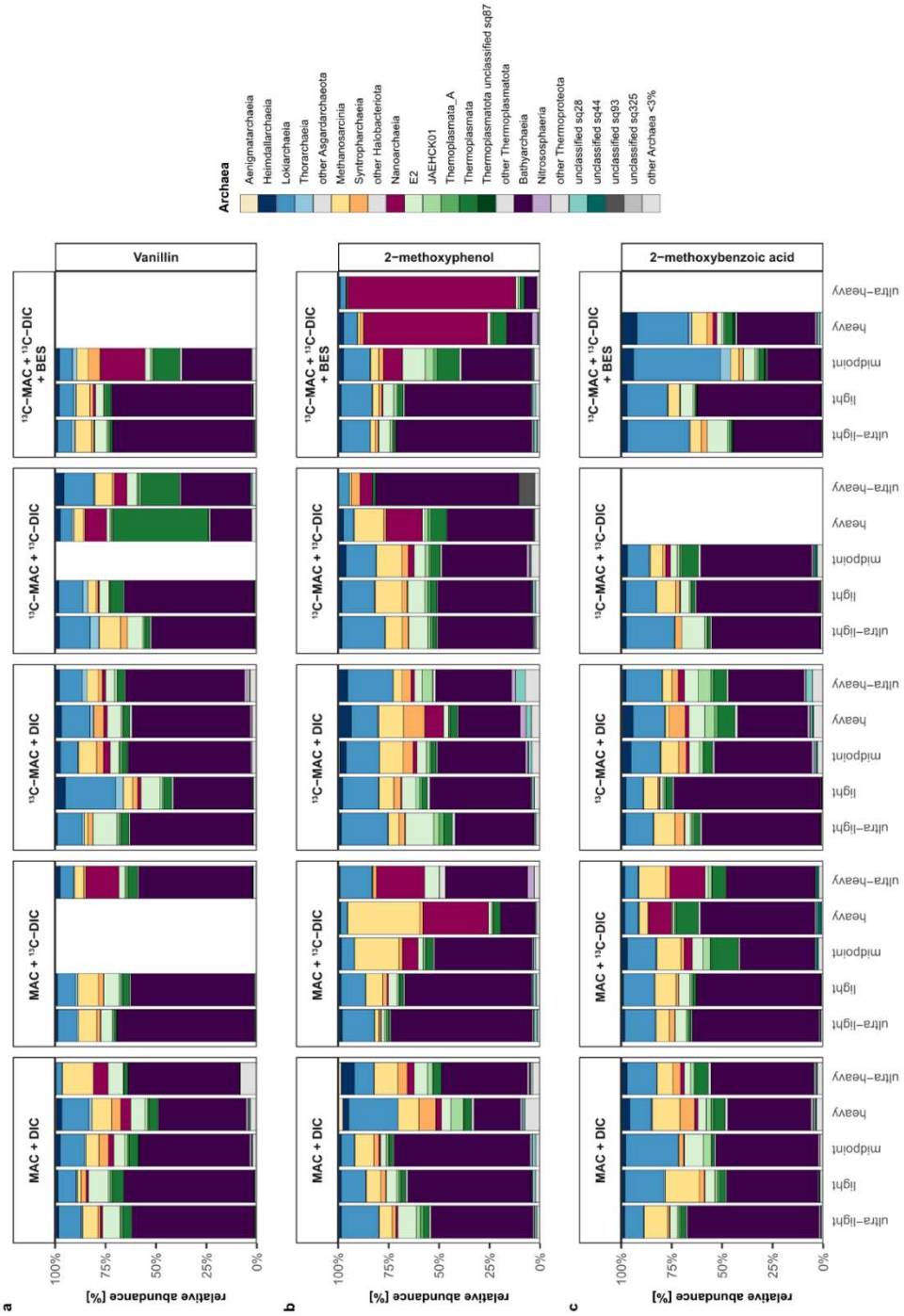

Figure S10 Relative 16S rRNA gene abundance of archaea in RNA-SIP enrichments of (**a**) Vanillin, (**b**) 2-methoxyphenol and (**c**) 2-methoxybenzoic acid amended samples. MAC refers to the used methoxylated aromatic compound. Triplicates were set up for each treatment, containing a combination of ^{13}C-labeled and unlabeled MACs (Vanillin, 2-methoxybenzoic acid, 2-methoxyphenol) and bicarbonate (i.e., dissolved inorganic carbon - DIC). DIC was supplied in a concentration of 10 mM, MACs were supplied in a concentration of 1 mM. One treatment for each of the tested MACs was additionally amended with 5 mM sodium 2-bromoethanesulfonate (BES) to inhibit methanogenic activity. Triplicates were pooled before density separation. For each treatment the RNA was separated into the fractions ultra-light, light, midpoint, heavy and ultra-heavy. Missing samples were removed due to low community coverage.

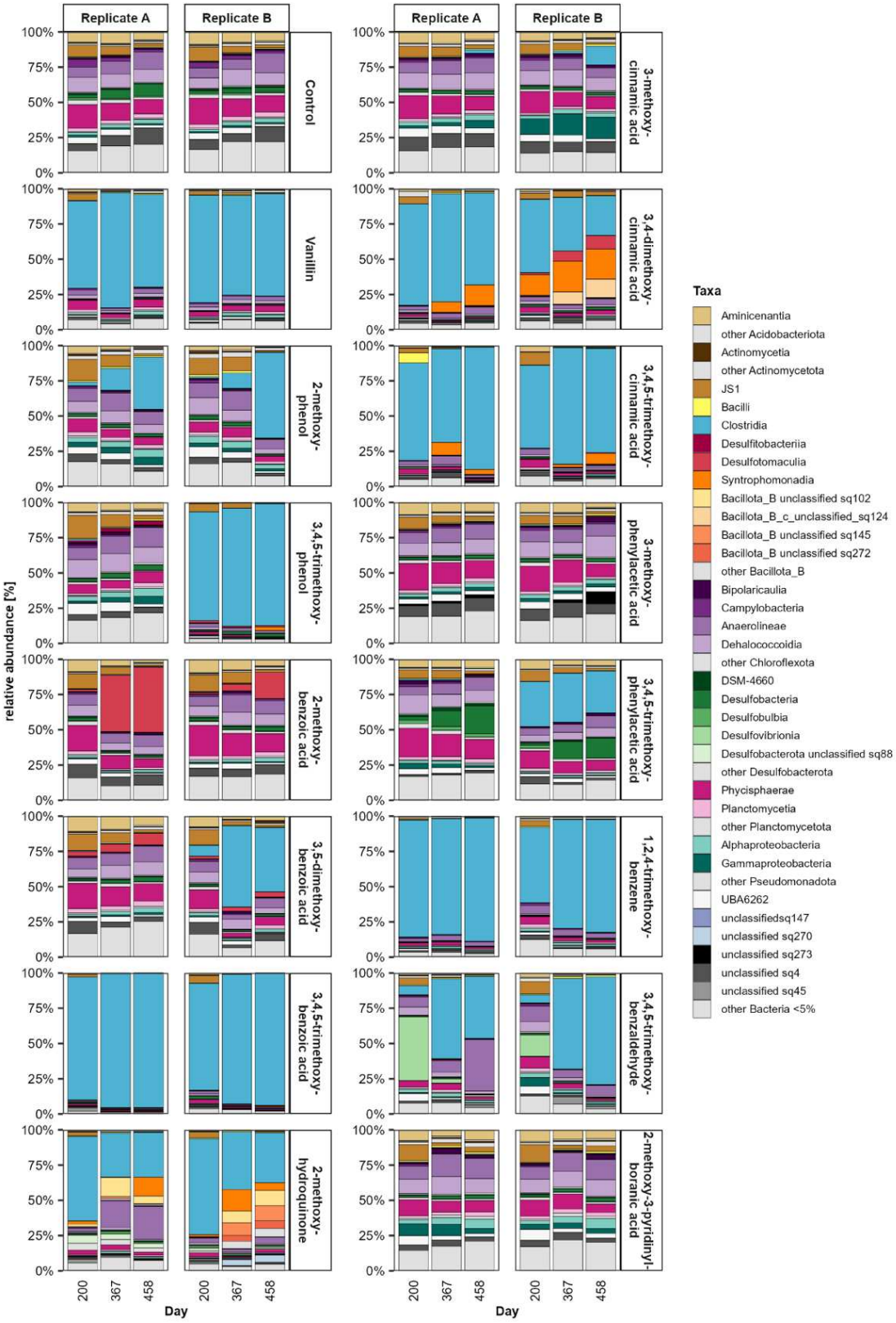

Figure S11 Relative 16S rRNA gene abundance of bacteria on days 200, 367 and 458 in two replicates A and B for single methoxylated aromatic compounds (MACs) treatments. Enrichments contain either 12 mM of one of the selected methoxylated aromatic compounds (MACs) or no additional carbon substrate as control. Replicates were additionally amended with 5 mM sodium 2-bromoethanesulfonate (BES) to inhibit methanogenesis.

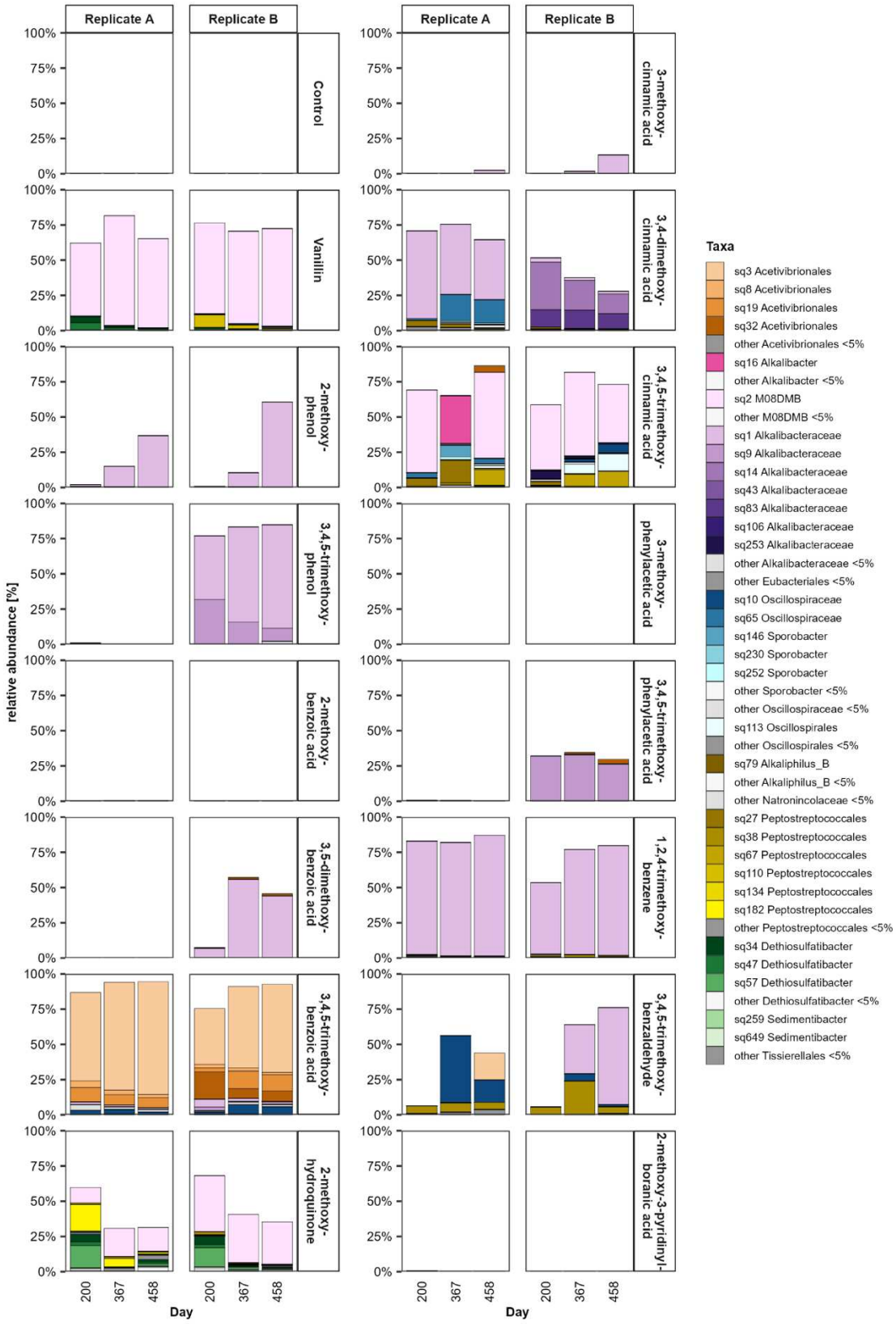

Figure S12 Relative 16S rRNA gene abundance of ASVs affiliated with the class Clostridia on days 200, 367 and 458 in two replicates A and B for single methoxylated aromatic compounds (MACs) treatments. Enrichments contain either 12 mM of one of the selected methoxylated aromatic compounds (MACs) or no additional carbon substrate as control. Replicates were additionally amended with 5 mM sodium 2-bromoethanesulfonate (BES) to inhibit methanogenesis.

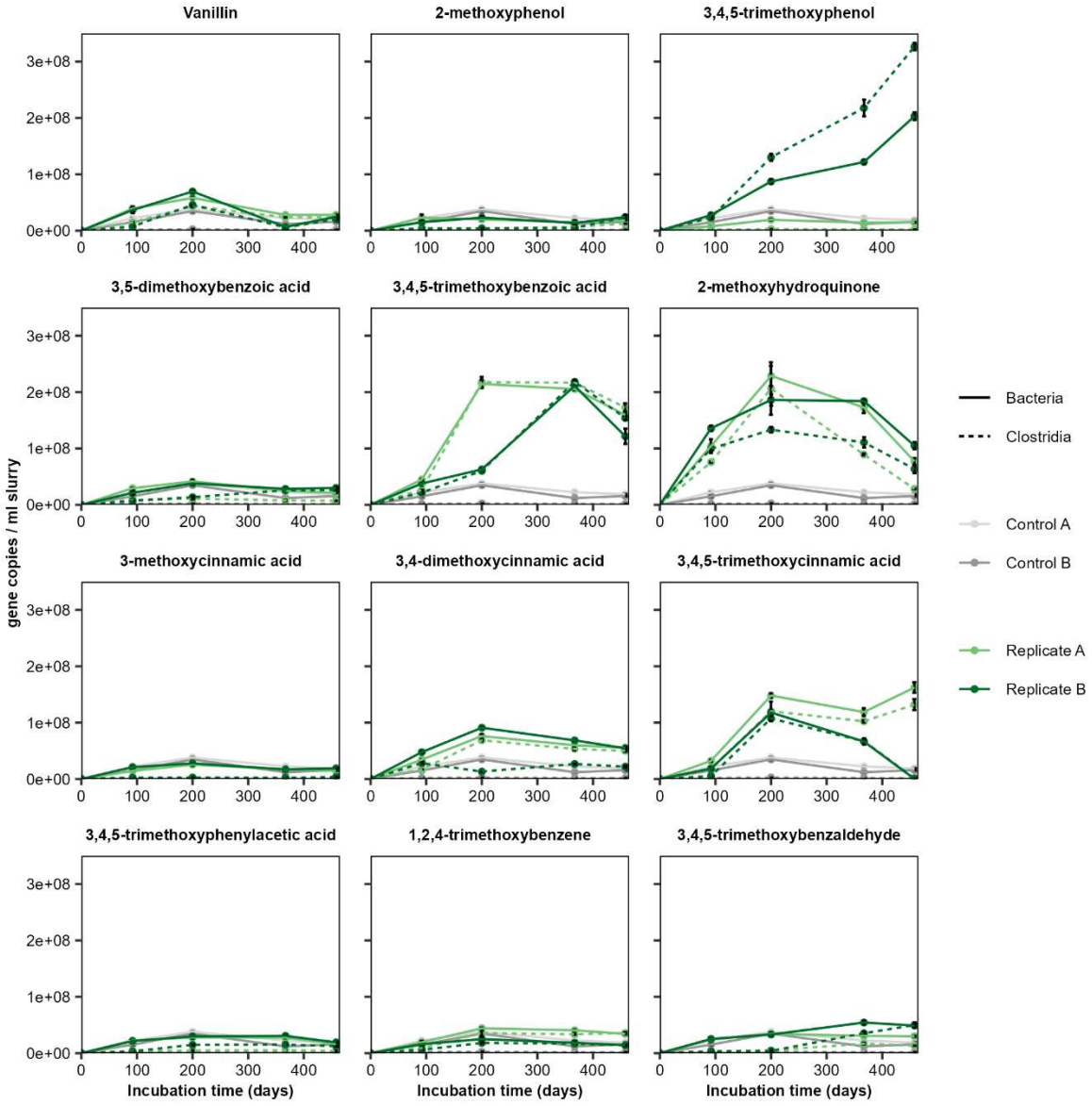

Figure S13 16S rRNA gene copies per ml slurry of all bacteria and Clostridia over 458 days in two replicates A and B in selected methoxylated aromatic compounds (MACs) enrichments. Grey-colored lines in the background show the control sample without any MAC amendment.

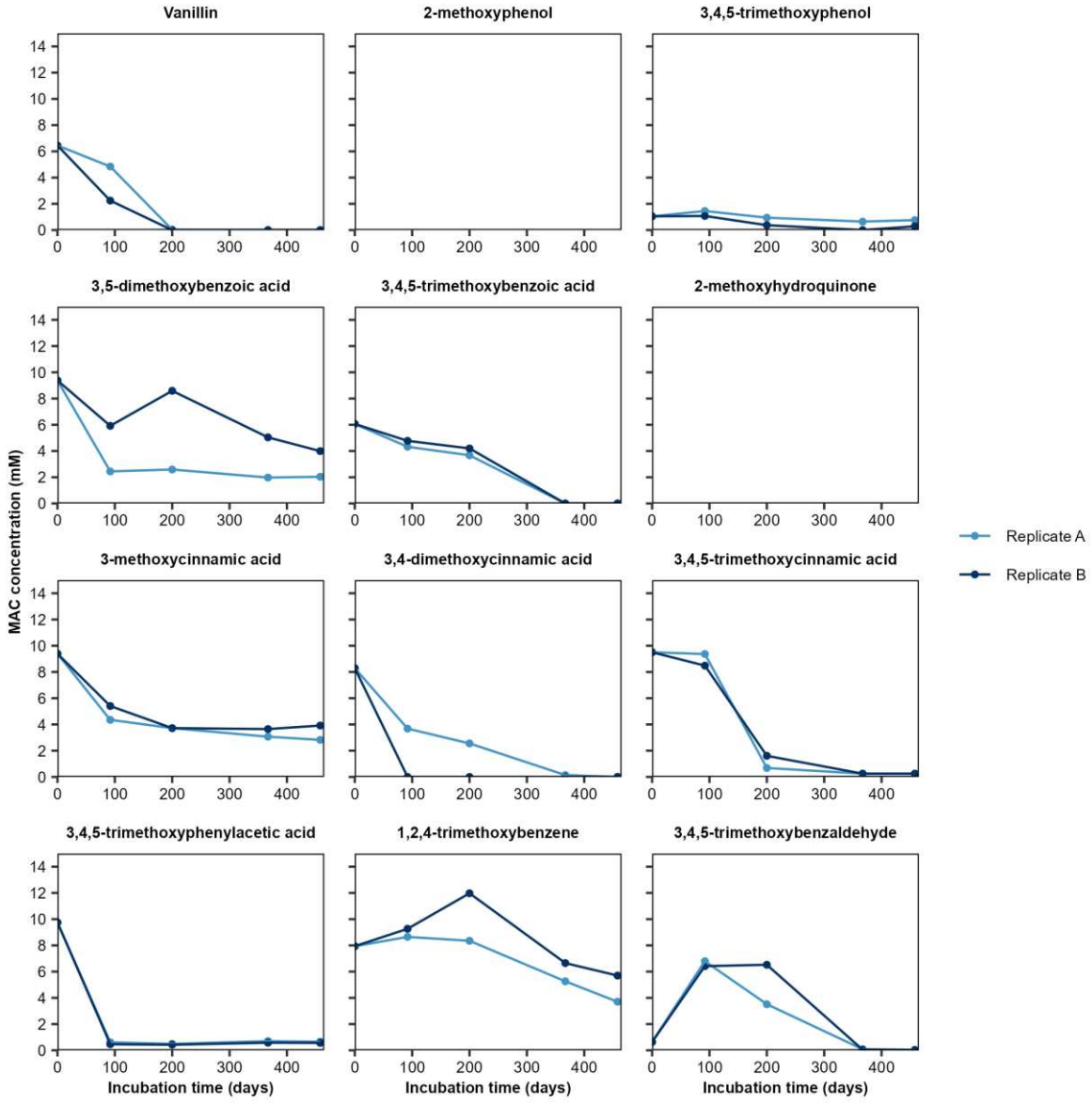

Figure S14 LC-MS/MS measurements of selected methoxylated aromatic compounds (MACs) in samples. The total MAC concentration was calculated using the concentration evaluated from the supernatant and sediment of each sample.

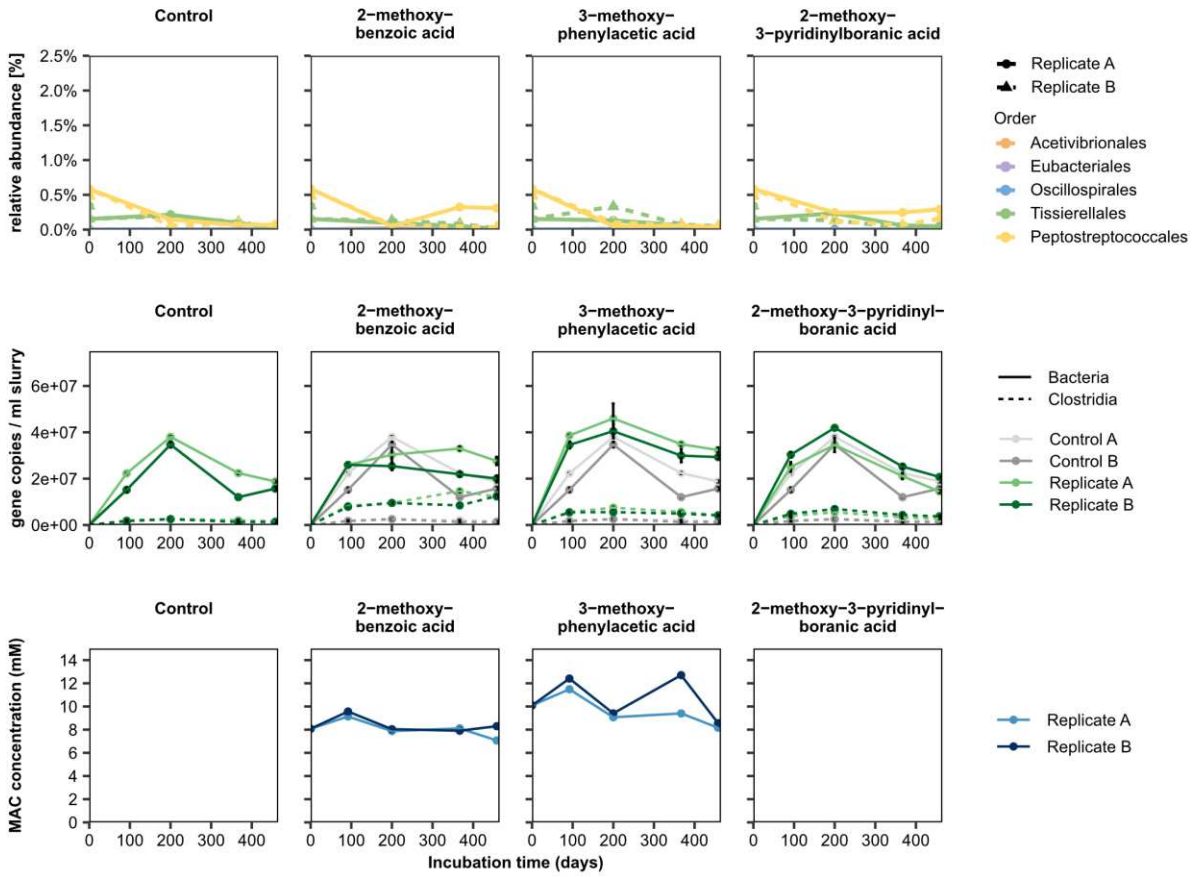

Figure S15 Relative 16S rRNA gene abundance, 16S rRNA gene copies of all bacteria and Clostridia and MAC concentration over 458 days in two replicates A and B in selected methoxylated aromatic compounds (MACs) enrichments without Clostridia enrichment. Grey-colored lines in the background show the control sample without any MAC amendment.

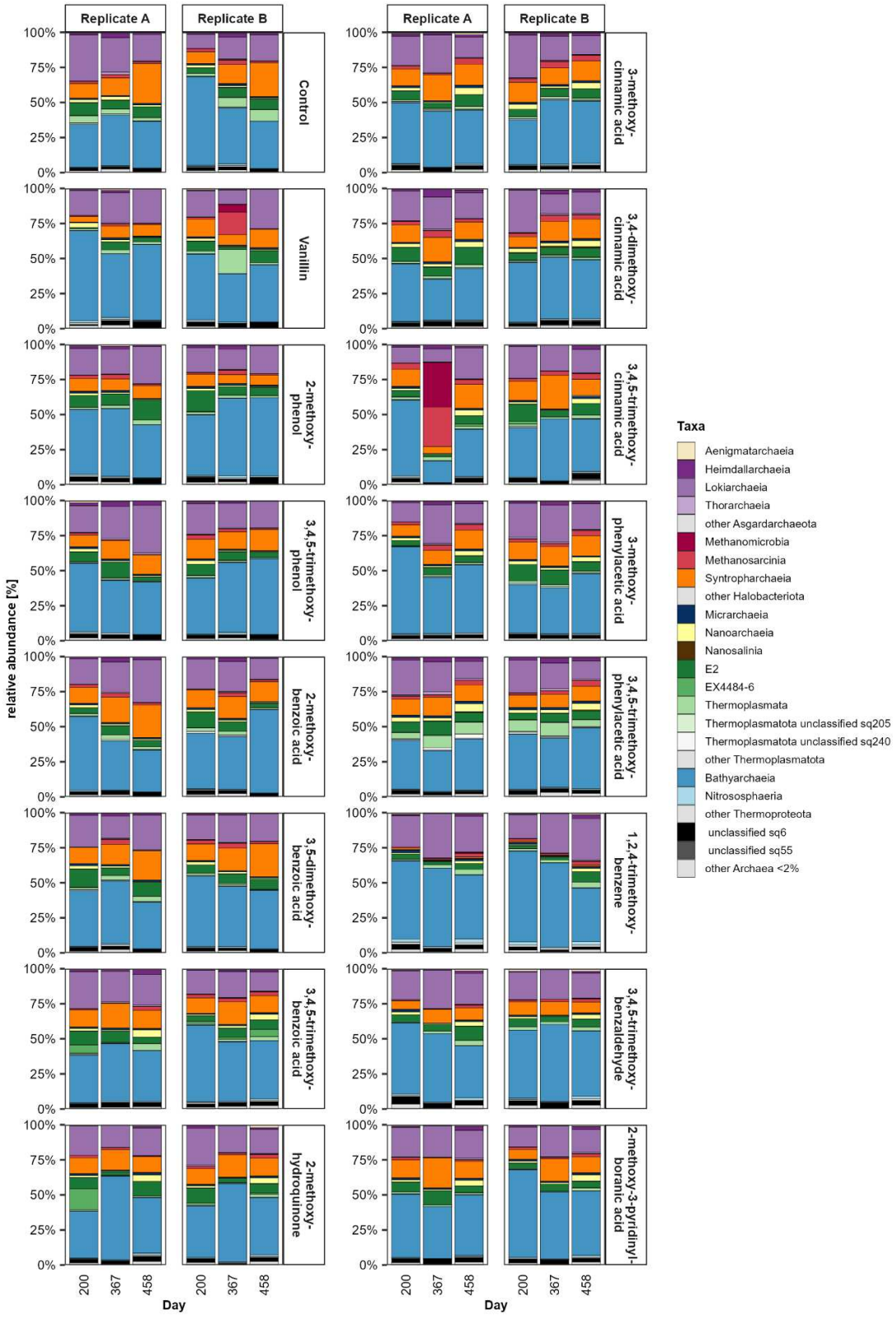

Figure S16 Relative 16S rRNA gene abundance of archaea on days 200, 367 and 458 in two replicates A and B for single methoxylated aromatic compounds (MACs) treatments. Enrichments contain either 12 mM of one of the selected methoxylated aromatic compounds (MACs) or no additional carbon substrate as control. Replicates were additionally amended with 5 mM sodium 2-bromoethanesulfonate (BES) to inhibit methanogenesis

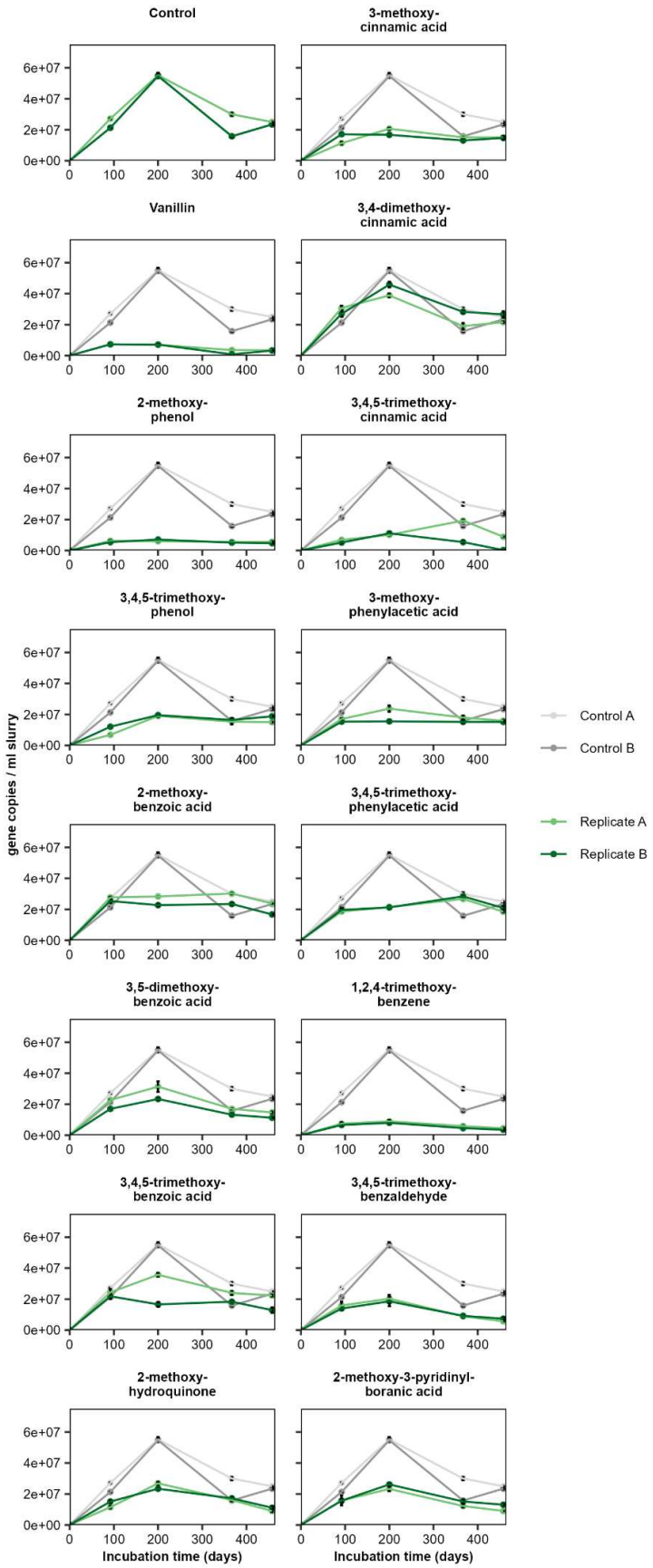

Figure S17 16S rRNA gene copies of archaea over 458 days in two replicates A and B in selected methoxylated aromatic compound (MAC) enrichments. Grey-colored lines in the background indicate the control sample without any MAC amendment.

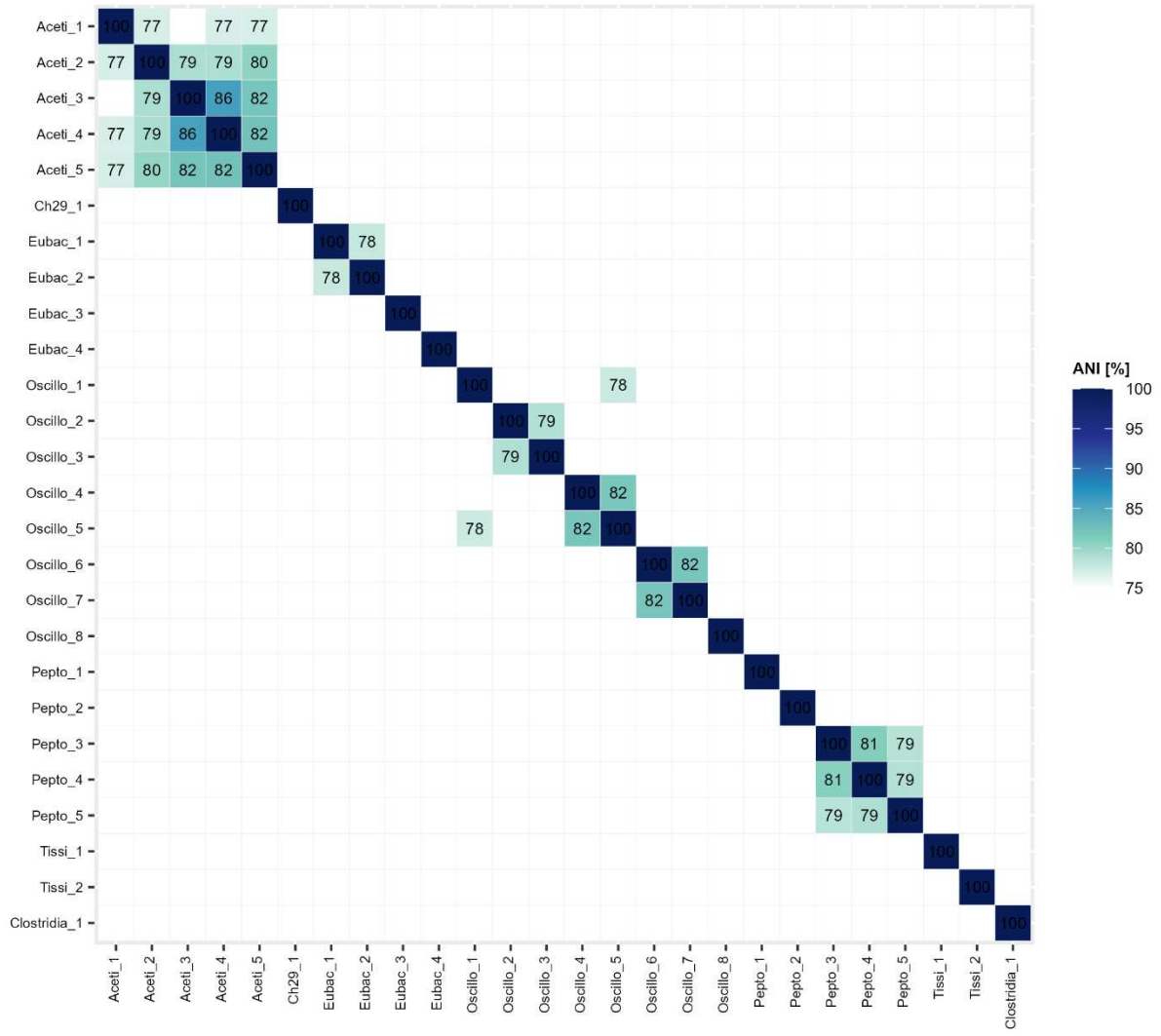

Figure S18 Heatmap of ANI between all 26 Clostridia MAGs obtained during the study. Numbers in colored tiles indicate the ANI percentage between the compared MAGs (threshold ANI > 75%).

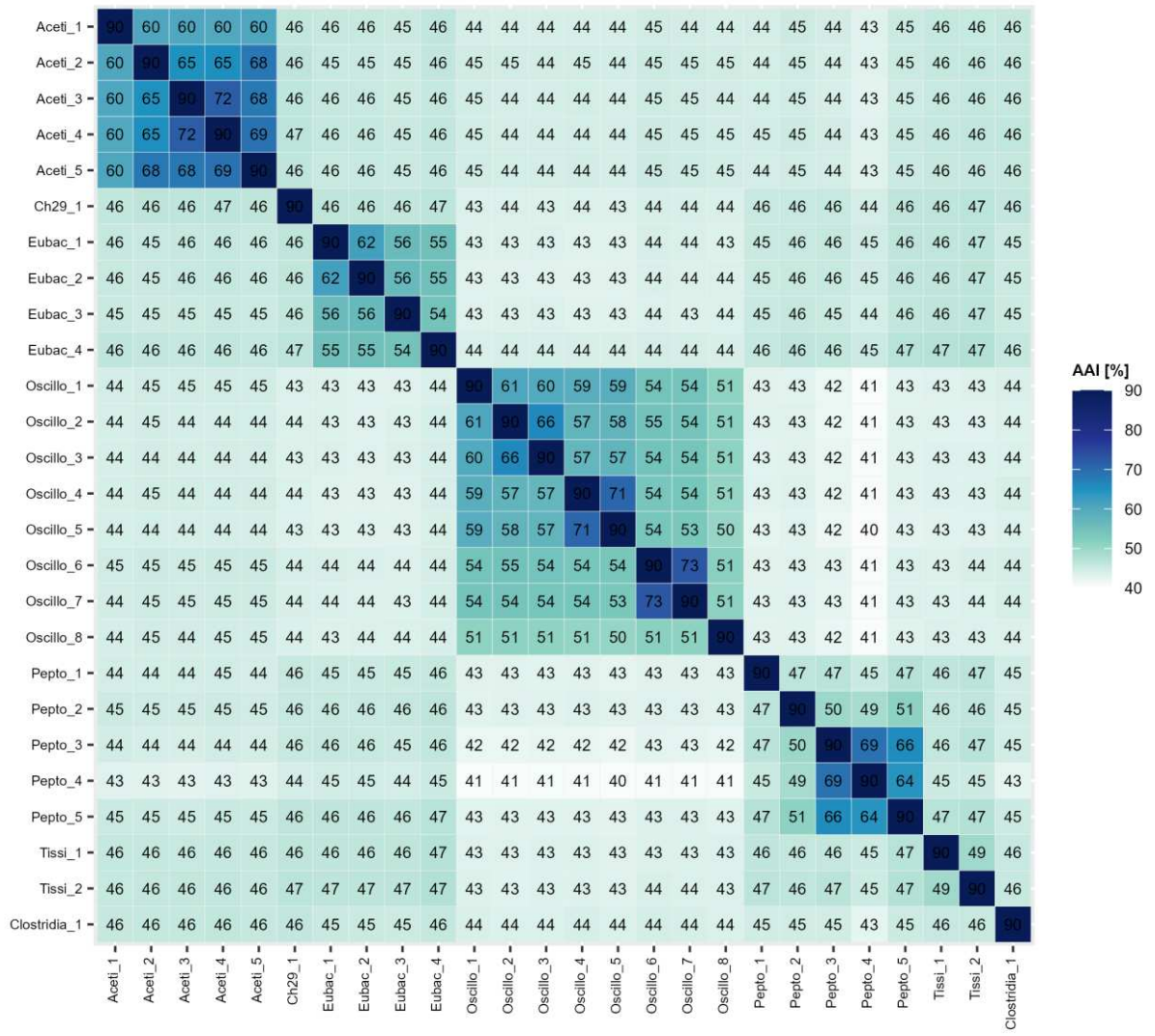

Figure S19 Heatmap of AAI between all 26 Clostridia MAGs obtained during the study. Numbers in colored tiles indicate the AAI percentage between the compared MAGs. The maximum value of 90 represents the highest values calculated by the program fastAAI (> 90%).

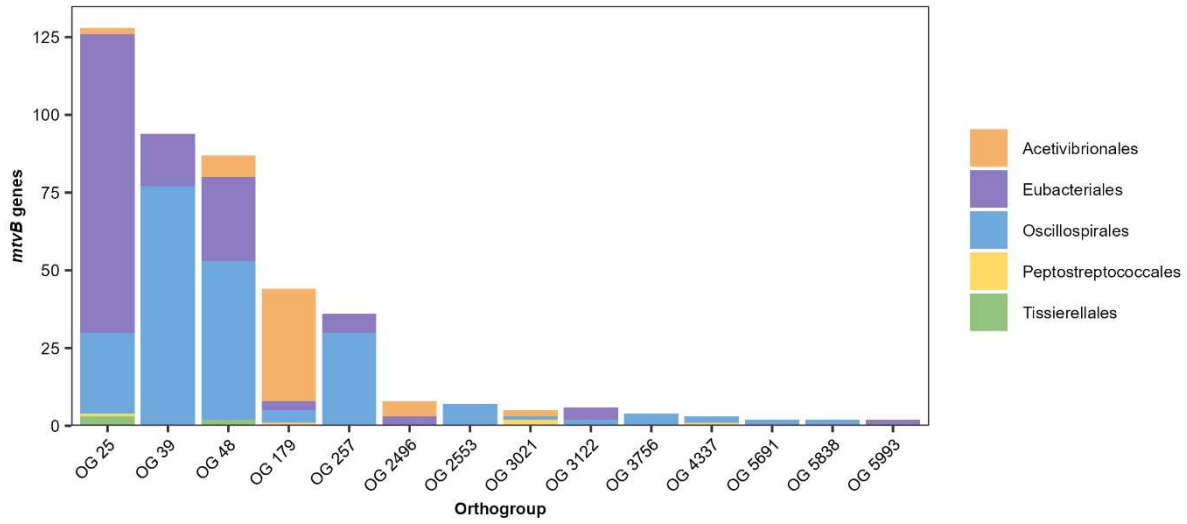

Figure S20 Gene counts of methyltransferase I (MTI) genes per orthogroup (OG). Colors indicate the order the MTI gene was detected in.

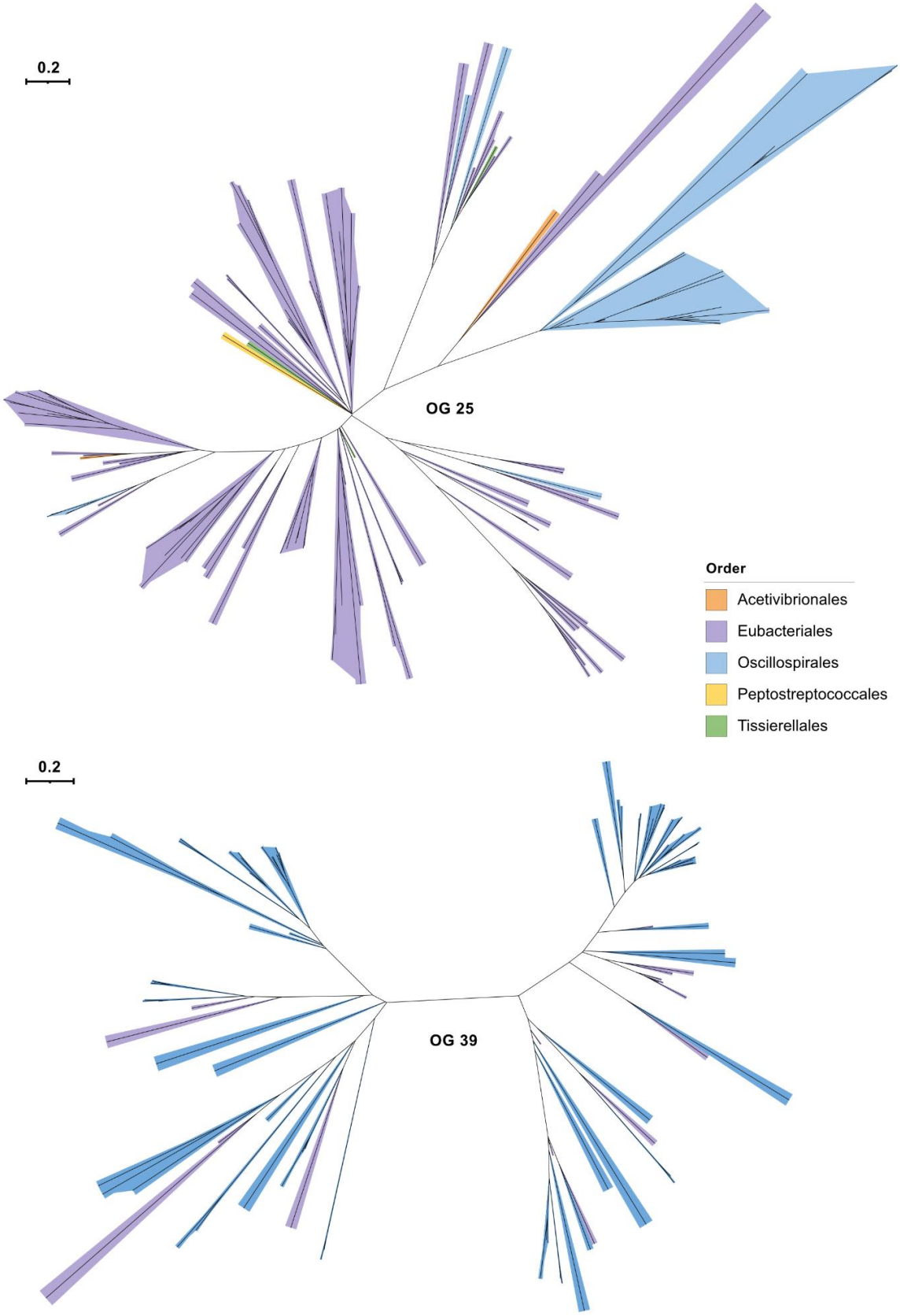

Figure S21 Unrooted orthogroup (OG) trees of the orthogroups 25 and 39 only containing methyltransferase I (MTI) genes. Colors indicate the order the MTI gene was found in. Orthogroup trees were computed using orthofinder v2.5.5 (Emms and Kelly 2019).

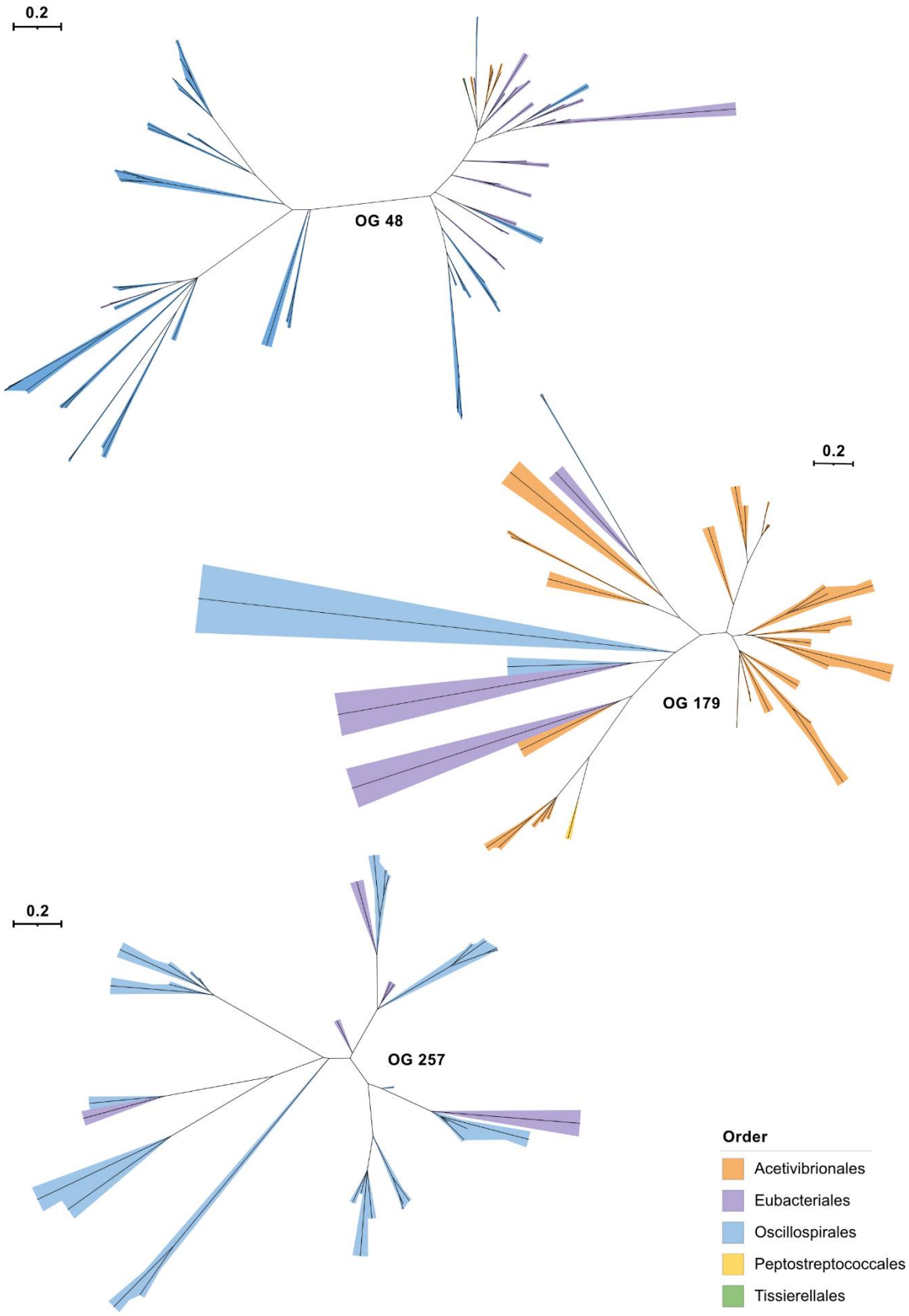

Figure S22 Unrooted orthogroup (OG) trees of the orthogroups 48, 179 and 257 only containing methyltransferase I (MTI) genes. Colors indicate the order the MTI gene was found in. Orthogroup trees were computed using orthofinder v2.5.5 (Emms and Kelly 2019).

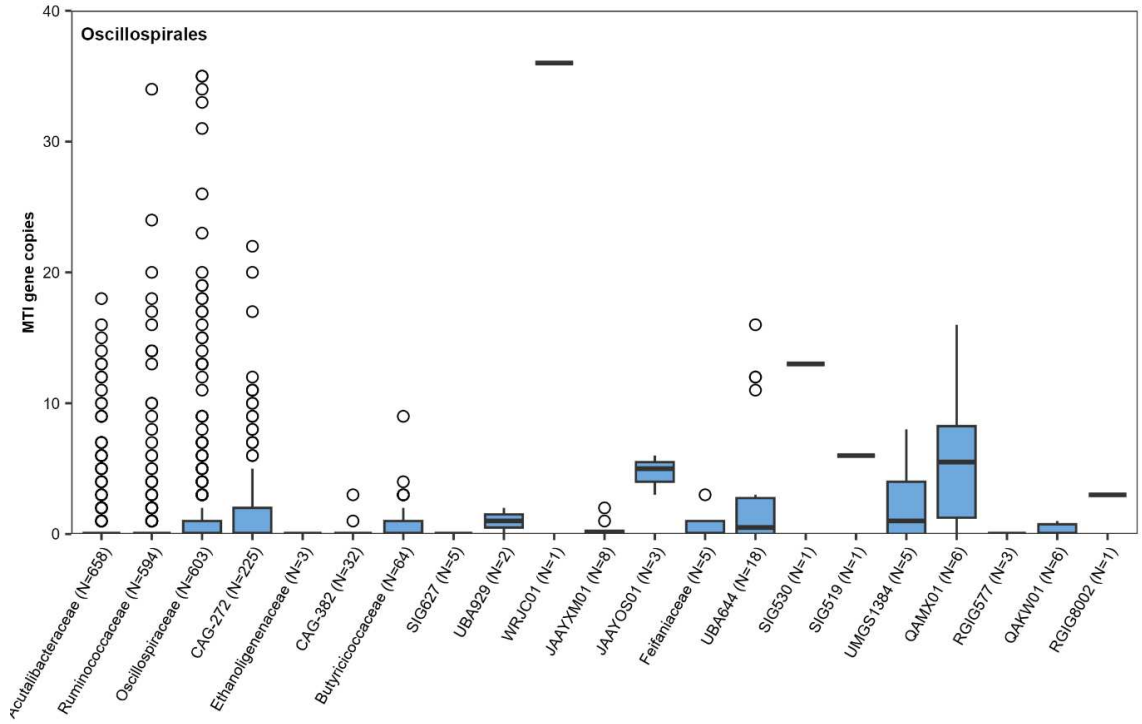

Figure S23 Methyltransferase I (MTI) gene copies in single families of species representatives in the order Oscillospirales. Boxplots are showing the interquartile range with indicated median. Whiskers indicate values below the first and above the third quantile ending in maximum and minimum values. Circles indicate outliers.

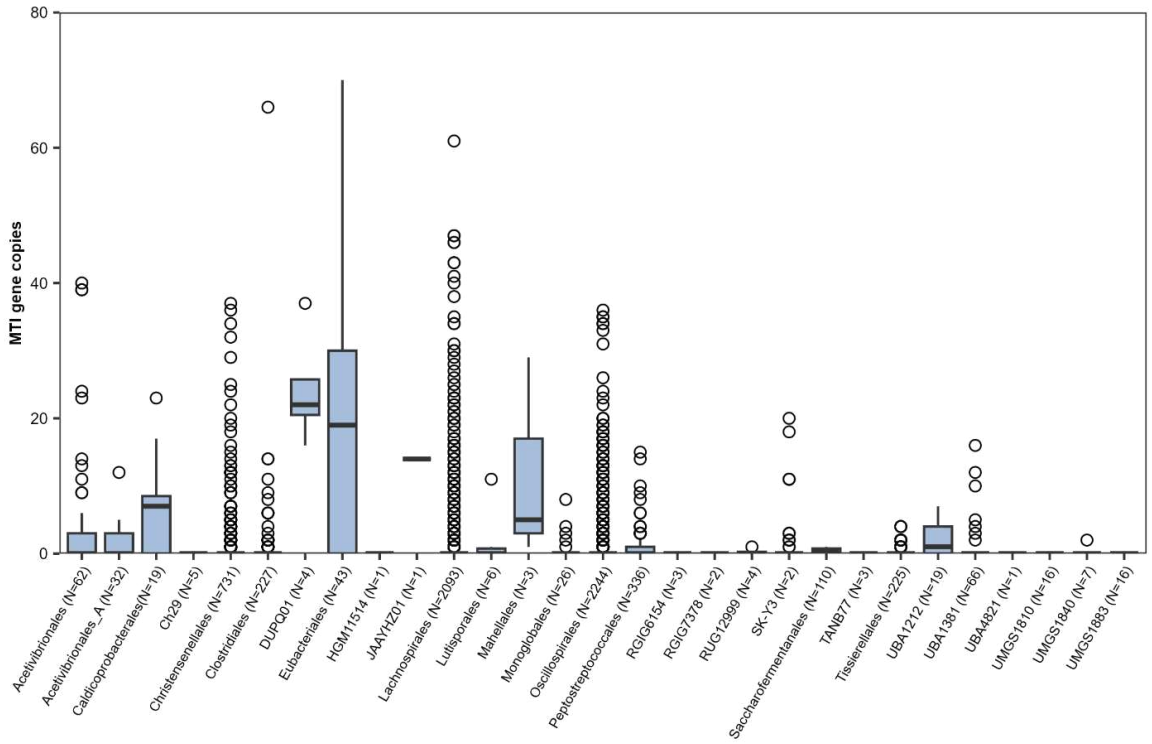

Figure S24 Methyltransferase I (MTI) gene copies in Clostridia species representatives of GTDB. Boxplots show the interquartile range with the indicated median. Whiskers indicate values below the first and above the third quantile ending in maximum and minimum values. Circles indicate outliers.

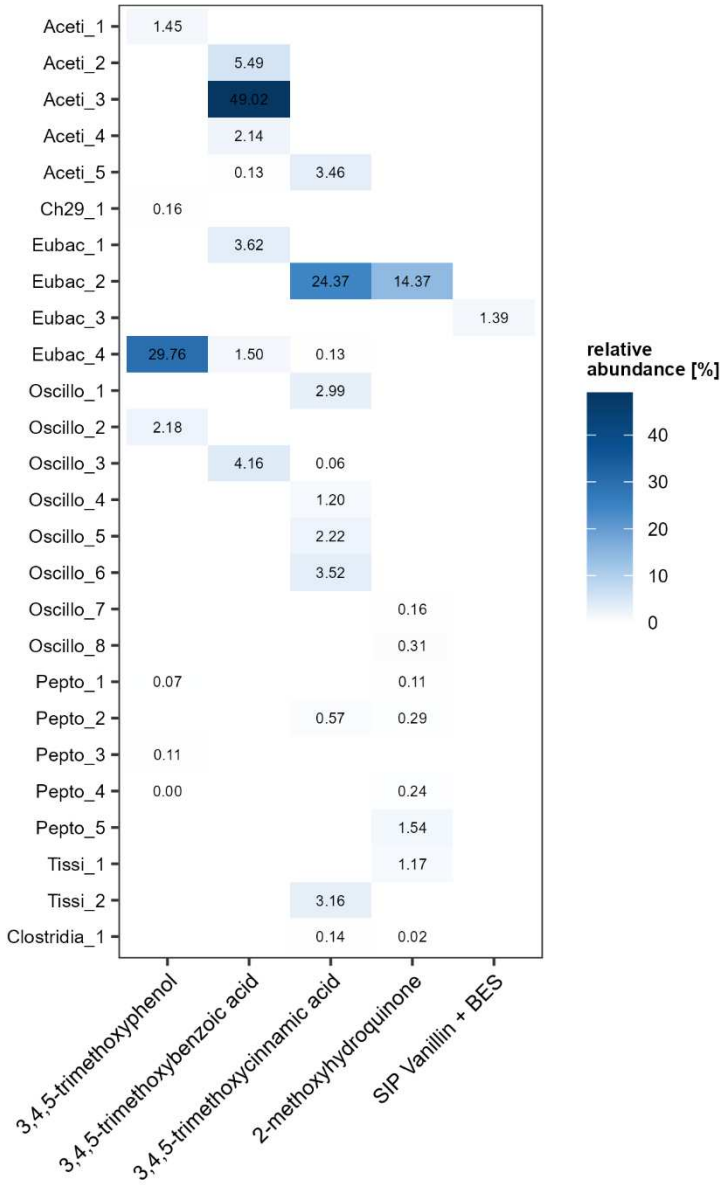

Figure S25 Relative abundance of 26 Clostridia MAGs in metagenomic samples used for metagenomic sequencing.

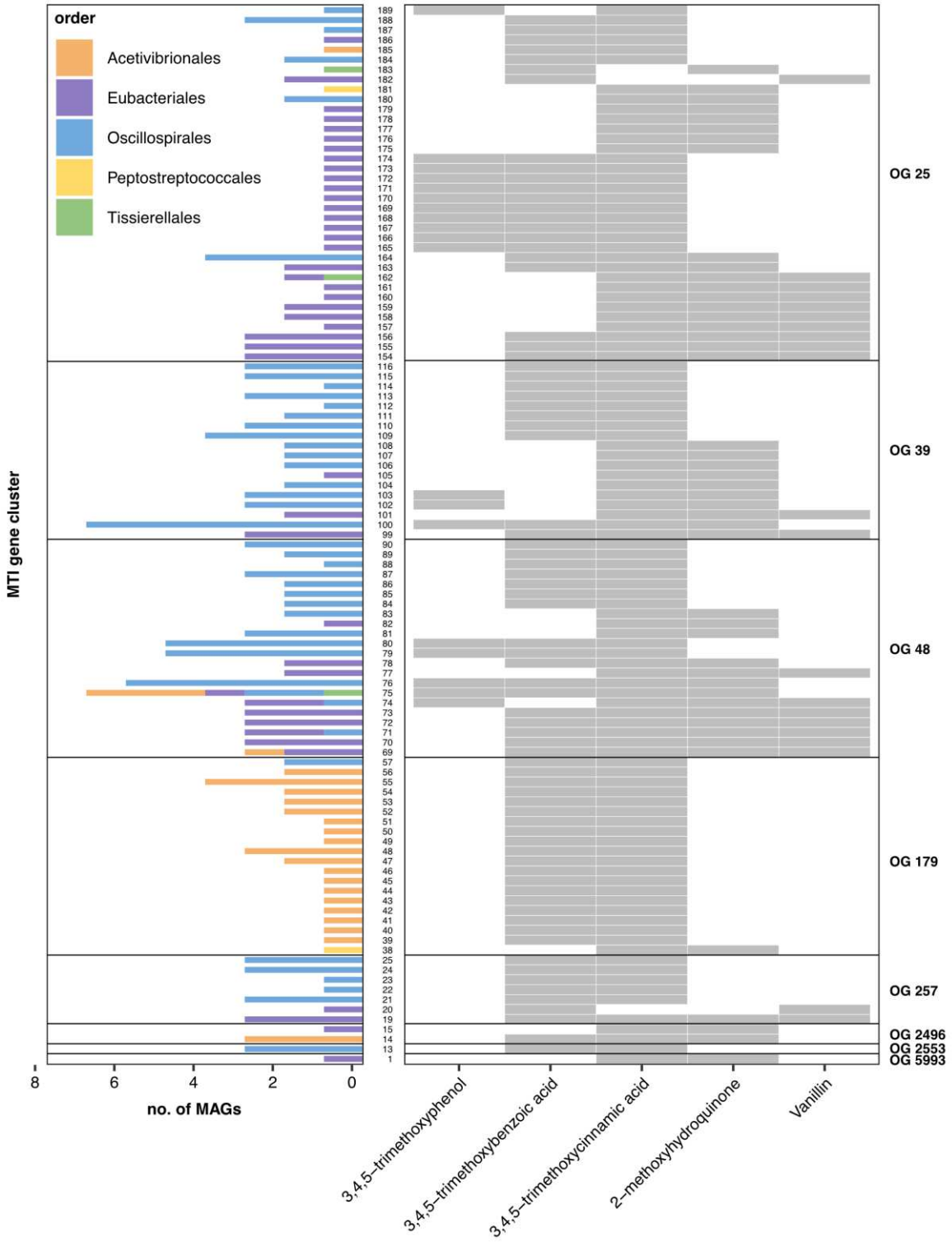

Figure S26 Presence of MTI gene cluster in different treatments. The barplot shows the number of unique MAGs per order, present per MTI gene cluster; horizontal lines in the plot separate orthogroups (OG). Substrate-unspecific gene clusters are indicated by grey color. Numbers between the plots indicate the gene cluster (Table S16).

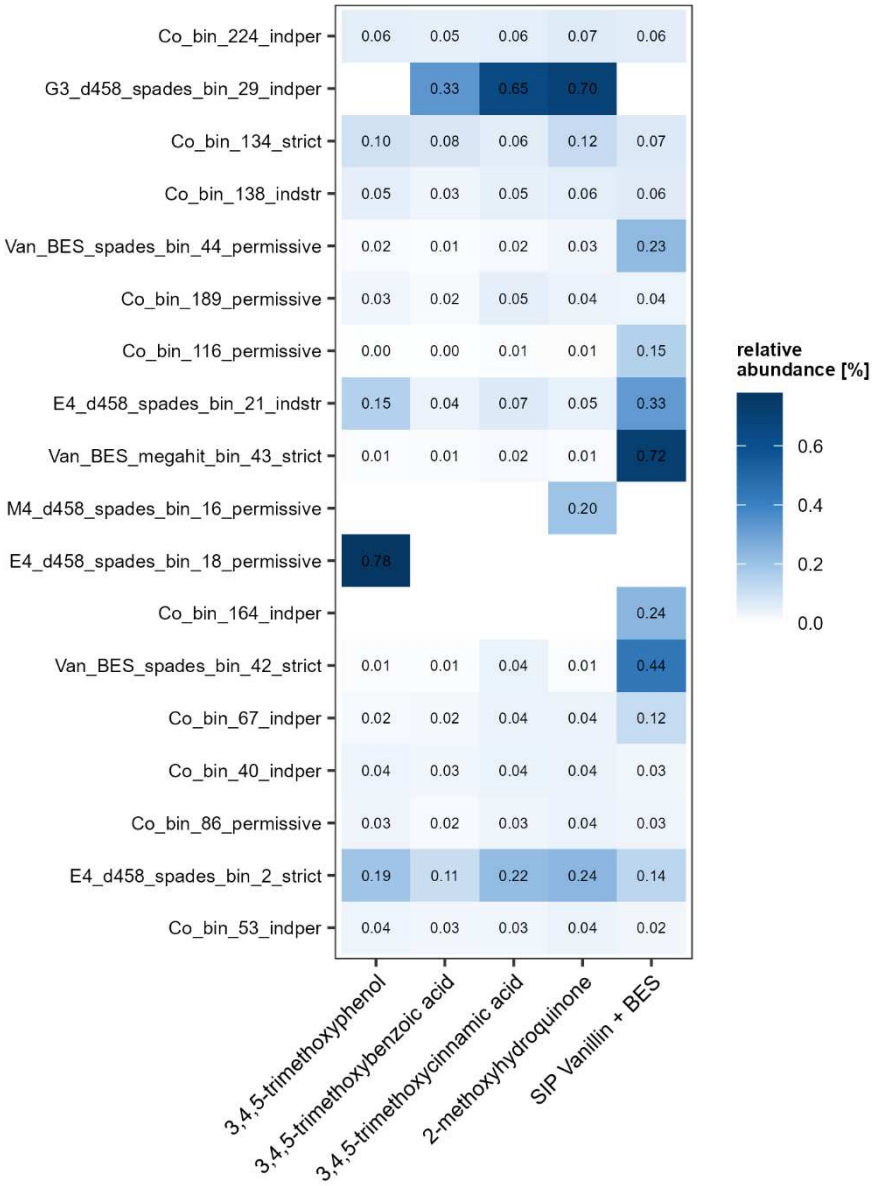

Figure S27 Relative abundance of MAGs involved in aromatic ring degradation in metagenomic samples used for metagenomic sequencing.

4.2.5 Supplementary References

Ben Aissa, F., Postec, A., Erauso, G., Payri, C., Pelletier, B., Hamdi, M., Fardeau, M. L. and Ollivier, B. (2015). Characterization of *Alkaliphilus hydrothermalis* sp. nov., a novel alkaliphilic anaerobic bacterium, isolated from a carbonaceous chimney of the Prony hydrothermal field, New Caledonia. Extremophiles **19**:183-188. doi: 10.1007/s00792-014-0697-y.

Defnoun, S., Labat, M., Ambrosio, M., Garcia, J. L. and Patel, B. K. (2000). *Papillibacter cinnamivorans* gen. nov., sp. nov., a cinnamate-transforming bacterium from a shea cake digester. Int. J. Syst. Evol. Microbiol. **50 Pt 3**:1221-1228. doi: 10.1099/00207713-50-3-1221.

Drent, W. J., Lahpor, G. A., Wiegant, W. M. and Gottschal, J. C. (1991). Fermentation of Inulin by *Clostridium thermosuccinogenes* sp. nov., a thermophilic anaerobic bacterium isolated from various habitats. Appl. Environ. Microbiol. **57**:455-462. doi: 10.1128/aem.57.2.455-462.1991.

Durand, G. A., Cadoret, F., Lagier, J. C., Fournier, P. E. and Raoult, D. (2017). Description of *Gorbachella massiliensis* gen. nov., sp. nov., *Fenollaria timonensis* sp. nov., *Intestinimonas timonensis* sp. nov. and *Collinsella ihuae* sp. nov. isolated from healthy fresh stools with culturomics. New Microb. New Infec. **16**:60-62. doi: 10.1016/j.nmni.2017.01.005.

Emms, D. M. and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**:238. doi: 10.1186/s13059-019-1832-y.

Garnova, E. S., Zhilina, T. N., Tourova, T. P., Kostrikina, N. A. and Zavarzin, G. A. (2004). Anaerobic, alkaliphilic, saccharolytic bacterium *Alkalibacter saccharofermentans* gen. nov., sp. nov. from a soda lake in the Transbaikal region of Russia. Extremophiles **8**:309-316. doi: 10.1007/s00792-004-0390-7.

Grech-Mora, I., Fardeau, M.-L., Patel, B. K. C., Ollivier, B., Rimbault, A., Prensier, G., Garcia, J.-L. and Garnier-Sillam, E. (1996). Isolation and characterization of *Sporobacter termitidis* gen. nov., sp. nov., from the digestive tract of the wood-feeding termite *Nasutitermes lujae*. Int. J. Syst. Evol. Microbiol. **46**:512-518. doi: 10.1099/00207713-46-2-512.

Heuer, H., Krsek, M., Baker, P., Smalla, K. and Wellington, E. M. (1997). Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. Appl. Environ. Microbiol. **63**:3233-3241. doi: 10.1128/aem.63.8.3233-3241.1997.

Khomyakova, M. A., Merkel, A. Y., Petrova, D. A., Bonch-Osmolovskaya, E. A. and Slobodkin, A. I. (2020). *Alkalibaculum sporogenes* sp. nov., isolated from a terrestrial mud volcano and emended description of the genus *Alkalibaculum*. Int. J. Syst. Evol. Microbiol. **70**:4914-4919. doi: 10.1099/ijsem.0.004361.

Konstantinidis, K. T., Rosselló-Móra, R. and Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. ISME J. **11**:2399-2406. doi: 10.1038/ismej.2017.113.

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L. and Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. U. S. A. **82**:6955-6959. doi: 10.1073/pnas.82.20.6955.

Lechtenfeld, M., Heine, J., Sameith, J., Kremp, F. and Müller, V. (2018). Glycine betaine metabolism in the acetogenic bacterium *Acetobacterium woodii*. Environ. Microbiol. **20**:4512-4525. doi: 10.1111/1462-2920.14389.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., et al. (2004). ARB: a software environment for sequence data. Nucleic Acids Res. **32**:1363-1371. doi: 10.1093/nar/gkh293.

Mei, N., Postec, A., Erauso, G., Joseph, M., Pelletier, B., Payri, C., Ollivier, B. and Quéméneur, M. (2016). *Serpentinicella alkaliphila* gen. nov., sp. nov., a novel alkaliphilic anaerobic bacterium isolated from the serpentinite-hosted Prony hydrothermal field, New Caledonia. Int. J. Syst. Evol. Microbiol. **66**:4464-4470. doi: 10.1099/ijsem.0.001375.

Podosokorskaya, O. A., Merkel, A. Y., Heerden, E. v., Cason, E. D., Kopitsyn, D. S., Vasilieva, M., Bonch-Osmolovskaya, E. A. and Kublanov, I. V. (2017). *Sporosalibacterium tautonense* sp. nov., a thermotolerant, halophilic, hydrolytic bacterium isolated from a gold mine, and emended description of the genus *Sporosalibacterium*. Int. J. Syst. Evol. Microbiol. **67**:1457-1461. doi: 10.1099/ijsem.0.001737.

Pruesse, E., Peplies, J. and Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics **28**:1823-1829. doi: 10.1093/bioinformatics/bts252.

Takii, S., Hanada, S., Tamaki, H., Ueno, Y., Sekiguchi, Y., Ibe, A. and Matsuura, K. (2007). *Dethiosulfatibacter aminovoran*s gen. nov., sp. nov., a novel thiosulfate-reducing bacterium isolated from coastal marine sediment via sulfate-reducing enrichment with Casamino acids. Int. J. Syst. Evol. Microbiol. **57**:2320-2326. doi: 10.1099/ijs.0.64882-0.

Walters, W. A., Caporaso, J. G., Lauber, C. L., Berg-Lyons, D., Fierer, N. and Knight, R. (2011). PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics **27**:1159-1161. doi: 10.1093/bioinformatics/btr087.

Chapter V

General discussion and concluding remarks

Marine sediments, especially coastal areas, receive large amounts of organic matter. The sources of this organic matter are either of marine or terrestrial origin. While marine organic matter is thought to be more labile, terrestrial organic matter is highly recalcitrant. Yet, both sources provide complex molecules, such as proteins, polysaccharides, lignin and cellulose. These complex molecules require an initial extracellular breakdown in order to be imported into the cell and become available for degradation. Primary fermenters then utilize monomers in cells for an initial degradation, converting them to fermentation products, including short-chain fatty acids, alcohols, H_2 and CO_2. Secondary fermenters and respiratory organisms thrive on fermentation products. Thus, numerous microorganisms in sediments occupying different ecological niches participate in complex organic carbon degradation, ultimately leading to the production of CO_2 and CH_4. These sediment microbes belong to diverse bacteria and archaea, which have been identified and characterized in the Helgoland mud area (HMA) by previous studies (Oni et al. 2015, Yin et al. 2020, Yin et al. 2022, Yin et al. 2024). Yet, these studies further revealed novel lineages involved in organic carbon degradation and demonstrated the potential diversity of microorganisms yet to be explored.

The primary goal of this dissertation was to identify and characterize yet undescribed microbial lineages involved in the degradation of organic matter in anoxic marine sediments. Throughout this dissertation, I identified yet undescribed taxa involved in marine and terrestrial organic matter degradation in the HMA, using enrichment techniques, stable isotope probing, metagenomic analyses and data mining. I identified active respiratory and fermentative microorganisms involved in the final breakdown of organic matter *in situ* (Chapter II). Various active undescribed taxa of the Desulfobulbia and Desulfuromonadia were involved in sulfur and iron-compound cycling, linking these element cycles with the final breakdown of organic matter. Primary fermenters supplying fermentation products to the *in situ* active microbial community were enriched by amending the same sediments with different carbon sources representing marine organic matter (protein) and terrestrial organic matter (methoxylated aromatic compounds). A novel rare class of Thermoplasmatota was identified as performing protein and amino acid degradation in the HMA and other coastal anoxic environments globally (Chapter III). Moreover, this novel class exhibited high metabolic functionality and was evolutionary distant from other known Thermoplasmatota. Terrestrial organic matter degrading

Clostridia were active in surface sediments and the sulfate reduction zone of the HMA (Chapter IV). Multiple unclassified members of five orders were involved in the methoxydotrophic growth, utilizing methoxylated aromatic compounds (MACs) and represent the first marine members of these orders capable of methoxydotrophy. High abundances of methyltransferases indicated a general C_1-carbon compound metabolism for this class, likely involved in carbon transformations in coastal areas.

The main findings of this dissertation are presented and discussed in the respective chapters. This general discussion combines the primary and supplementary results obtained in the single chapters regarding the expanded knowledge of organic matter remineralization in the Helgoland mud area and the importance of these findings for organic matter degradation in general. Further open questions and future perspectives are addressed in this discussion. Taxonomies mentioned in this discussion follow the Genome Taxonomy Database (GTDB) version r220 (Parks et al. 2018), unless indicated otherwise.

5.1 *In situ* activity of microbial communities in surface sediments of the Helgoland mud area is confined to sulfur and iron-cycling bacteria

DNA-based analyses offer an excellent tool for investigating the abundance of microorganisms in the environment and are used to gain first insights into microbial communities. However, this data does not accurately reflect the actual activity of microorganisms. Microorganisms in marine sediments are found in different metabolic states, including growing and active organisms, besides dormant and recently deceased organisms (Blazewicz et al. 2013, Wang et al. 2023). Ribosomal RNA (rRNA) was considered a suitable proxy for analyzing metabolically active microorganisms. In recent years, more studies rebutted this assumption due to inconsistencies between rRNA content and real-time activity (Blazewicz et al. 2013). Additionally, it was found that different environmental conditions can affect the longevity of rRNA molecules (Wang et al. 2023). To understand the *in situ* activity and identify actively metabolizing taxa in the marine anoxic sediment, methods other than 16S rRNA gene or rRNA sequencing are required. Identifying and characterizing the active microbial communities *in situ* is essential for understanding dominant pathways in marine anoxic sediments and evaluating the status and roles of microbes in element cycling.

To identify active players *in situ*, we combined the highly sensitive RNA-SIP technique with $H_2{}^{18}O$, as active cells incorporate oxygen atoms into their nucleic acids (Adair and Schwartz 2011). We developed a small-scale RNA-SIP incubation of only 7 days to track the *in situ* activity of diverse microbes in the anoxic surface sediment (0-25 cm) of the Helgoland mud

area (Chapter II). High microbial activity was apparent for sulfur and iron compound cycling bacteria of the Desulfobulbia (Desulfocapsaceae, Desulfurivibrionaceae, Desulfobulbaceae) and Desulfuromonadia (Sva1033). Some key microbes were autotrophs, as revealed from control RNA-SIP incubations of the surface sediments using ^{13}C-labeled DIC as the sole carbon source in Chapter IV. We identified members of the Desulfobulbia amongst the most active autotrophic taxa in anoxic marine sediment, suggesting the involvement of sulfur disproportionation. These SIP analyses have led to conclusions different from those of 16S rRNA gene amplicon sequencing analysis alone. For example, the class Desulfobulbia was thought to be involved in the oxidation of low-molecular-weight compounds derived from primary fermentation previously (Oni et al. 2015). Specifically, the genus MSBL7 (JAJRUT01), unclassified Desulfocapsaceae and unclassified Desulfobulbales affiliated with the class Desulfobulbia were identified as autotrophic community members, indicated by the labeling with ^{13}C-DIC (Chapter IV). A comparison of the detected active Desulfobulbia ASVs in Chapters II and IV revealed close identities between ASVs of the MSBL7 (sq12, sq34), Desulfocapsaceae (sq22, sq50, sq13) and unclassified Desulfobulbales (sq31, sq27) obtained in Chapter II and those obtained in Chapter IV (Fig. 1).

The highest proportion of active taxa was found for the MSBL7 cluster (Chapter II). Based on the SILVA taxonomy, MSBL7 was described as genus (Quast et al. 2012). Yet, the relative evolutionary divergence computed by GTDB suggests two different genera for those MSBL7-assigned ASVs (Parks et al. 2018). One ASV with a close identity to the genus JAJRUT01 (sq19) of the MSBL7 cluster could be identified as autotrophic by labeling with ^{13}C-DIC. In contrast, a second genus, BM506, could not be identified among autotrophic active taxa (Chapter IV). These findings suggest that among the MSBL7 cluster, different lifestyles persist. Previous studies described MSBL7 as a sulfur compound reducing bacteria due to its taxonomic affiliation with the phylum Desulfobacterota, yet no cultured representative exists, confirming these findings (Pachiadaki et al. 2014, Wegener et al. 2016, Barnum et al. 2018, Deja-Sikora et al. 2019, Zhong et al. 2022). An earlier study furthermore suggested the possibility of sulfur compound disproportionating activities for a strain closely related to the MSBL7 cluster (Wegener et al. 2016). The finding of different lifestyles could indicate different roles of the MSBL7 cluster members.

A high proportion of active sulfur compound cycling *Desulfocapsa* represented further sulfur compound disproportionating microorganisms. *Desulfocapsa* species grow by disproportionating sulfur compounds, e.g. thiosulfate, sulfite, and elemental sulfur (Finster

2008). By ^{13}C-DIC labeling, we identified two autotrophic Desulfocapsaceae ASVs (sq34, sq37, Chapter IV) of the genus *Desulfocapsa* with high similarities to previously identified Desulfocapsaceae (sq22, sq50, sq13, Chapter II), supporting the growth by sulfur disproportionation. Further active Desulfocapsaceae (sq15, sq19, sq20, sq21, sq26, sq53, Chapter II) were not found to incorporate ^{13}C-DIC (Chaper IV), indicating a heterotrophic lifestyle. These ASVs were closest related to members of the genera *Desulfomarina*, *Desulforhopalus* and *Desulfopila*. All these genera have the metabolic ability to reduce sulfur compounds, coupled with the oxidation of various fermentation products, e.g., acetate, propionate, lactate, pyruvate, ethanol, propanol, butanol and fumarate as electron donors (Isaksen and Teske 1996, Lie et al. 1999, Suzuki et al. 2007, Gittel et al. 2010, Hashimoto et al. 2021, Song et al. 2021). Fumarate was also used as an electron acceptor coupled to the oxidation of lactate by *Desulfopila aestuarii* (Suzuki et al. 2007) and fermented by *Desulfopila inferna* (Gittel et al. 2010). Some of these genera (*Desulforhopalus, Desulfopila*) also showed the ability to ferment pyruvate or taurine (Isaksen and Teske 1996, Lie et al. 1999, Suzuki et al. 2007). Only a few Desulfocapsaceae species could use H_2 as an electron donor, coupled with CO_2 (Hashimoto et al. 2021) or acetate as a carbon source (Isaksen and Teske 1996). Only recently, a study showed that unclassified sulfate-reducing members of the Desulfocapsaceae grow possibly mixotrophically using either CO_2 or the fermentation products lactate, propionate and butyrate (Yin et al. 2024). Here, we could not detect an autotrophic metabolism besides for the genus *Desulfocapsa*, suggesting these identified Desulfocapsaceae ASVs require the presence of fermentation products to thrive.

Active unclassified Desulfobulbales (sq31, sq44, Chapter IV) could be identified as autotrophic. These ASVs, along with those observed in Chapter II (sq31, sq27) and the genome JAKITW01 sp021647905 (GCA_021647905.1) form a monophyletic group in the constructed 16S rRNA gene phylogenetic tree and are likely to form the yet undescribed family BM004. So far, no information is available on the metabolic capabilities of this group. However, as part of the Desulfobacterota phylum, Desulfobulbales are likely to be involved in reducing sulfur compounds (Slobodkin and Slobodkina 2019).

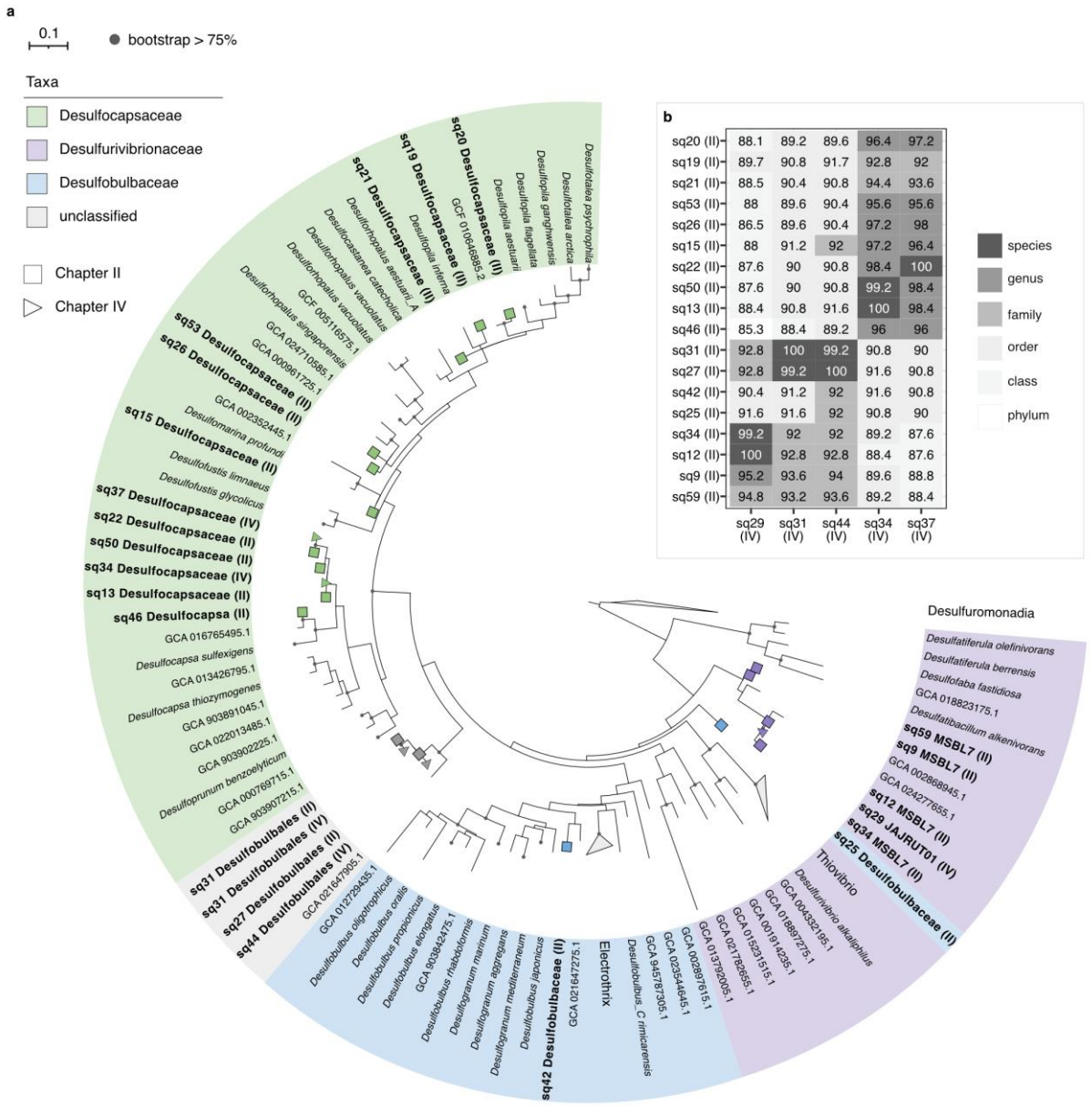

Figure 1 Phylogenetic placement of active ASVs of the Helgoland mud area. **(a)** 16S rRNA gene phylogenetic tree of active taxa of the class Desulfobulbia in the Helgoland mud area surface sediments. The 16S rRNA gene sequences for tree reconstruction were retrieved from available good-quality genomes (completeness >80%, contamination < 5%) present in the GTDB version r214 using barrnap. ASVs of the class Desulfobulbia found in Chapter II and Chapter IV with relative abundance of > 2% were added to the sequence set. The tree was calculated using raxml-ng v1.2.1 (70 starting trees, bootstrap convergence with a bootstrap cutoff of 0.03 was reached after 800 trees). **(b)** Similarity matrix of ASVs derived from Chapter II and Chapter IV computed with blastn of ncbi blast v2.16.0+. For the distinction of taxon levels, the taxon level thresholds of Konstantinidis et al. (2017) were applied (species: > 98.6%, genus: 95-98.6%, family: 92-95%, order: 89-92%, class: 86-89%, phylum: 83-86%). Code used to create this figure is deposited at https://github.com/mmaeke/Thesis_Discussion.

The results obtained in this dissertation show that members of the class Desulfobulbia are actively involved in sulfur cycling in the HMA surface sediments and act on remaining fermentation products and available CO_2 (Fig. 2). Sulfate-reducing bacteria play the primary role in the degradation of organic matter in anoxic marine sediments by facilitating the final

oxidation of organic matter to CO_2 (Jørgensen et al. 2019). Altogether, they enable about 50% of the carbon remineralization in anoxic marine sediments, making them essential players in organic matter rich sediments and organic matter turnover (Jørgensen 1982). The most active taxa identified in the HMA have all been attributed to sulfur-cycling organisms, indicating the process of sulfur cycling to be a dominant process in the HMA *in situ*. Dominantly autotrophic and less abundant heterotrophic microorganisms were identified. Some of the autotrophic taxa could be identified as sulfur compound disproportionating organisms. During sulfur compound disproportionation, elemental sulfur, thiosulfate, and sulfite are electron donors and acceptors (Bak and Cypionka 1987, Thamdrup et al. 1993). Yet, the process of sulfur compound disproportionation and the biochemical pathways used are hardly understood (Slobodkin and Slobodkina 2019). So far, no functional marker genes involved in the process have been identified (Slobodkin and Slobodkina 2019). As most sulfur compound disproportionating microorganisms can grow by other processes, the distinction between sulfur compound disproportionating and other sulfur compound cycling organisms remains difficult (Slobodkin and Slobodkina 2019). Therefore, the importance and abundance of this process in anoxic marine sediments remain a critical ongoing research topic.

Heterotrophic sulfur compound cycling organisms were likely involved in reducing sulfate, thiosulfate, and sulfite coupled to the oxidation of fermentation products. Hydrogen sulfide produced by these sulfur compound reducers is either reoxidized to intermediate sulfur compounds, such as sulfite, elemental sulfur, polysulfides, or thiosulfate, or buried in sediments as complexes with iron or organic matter (Berner 1982, Jørgensen 1982, Brüchert 1998). Due to the rapid turnover of sulfur compounds, analyzing these concurrently ongoing biogeochemical processes remains challenging.

Apart from active sulfur-cycling microorganisms of the Desulfobacterota, the labeling by $H_2{}^{18}O$ revealed further active groups. We could not detect these groups in ^{13}C-labeled DIC enrichments, so these organisms must also be involved in heterotrophic activity. The most active group was identified as the family Sva1033, which has been described as an acetate-utilizing dissimilatory iron reducer (Aromokeye et al. 2021b, Wunder et al. 2021). The group of Sva1033 is regularly found in cold marine sediments of the Arctic and Antarctic environments (Ravenschlag et al. 1999, Aromokeye et al. 2021b, Wunder et al. 2021, Baloza et al. 2023). Despite being abundant in these environments, the knowledge of this group remains scarce and limited to RNA-SIP incubations conducted in these environments

(Aromokeye et al. 2021b, Wunder et al. 2021). To date, no genomic analyses of this group exist.

The ^{18}O-labeling also revealed lower activity for Bacteroidia (*Marinifilum*) and Fusobacteriia (*Propionigenium*). Bacteroidia of the genus *Marinifilum* have the metabolic capability to ferment various carbohydrates, polysaccharides and amino acids (Na et al. 2009, Ruvira et al. 2013, Fu et al. 2018, Dai et al. 2022). In a previous study, the phylum Bacteroidota was suggested to be involved in the degradation of algal-derived organic matter, such as polysaccharides (Oni et al. 2015). The results obtained in this study confirmed the active participation of this group in organic matter degradation. We further identified the active *Propionigenium.* This genus is also involved in the fermentation of succinate, fumarate, pyruvate (Schink and Pfennig 1982), besides carbohydrates and amino acids (Janssen and Liesack 1995). Less abundant Clostridia of the genus *Fusibacter* were potentially involved in respiratory pathways reducing sulfate, thiosulfate or elemental sulfur coupled with the oxidation of carbohydrates (Ravot et al. 1999, Smii et al. 2015, Qiu et al. 2021). For some members the reduction of Fe(III) compounds was tested (Qiu et al. 2021). Yet, also, fermentation of carbohydrates was observed (Ravot et al. 1999, Smii et al. 2015). Due to their ability to ferment carbohydrates, all three members might be involved in the degradation of algal-derived polysaccharides.

Gammaproteobacteria described to be highly abundant in surface sediments of the HMA (Oni et al. 2015) were not active during seven days of incubation. In fact, Gammaproteobacteria were primarily abundant in the ultra-light fraction of our experiment (14% relative abundance), indicating no or shallow activity (Chapter II). In our long-term incubation using ^{13}C-DIC as sole carbon source, Gammaproteobacteria were labeled beside the Desulfobulbia mentioned above, hinting towards an autotrophic activity of these members in the sediments of the HMA (Chapter IV). However, due to long incubation times of 100 days, cross-feeding cannot be excluded. Gammaproteobacteria have been previously suggested to be involved in the degradation of algal-derived organic matter in the surface sediments of the HMA (Oni et al. 2015). A more recent study on the sediments of the HMA identified metal-reducing Gammaproteobacteria, e.g., Sedimenticolaceae, *Ferrimonas,* and *Amphritea*, as respiratory microorganisms using fermentation products, such as acetate, lactate, propionate, butyrate, and ethanol (Yin et al. 2024). Yet, no autotrophic activity was observed (Yin et al. 2024). Due to their potential autotrophic activity in this study, the Gammaproteobacteria are likely different from those metal-reducers or organisms involved in the degradation of algal-derived organic

matter. In some studies, Gammaproteobacteria were identified as sulfur-oxidizers, accounting for up to 70% of the dark carbon fixation in coastal surface sediments (Lenk et al. 2011, Dyksma et al. 2016). Thus, Gammaproteobacteria, identified here as autotrophic, might be involved in sulfur cycling.

Acidobacteria, Verrucomicrobia and Cyanobacteria, which were detected in low relative abundances in surface sediments by previous studies, were not active *in situ* (Oni et al. 2015). Further, no activity of primary fermenters and autotrophic archaea was detected in short-term incubations, likely due to the depletion of complex organic carbon in these sediments after storage since 2019 or slower metabolic activity. Instead, archaea were likely to be more active in deeper sediments.

Thus, to conclude, the dominantly active sulfur compound and iron-cycling microorganisms detected in organic carbon-depleted sediments of the HMA are autotrophic and heterotrophic organisms acting on DIC and fermentation products, and with this, the first hypothesis raised in this dissertation is true. However, we also observed a lower activity of *Marinifilum, Propionigenium* and *Fusibacter*, potentially involved in the fermentation of higher molecular weight compounds, such as carbohydrates. Furthermore, we observed that both autotrophic and heterotrophic taxa became active in carbon-depleted sediments. While *Desulfocapsa* sp., the genus JAJRUT01 of the MSBL7 cluster and the family BM004 of the order Desulfobulbales were likely autotrophic, other heterotrophic members, such as unclassified Desulfocapsaceae, the genus BM506 of the cluster MSBL7 and Sva1033 were equally active. Therefore, these results suggest that autotrophic taxa are not predominantly active in the carbon-depleted sediments used and the second hypothesis raised was false. Nevertheless, results on microbial activity are derived from two different studies and SIP incubations conducted throughout this dissertation. Additional genomic data is required to gain more insights into the metabolic potential of the found microorganisms.

5.2 Multiple microbial groups contribute to protein fermentation in subsurface sediments of the HMA

High concentrations of marine organic matter characterize the western HMA surface sediments (Oni et al. 2015, Müller et al. 2024). Such marine organic matter derives from phytoplankton debris, detritus and the remains of other autotrophic and heterotrophic microorganisms inhabiting the overlying water column (Lomstein et al. 2012, Braun et al. 2017). The marine organic matter contains simple molecules such as amino acids, sugars, vitamins, nucleic acids

or fatty acids and more complex molecules such as proteins and polysaccharides (Neidhardt et al. 1990, Burdige 2007, Lengeler et al. 1999, Repeta 2015). Especially proteins and amino acids constitute about 50% of the exported marine organic matter and provide a large pool of carbon for primary fermenting organisms (Burdige 2007).

Previous studies in the HMA identified different microbial groups as the predominant microbes involved in protein breakdown, including bacteria and archaea (Yin et al. 2020, Yin et al. 2022). For example, bacterial groups such as Clostridia (*Fusibacter* and unclassified Clostridia), Enterobacterales (*Psychromonas, Photobacterium* and *Vibrio*) and Fusobacteriia (*Propionigenium*) have been shown to participate in crude protein degradation in the HMA and other environments (Pelikan et al. 2021, Zhu et al. 2024). As mentioned above, *Fusibacter* and *Propionigenium* were active *in situ* in the surface sediments of the HMA. As the HMA surface sediments are rich in marine derived organic matter, these organisms are likely involved in the ongoing crude protein degradation in these sediments *in situ*. However, these bacteria are not detected in the same sample when pure protein is used as the substrate, whereas multiple archaeal groups are identified instead. Thus, it appears that archaea are the key microbial group responsible for protein degradation in HMA sediments (Fig. 2).

In Chapter III, we identified the known protein-degrading *Ca*. Prometheoarchaeum syntrophicum strain MK-D1 in lower sediment zones (40-75 cm) as highly abundant, reaching more than 85% relative abundance in the archaeal community of the enrichments. Lokiarchaeia of the Loki-2c subgroup were previously identified as significant protein-degrading archaea in the sulfate reduction zone of the HMA (16-41 cm) (Yin et al. 2022). In contrast, the subgroups Loki-2a and Loki-2b showed much lower activity (Yin et al. 2020). Thus, while we could not directly show the activity of the strain MK-D1, such high relative abundance indicates the prevalence of this group in the degradation of protein in our enrichment. However, it needs to be noted that the conducted protein-enrichments were amended with antibiotics and the bacterial community therefore suppressed. Overall, the finding of Lokiarchaeia being involved in protein breakdown in these marine organic carbon rich sediments is not surprising. Among protein-degrading organisms, Asgardarchaeota, including Lokiarchaeia were detected as an abundant group (Lazar et al. 2017, Imachi et al. 2020, Yin et al. 2022). Analyzed genomes of Lokiarchaeia, Thorarchaeia, Heimdallarchaeia, Helarchaeia and Gerdarchaeia carried genes for extracellular peptidases required for primary protein hydrolysis and subsequent genes for degradation of amino acids (Seitz et al. 2016, Liu et al. 2018, Fraser et al. 2019, Cai et al. 2020, Imachi et al. 2020, Yin et al. 2022).

Chapter III further revealed an undescribed class within the phylum of Thermoplasmatota (EX4484-6) to be involved in the degradation of proteins in enrichments. The EX4484-6 became highly abundant, reaching more than 90% relative abundance in one of the replicates. Though we could not evaluate why the EX4484-6 became much more abundant than the Lokiarchaeia in only one of the biological replicates, the high abundance of this group in the enrichments gave rise to the hypothesis that the novel group might be involved in the degradation of proteins. Metabolic characterization of this novel group using metagenomic analyses further supported this finding.

The active protein-degrading members of the Thermoplasmatota phylum in the HMA were so far confined to members of the SG8-5 order, while also members of the *Ca*. Thermoprofundales, formerly known as MBG-D, were identified as present in the HMA sediments (Yin et al. 2022). Especially *Ca*. Thermoprofundales are described as one of the predominant protein-degrading microbes in marine sediments (Lloyd et al. 2013, Lazar et al. 2017, Zhou et al. 2019). Thermoplasmatota, in general, are known for their amino acid and protein degrading capabilities, for example, protein and amino acid degrading Marine Group II archaea dominate the water column (Orsi et al. 2016). Further studies revealed the protein and amino acid degradation potentials by members of the order *Ca*. Sysuiplasmatales, *Ca*. Angelarchaeales and *Ca*. Lutacidiplasmatales, inhabiting diverse marine and terrestrial environments (Yuan et al. 2021, Diamond et al. 2022, Sheridan et al. 2022).
Despite these recent findings, studies regularly detect unclassified members of the Thermoplasmatota with the ability to utilize proteins (Yin et al. 2022). However, the abundance of these organisms is much lower than that of the predominant groups of Lokiarchaeia or *Ca*. Thermoprofundales. Like that, the group EX4484-6 was only of low relative abundance in global marine coastal sediments and was detected as a member of the rare biosphere rather than a dominant player in the cycling of proteins. The importance this group might play in the anoxic marine sediments is yet to be explored. Nevertheless, these findings prove that regularly overlooked taxa in marine sediments might play essential roles in carbon cycling.

Bathyarchaeia of subgroup Bathy-15 were identified to be involved in protein degradation in the HMA once antibiotics suppressed the bacterial community (Yin et al. 2022). A recent study showed the presence of extracellular peptidases in genomes of all but one order of the Bathyarchaeia (Hou et al. 2023). The highest numbers of peptidases and extracellular peptidases were found in the order Wuzhiqiibiales, formerly known as subgroup 15 (Hou et al. 2023). Thus, the previous finding of the subgroup Bathy-15 to be involved in protein

degradation in the HMA is following the more recent observations. In the study on protein degradation in this dissertation (Chapter III), Bathyarchaeia could not be enriched. However, high abundances of ~50% were observed in unamended control samples. Overall, Crenarchaeota, including Bathyarchaeia dominate the archaeal community in marine anoxic sediments (Hoshino et al. 2020). Bathyarchaeia are a globally distributed group accounting for up to 92% of total archaea in sediments (Kubo et al. 2012, Zhou et al. 2018). Within eight orders, multiple metabolisms were detected by analysis of the metabolic capabilities of single metagenome assembled genomes (MAGs), including the degradation of detrital proteins, polymeric carbohydrates, fatty acids, aromatic compounds, methane, short-chain alkanes, and methylated compounds (Lloyd et al. 2013, Meng et al. 2014, Evans et al. 2015, Lazar et al. 2015, He et al. 2016, Lazar et al. 2016, Zhou et al. 2018, Yin et al. 2022). Thus, while we did not detect protein-degrading Bathyarchaeia, their role in the anoxic marine sediment could involve the fermentation of many different substrates available in the *in situ* sediment.

5.3 Many unclassified bacterial members participate in the degradation of methoxylated aromatic compounds (MACs) in subsurface sediments of the HMA

High levels of terrestrial organic matter characterize deeper sediments of the HMA, especially in the western HMA (Oni et al. 2015, Müller et al. 2024). Such terrestrial deposits may derive from the erosion processes of the Helgoland island during storms and from the input of the rivers Elbe and Weser (Müller et al. 2024). Overall, terrestrial organic matter contains polysaccharides, polyaromatic, aromatic, and unsaturated structures derived from living plant biomass, plant litter, and soil organic matter, such as humus (Burdige 2007). Terrestrial organic matter is much more recalcitrant than marine organic matter and, thus, more likely to remain in the subseafloor sediments.

Methoxydotrophy

The terrestrial organic matter degradation in the HMA has so far been restricted to one *in situ* study and the degradation of the model compound benzoate (Oni et al. 2015, Aromokeye et al. 2021a). In our study on terrestrial organic matter degradation in the HMA we used methoxylated aromatic compounds (MACs), a significant lignin constituent (Chapter IV). The study revealed Clostridia as the most abundant and active organisms in the incubations of surface sediments down to a sediment depth of 70 cm (Chapter IV). Interestingly, previous studies only identified Clostridia to be of low abundance, even in deeper sediment layers with higher concentrations of terrestrial organic matter (Oni et al. 2015). In the surface sediments, we detected members of the class Clostridia with low activity; the *in situ* abundant *Fusibacter*

was most likely involved in the fermentation of crude protein or low molecular weight carbohydrates rather than the degradation of recalcitrant organic matter. However, upon adding MACs, mostly unclassified Clostridia of the orders Acetivibrionales, Eubacteriales, Oscillospirales, and Tissierellales were identified as the sole community members involved in methoxydotrophy. The dominance of Clostridia in the enrichments suggests that Clostridia might be key players in methoxydotrophic carbon transformations in the anoxic coastal marine sediments of the HMA and drive the preliminary transformation of lignin-derived monoaromatic compounds (Fig. 2).

So far, more than 40 cultured species have been identified to use MACs anaerobically, most of which belong to the class of Clostridia (Khomyakova & Slobodkin 2023). Similar to our findings, most members of Clostridia perform methoxydotrophic acetogenesis, using the methoxy groups as electron donors and CO_2 as electron acceptors (Khomyakova and Slobodkin 2023). The annotation of metabolic pathways required for methoxy group utilization was limited to the pathway of methoxydotrophic acetogenesis via a reductive acetyl-CoA pathway in our study. Yet, some members of the identified Oscillospirales will likely use a different methoxy group utilization strategy. Further, more in-depth analyses of the metabolic potential are required.

In Chapter IV, we could detect different species in up to 9 different MAC enrichments and thus showing a high substrate spectrum for different MACs, which is in accordance with previous findings on MAC degradation by Clostridia. The first cultivated acetogenic methoxydotroph was *Acetobacterium woodii* of the class Clostridia, which utilizes more than 10 different MACs (Bache and Pfennig 1981). The genome contained 23 different methyltransferase I (MTI) genes, suggesting substrate-specificity of these towards single MACs or other methylated compounds, which are yet to be identified (Lechtenfeld et al. 2018, Khomyakova and Slobodkin 2023). The Clostridia species *Moorella thermoacetica* was found to metabolize more than 20 different MACs and contained multiple MTI homologs (Pierce et al. 2008). Other members of the Clostridia were only tested for fewer MACs but showed equal metabolic characteristics (Tanaka and Pfennig 1988, Greening and Leedle 1989, Lux and Drake 1992, Parekh et al. 1992, Kaufmann et al. 1998, Paarup et al. 2006, Allen et al. 2010, Khomyakova et al. 2020, Khomyakova et al. 2021).

MAGs analyzed in Chapter IV contained high numbers of MTI genes. Due to the lack of annotation for these identified MTI genes, we could not identify specific methoxytransferases, only the general methyltransferases. Thus, further research is required to answer these open

questions in Chapter IV. The protein structures of MTI genes need to be compared to those of the MTI genes involved in methoxy group utilization to predict methyltransferases involved in only the methoxy group utilization. As neither the MTI nor corrinoid proteins and methyltransferase II (MTII) genes were annotated, further comparison of hypothetical proteins to known corrinoid proteins and MTII could be used to identify complete *O*-demethylase gene clusters and from this evaluate the gene synteny.

While sample-specific and genome-specific MTI genes could be identified in Chapter IV, none of these could be directly linked to methoxy group utilization of the identified substrate. Since only 5 of the tested MAC treatments were used for metagenomic sequencing, sample-specific methyltransferases might not be specific for the substrate analyzed. Using the protein structure of MTI genes involved in demethoxylation, substrates could be fitted into protein structures to prove substrate-specificity. Further, an analysis of orthogroups of MTI genes revealed the potential evolution of methyltransferases towards different substrate spectra. The evolutionary history of MTI genes could be further investigated by analyzing the codon usage of MTI genes within single orthogroups. A possible lateral gene transfer could be identified by comparing the codon usage among genes of the same orthogroup and the codon usage of the methyltransferase-derived genome. These analyses can investigate the functional divergence of detected methyltransferases within orthogroups and could indicate genes with similar substrate-specificity. To ultimately link MTI genes with methoxy group specificity, metaproteomic analyses would be required to identify expressed genes involved in the degradation of MACs in single treatments.

Despite this vast knowledge of MAC degrading Clostridia, only one member of the Clostridia, *Acetobacterium carbinolicum*, has been isolated from the marine environments (Paarup et al. 2006). Other methoxydotrophic Clostridia have so far only been associated with terrestrial and host-associated environments (Khomyakova and Slobodkin 2023). While this demonstrates the widespread trait of Clostridia for terrestrial organic matter degradation, the finding of diverse groups involved in the methoxydotrophy in the anoxic marine coastal sediment is novel. Overall, Clostridia are common in coastal sediments, reaching ~10% relative abundance (Zinger et al. 2011). Thus, Clostridia likely play an essential role in C_1-carbon compound transformations in environments other than the HMA.

Apart from Clostridia, only some members of the phylum Pseudomonadota and the archaea are known to degrade MACs anaerobically (Taylor 1983, DeWeerd et al. 1990, Kurth et al. 2021,

Welte et al. 2021, Yu et al. 2023a). Pseudomonadota couple the oxidation of methoxylated aromatic compounds to nitrate or sulfur compound reduction (Taylor 1983, DeWeerd et al. 1990). *Archaeoglobus fulgidus*, *Methermicoccus shengliensis* and Bathyarchaeia like *Ca.* Baizosediminiarchaeum ligniniphilus and *Bathyarchaeum tardum* were found to use methoxy groups, coupling the methyl group transfer to the electron acceptor tetrahydromethanopterin (Kurth et al. 2020, Welte et al. 2021, Khomyakova et al. 2023, Yu et al. 2023a). The metabolic capability was even detected in five of eight Bathyarchaeia orders, thus indicating a rather widespread utilization of methoxy groups among this class (Yu et al. 2023a). While the sulfate-reducing *Archaeoglobus fulgidus* and the acetogenic Bathyarchaeia convert the methoxy group to CO_2 and acetate via a reductive acetyl-CoA pathway, the methanogen *Methermicoccus shengliensis* performs methoxydotrophic methanogenesis (Mayumi et al. 2016, Kurth et al. 2020, Welte et al. 2021, Khomyakova et al. 2023, Yu et al. 2023a). Besides, further archaea belonging to Lokiarchaeia, Verstratearchaeota, Korarchaeota, Helarchaeota, and Nezhaarchaeota were suggested to have the genetic potential for the growth with MACs (Welte et al. 2021). In the HMA, the Lokiarchaeia subgroup Loki-3 was detected to be active in lignin-amended incubations (Yin et al. 2020). Despite these initial findings, no proof of the involvement of Lokiarchaeia in lignin degradation has been found. Our results of the archaeal community in MAC amended enrichments indicate that no archaeal community member was involved in the MAC breakdown via methoxydotrophy. Furthermore, none identified taxa had the metabolic potential for aromatic ring cleavage. Hence, the capability of using terrestrial-derived monoaromatics is likely restricted to single subgroups within taxa, like Bathyarchaeia or Lokiarchaeia and is not a trait shared widely.

Aromatic ring cleavage

As of earlier studies (Oni et al. 2015, Aromokeye et al. 2021a), we identified members of the Desulfobacterota (*Malonomonas*, Desulfobacteraceae, Desulfocapsaceae, Desulfatiglandales and unclassified Desulfobacterota) and Chloroflexota (Anaerolinea and Dehalococcoidia) to potentially be involved in the degradation of aromatic compounds via the benzoyl-CoA intermediate in the HMA (Chapter IV). The oxidation of aromatic compounds via the benzoyl-CoA pathway is a common trait among sulfate-reducing bacteria (Davidova et al. 2007, Musat and Widdel 2008, Ahn et al. 2009, Musat et al. 2009, Junghare and Schink 2015, Zhuang et al. 2019).

Members of the phylum Chloroflexota, such as Dehalococcoidia are a widespread group in marine sediments and have been reported to degrade benzoate in various environments

(Wasmund et al. 2014, Wasmund et al. 2016, Yu et al. 2023b). Anaerolinea that have been lacking multiple genes of the lower pathway in the study by Aromokeye et al. (2021a) did at least contain the complete upper benzoyl-CoA pathway converting benzoyl-CoA into alipathic C_7-dicarboxyl-CoA compounds as described in Carmona et al. (2009) in our study. Anaerolinea have been described as a broadly distributed group in diverse environments, reaching high abundances (Blazejak and Schippers 2010). Some members of the Anaerolinea have been described as fermentative bacteria utilizing a broad spectrum of carbohydrates but were incapable of using benzoate (Sekiguchi et al. 2003, Yamada et al. 2006). Anaerolinea detected in Chapter IV are of yet undescribed orders E26-bin7 and UBA7937. Thus, these groups might have the potential for benzoyl-CoA degradation, although there is no knowledge on benzoate degrading Anaerolinea.

Despite detecting some members involved in aromatic ring cleavage, previous research investigating the benzoate degradation in sediments of the HMA by incubations and metagenomics identified a much higher diversity of organisms involved in the degradation of aromatic compounds. Members of the phyla Bacillota_B (Desulfotomaculia, Syntrophomonadia), Bacillota_D (Dethiobacteria) and Bacteroidota (Ignavibacteria) were identified to take part in the aromatic ring cleavage (Aromokeye et al. 2021a). Of these, only Bacillota_B were detected among abundant bacteria in the enrichments of Chapter IV yet could not be identified as aromatic compound degraders by metagenomic analyses. In our study, we performed metagenomic sequencing on some samples only; hence, we likely missed the Desulfotomaculia present in high relative abundances only in the 2-methoxybenzoic acid treatment. Desulfotomaculia in those enrichments were assigned to the family Desulfallaceae (*Sporotomaculum*) and, therefore, of the same genus as previously detected Desulfotomaculia in the HMA (Aromokeye et al. 2021a). The genus *Sporotomaculum* could be identified as a benzoate degrader previously (Qiu et al. 2003) and the ability to utilize benzoate has also been detected in other species of the Desulfotomaculia (Plugge et al. 2002). As we did not observe methoxydotrophic activity in the 2-methoxybenzoic enrichment, the most abundant Desulfotomaculia likely performed aromatic compound degradation on this methoxylated aromatic compound.

We further reconstructed eight MAGs of the Syntrophomonadaceae, none of which had the metabolic potential to degrade aromatic compounds. Syntrophomonadaceae are known for their ability to degrade fatty acids of various lengths syntrophically (McInerney et al. 1981, Stieb and Schink 1985, Roy et al. 1986, Zhang et al. 2004, 2005). In enrichments with MACs,

this group might, be involved in the attack of acyl side chains of aromatic compounds via β-oxidation (Gibson and Harwood 2002) or the degradation of short-chain fatty acids derived from aromatic ring cleavage rather than aromatic compound degradation.

Overall, taxa with the ability to degrade benzoyl-CoA were only of low abundance in the enrichments of Chapter IV. Earlier studies in the HMA hinted that Clostridia, the most abundant class in our enrichments, potentially participate in aromatic compound degradation (Aromokeye et al. 2021a). Yet, Clostridia detected in Chapter IV could not degrade aromatic compounds via the intermediate benzoyl-CoA based on metagenomic analyses. Moreover, we did not detect genes for the pathway of phloroglucinol degradation, a pathway quite common in multiple members of Clostridia, including *Sporobacter termitidis* and *Sporobacterium olearium* (Kreft and Schink 1993, Grech-Mora et al. 1996, Mechichi et al. 1999, Zhou et al. 2023b). Also, genes for the reductive degradation of resorcinol were observed in fermenting bacteria of Clostridia but were lacking in any of the screened MAGs (Tschech and Schink 1985, Schnell et al. 1989). Despite the lack of genes for these pathways in Clostridia MAGs, we cannot exclude the ability of Clostridia to be involved in these pathways. Genes for these pathways have only recently been characterized (Zhou et al. 2023b) or are yet to be described (Schink et al. 2000). While there is knowledge of the functionality of these genes, exact annotations are still lacking from standard databases, such as KEGG (Kanehisa et al. 2016) and NCBI NR (Sayers et al. 2022). Further protein family models of detected involved genes could be used for a more in-depth analysis of these two pathways to investigate the possible participation of Clostridia in aromatic compound degradation.

5.4 Fermentation of other complex organic matter

Though not studied in the Chapters of this dissertation, more complex organic matter, such as polysaccharides, is available in marine environments. As we identified members of the Bacteroidetes and Fusobacteriia active *in situ,* possibly involved in the fermentation of carbohydrates, the following part briefly covers the sources of polysaccharides in the marine environment and microorganisms involved in the fermentation of these, leading to the availability of carbohydrates.

Polysaccharides present in the marine environment include starch, laminarin, fucoidan, alginate, and cellulose, commonly found in the cell walls or as storage products of green, brown, and red algae (Davis et al. 2003); chitin from crustacean exoskeletons, zooplankton and diatom surfaces (Gooday 1990, Durkin et al. 2009) and cellulose, derived from terrestrial

environments (Lynd et al. 2002). Microorganisms affiliated with polysaccharide degradation are Pseudomonadota, predominantly of the class Gammaproteobacteria, Clostridia, Bacteroidota, Actinomycetota and Verrucomicrobiota (Lynd et al. 2002, Fernández-Gómez et al. 2013, Mann et al. 2013, Jain and Krishnan 2017, Sichert et al. 2020, McKee et al. 2021, Sajeela et al. 2023, Zhang et al. 2024). Pseudomonadota and Bacteroidota were detected in the community of the HMA and previously affiliated with the degradation of algal-derived polysaccharides (Oni et al. 2015). As described above, we only detected Bacteroidota as active *in situ* (Chapter II). Though not detected in this dissertation, Verrucomicrobiota were abundant in a previous 16S rRNA gene-based community study in the HMA surface sediments and, therefore, could be involved in polysaccharide degradation in the HMA (Oni et al. 2015). During the degradation of polysaccharides, these microorganisms use carbohydrate active enzymes (CAZymes), such as glycoside hydrolases, glycosyltransferases, polysaccharide lyases and carbohydrate esterases (Drula et al. 2021).

The degradation of marine and terrestrial derived cellulose was affiliated with many Clostridia (Hungate 1944, Patel et al. 1980, Khan et al. 1984, Sleat et al. 1984, Schwarz 2001, Kato et al. 2004, Shiratori et al. 2009, Rettenmaier et al. 2019, Rettenmaier et al. 2021). Besides Clostridia, members of the Actinomycetota and Bacteroidota are involved in cellulolytic metabolisms (Goksøyr 1988, Li and Gao 1997, Lykidis et al. 2007, Abt et al. 2010). During cellulose degradation, the large polysaccharides are converted by cellulases, including endoglucanases, exoglucanases and cellobiases into monomeric glucose (Lynd et al. 2002). In this dissertation, we did not detect specific cellulolytic microorganisms. However, some members of the genus *Marinifilum* and *Fusibacter* that were active *in situ* have the potential to degrade cellobiose, a cellulose degradation product and thus are involved in the degradation of organic matter of marine and possible terrestrial origin (Ravot et al. 1999, Ruvira et al. 2013, Smii et al. 2015, Fu et al. 2018).

The results on the primary fermentation of complex organic carbon obtained in this dissertation identified known primary fermenters, such as the *Ca*. Prometheoarchaeum syntrophicum strain MK-D1, *Propionigenium, Marinifilum* and *Fusibacter*. Besides these known fermenters, the amendment of protein and methoxylated aromatic compounds as carbon sources revealed uncharacterized primary fermenting organisms, namely multiple uncharacterized members of the class Clostridia and the class EX4484-6 of the phylum Thermoplasmatota. Thus, the initial third hypothesis that due to the lack of knowledge on primary fermenting microorganisms,

using different carbon sources in anoxic marine sediments will reveal uncharacterized primary fermenting organisms besides already known taxa is true.

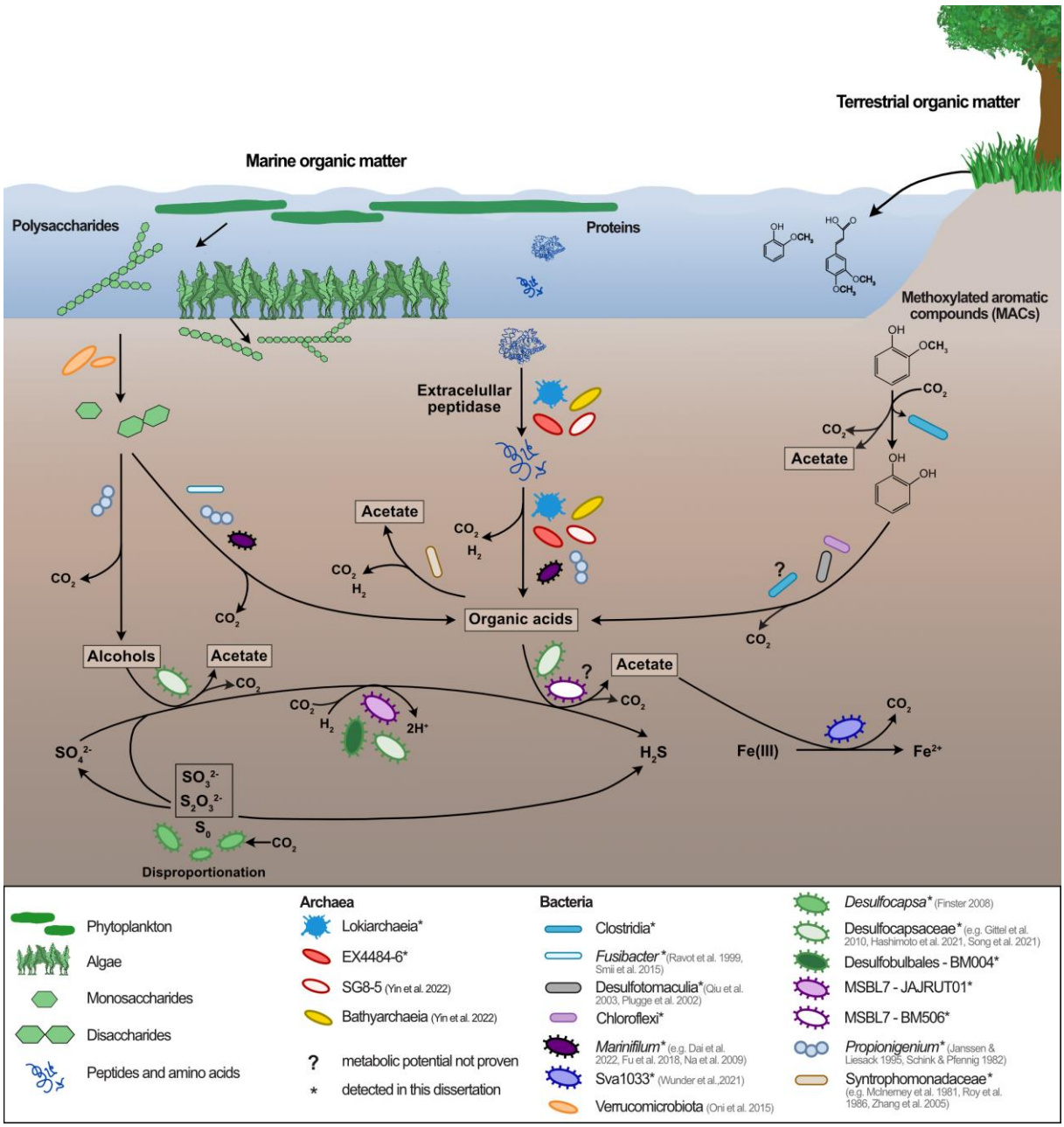

Figure 2 Schematic overview of biotic processes in the HMA as described in this dissertation. Taxa marked with an asterisk have been identified in this dissertation. Bacteria marked with a question mark were suggested to participate in these processes; however, their metabolic potential has yet to be described. Literature is indicated for taxa we did not analyze the metabolic potential of.

Moreover, these uncharacterized primary fermenting organisms were not detected in the sediment controls and became abundant upon substrate addition only. Hence, these members are expected to be present in the sediment matrix; however, they are only present in low abundances. Nevertheless, Clostridia were identified as key players in transforming methoxylated aromatic compounds. Overall, these results suggest that members of the

Clostridia are low-abundant microbial taxa that might play an important role in the degradation of organic matter. Therefore, for the group of Clostridia the fourth hypothesis that besides main players in anoxic marine sediments, low abundant microbial taxa are important for the degradation of organic matter is true. The EX4484-6 became abundant in antibiotics-amended enrichments, thus requiring the suppression of bacteria. Further analyses indicated that members of the class EX4484-6 are rare biosphere members. The importance and role of the EX4484-6 class in marine sediments are yet to be determined. However, the presence of this class in organic matter rich environments globally suggests that this group plays an essential role in such environments, despite its low abundance.

5.5 Novelty among well-represented phyla

As the main objective of this dissertation, I aimed to detect undescribed microorganisms involved in the degradation of organic matter. As previously mentioned, throughout the Chapters we detected many uncharacterized and unclassified microorganisms. In Chapter II, we identified multiple members of the class Desulfobulbia that could not be classified at the genus level, following the SILVA taxonomy. In total, we identified 368 ASVs affiliated with the class Desulfobulbia. Yet, only 92 (25%) of these ASVs were classified at the genus level (Fig. 3a). Of the three identified families, Desulfobulbaceae had the highest numbers of unclassified ASVs on the genus level (138 ASVs, 51%), followed by Desulfocapsaceae (52 ASVs, 14%) and Desulfurivibrionaceae (5 ASVs, 2%). A total of 81 (22%) of the identified ASVs were unclassified on family level, indicating high taxonomic novelty in this order. We later compared some highly abundant unclassified ASVs with genomes present in the GTDB version r214 (Parks et al. 2018). We could retrieve a more resolved taxonomic classification as described before (e.g. unclassified Desulfobulbales were sorted into the family BM004). Yet, most of these taxa remain uncharacterized, and their metabolic potential has not been explored.

Similarly, Clostridia are among the best-represented taxonomic groups in the GTDB version r220, comprising 80,141 genomes (Parks et al. 2018). By SILVA Clostridia are sorted into the phylum Firmicutes. The Firmicutes are among the phyla having the highest OTU richness in the SILVA database (Louca et al. 2019). This high diversity was connected with their ability to colonize many different environments and hosts (Louca et al. 2019). Firmicutes were especially abundant in humans, but new undescribed lineages were also detected in insects, indicating the hidden diversity of this phylum (Schulz et al. 2017). However, despite the high representation of Clostridia in the GTDB and other databases, most ASVs and MAGs detected in Chapter IV could not be classified below the family level and are therefore regarded as

unclassified (Fig. 3b). We could detect 620 ASVs, of which 20% could not be assigned to any family and 27% could not be assigned to any genus. 192 ASVs (31%) were even unclassified on the order level. In all five orders, identified in Chapter IV as the most relative abundant, high taxonomic novelty was observed (21 ASVs Acetivibrionales, 48 ASVs Eubacteriales, 37 ASVs Oscillospirales, 77 ASVs Peptostreptococcales and 9 ASVs Tissierellales). This high taxonomic novelty might derive from Clostridia not being prevalent in the marine environment and thus understudied. Based on Zinger et al. (2011) Clostridia are abundant mostly in coastal benthic areas. A study investigating the global diversity of microbial communities in marine sediments identified Firmicutes as being present in both anoxic and oxic environments. Yet, this group showed low relative abundances in most studied sediment cores (Hoshino et al. 2020). Overall, Clostridia are thought to be more common in terrestrial habitats (Ruff et al. 2024).

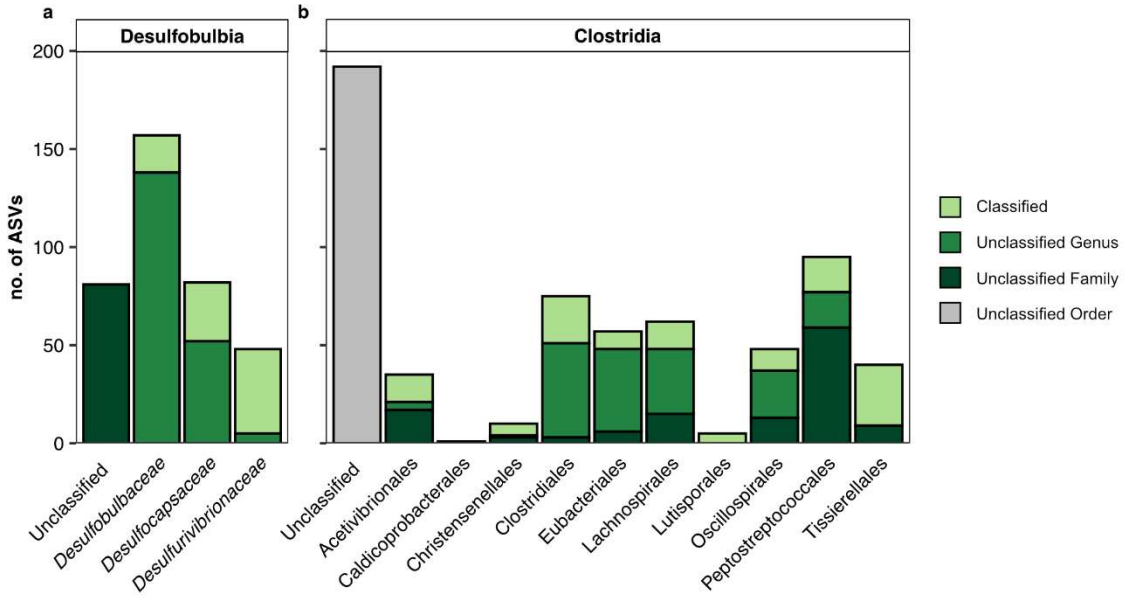

Figure 3 Classified and unclassified ASVs detected in (**a**) Chapter II and (**b**) Chapter IV. All Desulfobulbia ASVs of Chapter II were affiliated with the order Desulfobulbales. Code used to create this figure is deposited at https://github.com/mmaeke/Thesis_Discussion.

The representation of bacterial genomes on public databases, such as the GTDB, is far greater than that of archaea (bacteria: 584,382 vs. archaea: 12,477) (Parks et al. 2018). While the best represented bacterial phylum Pseudomonadota counts 214,930 genomes, the best represented archaeal phylum Thermoplasmatota, only counts 3,169 genomes (Fig. 4). Such contrasts between bacterial and archaeal diversity have been described before (Aller and Kemp 2008). In most environments, bacteria are thought to be more diverse than archaea (Aller and Kemp 2008). However, archaea are thought to play a significant role, especially in subseafloor

environments (Arndt et al. 2013, Hoshino and Inagaki 2018, Orsi et al. 2020), hinting towards a yet understudied diversity in the domain of archaea.

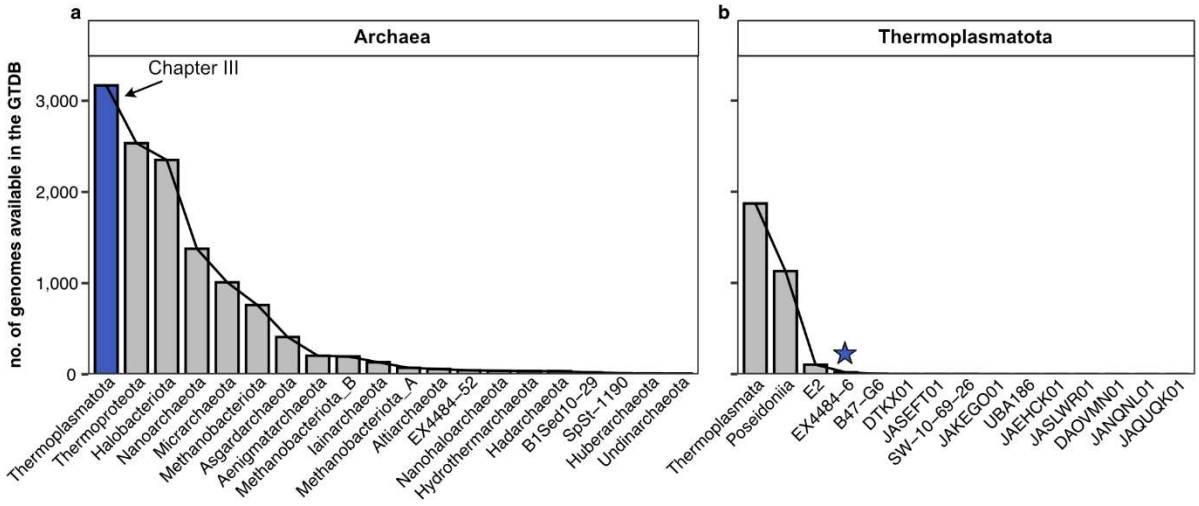

Figure 4 Rank-abundance of **(a)** the 20 most abundant archaeal phyla and **(b)** all classes of the phylum Thermoplasmatota. The star indicates the detected EX4484-6 class. The data is derived from the Genome Taxonomy Database (GTDB) version r220 and shows the counts of genomes available per taxonomic group (Parks et al. 2018). Code used to create this figure is deposited at https://github.com/mmaeke/Thesis_Discussion.

The representation of Thermoplasmatota as the best-represented archaeal phylum likely derives from the abundance and prevalence of this phylum in many different environments, such as acidic soils and acid mine drainages (Baker and Banfield 2003, Yuan et al. 2021, Sheridan et al. 2022), hydrothermal vent fields (Reysenbach et al. 2006, Dombrowski et al. 2018), marine sediments (Lloyd et al. 2013, Zhou et al. 2019, Zheng et al. 2022) and the marine water column (Li et al. 2015, Martin-Cuadrado et al. 2015, Zhang et al. 2015, Rinke et al. 2019). Upon closer inspection of the rank abundances, only the classes Thermoplasmata (1,874) and Poseidoniia (1,132) contribute to this high representation (Fig. 4b). Various orders of the class Thermoplasmata have been described, including the Aciduliprofundales (Reysenbach et al. 2006), Halarchaeoplasmatales (Zhou et al. 2022, 2023a), Methanomassiliicoccales (Iino et al. 2013), Sysuiplasmatales (Yuan et al. 2021), Thermoplasmatales (Huber and Stetter 2006), *Ca*. Gimiplasmatales (Hu et al. 2021), *Ca*. Angelarchaeales (Diamond et al. 2022), *Ca*. Lutacidiplasmatales (Sheridan et al. 2022) and SG8-5 (Yin et al. 2022). Poseidoniia, with its two orders, is predominantly found in the marine water column (Massana et al. 1997, Massana et al. 2000, Li et al. 2015, Rinke et al. 2019, Tully 2019). Members of Thermoplasmata and Poseidoniia are highly abundant in their respective environments (Parkes et al. 2014, Compte-Port et al. 2017, Rinke et al. 2019, Tully 2019). While analyzing the abundance of classes within the Thermoplasmatota in Chapter III, we identified the orders *Ca*. Lutacidiplasmatales

and Thermoplasmatales of the class Thermoplasmata and the order *Ca*. Poseidoniales of the class Poseidoniia to be abundant with up to 40% relative abundance in various samples, thus indicating similar observations as previously described. However, we also detected the undescribed class EX4484-6 of the phylum Thermoplasmatota in protein-amended enrichments. The representation of this class now comprises 22 genomes in the GTDB and has increased from previously five published genomes since our study in Chapter III (Fig. 4b). Through extensive screening (Chapter III), we identified three novel orders, increasing the representation of this class. Still, present genomes in the GTDB are, so far, only of one order. With 22 publicly available genomes in the GTDB (Parks et al. 2018) and additional 35 genomes retrieved in our study (Chapter III), this class represents the fourth most abundant class of the phylum Thermoplasmatota. Additional 11 classes of the Thermoplasmatota phylum are only represented by one to seven genomes and thus, together with the EX4484-6, remain underrepresented in this public database.

The finding of large numbers of ASVs of unclassified members among well-represented phyla, like sulfur-compound utilizing Desulfobacterota, fermenting Bacillota_A or Thermoplasmatota as described in this dissertation, hints that despite continuing advances in sequencing technologies, including deeper sequencing, significant phylogenetic diversity remains to be explored. The high genome representation of different taxonomic groups likely indicates high abundance in the environment, leading to the primary discovery of these groups. For example, the phylum Pseudomonadota comprises the well-represented classes Alphaproteobacteria (31,226 genomes) and Gammaproteobacteria (183,508 genomes). Both of these classes are among the most abundant groups in marine habitats (Zinger et al. 2011, Bienhold et al. 2016) but are also abundant in other habitats, such as soil (Spain et al. 2009). Most likely, this high genome representation of some groups is also derived from the culturability of these microorganisms, e.g., *Escherichia coli,* counting 38,926 genomes present in the GTDB (Parks et al. 2018).

On the other hand, low abundant taxa likely remain in the understudied fractions, such as the rare biosphere and are regularly overlooked. Such underrepresentation of taxa might derive from factors such as high intraspecific competition, a narrow environmental niche, or predation (Jousset et al. 2017). These factors can influence the abundance of microorganisms, leading to conditionally rare taxa (Shade et al. 2014). Similarly, we observed that most detected undescribed taxa, such as the Clostridia and EX4484-6 were of low abundance or not detected in sediment controls and only stimulated upon substrate addition. The group EX4484-6 was

overall of low relative abundance in various habitats and could, therefore, be classified as a rare biosphere member (Chapter III). These detected taxa likely inhabit a narrow environmental niche, being dependent on substrate availability. Results obtained in this dissertation indicate that *in situ* studies alone are insufficient to target the unknown phylogenetic diversity. As stated in Chapter III, analyses of time series could be beneficial to detect conditionally rare taxa. Further methods, as applied in this dissertation, including ^{13}C-RNA stable isotope probing or enrichment experiments with single carbon substrates, could reveal *in situ* not detectable taxa and shed more light onto novel members in otherwise well-represented phyla. Following up on the detection of undescribed and rare microorganisms, data-mining was suggested to improve the recovery and characterization of taxa (Lynch and Neufeld 2015).

5.6 Data-integrative studies as a chance to characterize undescribed taxa

The Genomic Encyclopedia of Bacteria and Archaea was initiated as the first study to illuminate the diversity of underrepresented strains (Wu et al. 2009). Since then, multiple studies have been conducted, performing large-scale sampling to reconstruct novel genomes from environmental samples (Nelson et al. 2010, Anantharaman et al. 2016, Liu et al. 2022, Han et al. 2023). However, such sampling campaigns are highly time- and cost-consuming. Only in recent years have people become more aware of the available data in databases and recognized the opportunities lying in this data. As the first large-scale data-driven study, 8000 metagenome assembled genomes (MAGs) were reconstructed from metagenomic samples of diverse environments available in the International Nucleotide Sequence Database Collaboration databases (INSDC) and expanded the bacterial and archaeal phylogeny by more than 30% (Parks et al. 2017). From this initial study, the Genome Taxonomy Database was created, including 87,106 bacterial genomes available on INSDC databases and 11,603 MAGs reconstructed from openly published metagenomic samples (Parks et al. 2018). Since then, the representation of genomes on the GTDB has increased, covering 596,859 genomes of the domains archaea and bacteria as of release r220 (December, 2024) (Parks et al. 2021). After this first model study, the genomic catalog of Earth's microbiomes (Nayfach et al. 2021), the Ocean Microbiomics Database (Paoli et al. 2022) and SPIRE (Schmidt et al. 2023) added numerous novel groups to the tree of life. Data mining, as has been used by these studies, can best be described as analyzing large amounts of data and transforming information retrieved from this data into knowledge. The mining of data allows the utilization of data that regularly remains unanalyzed, as most studies only utilize fractions of published data. Such data mining can be focused on 'omics data, utilizing genome assemblies, metagenomic short read data,

amplicon sequencing and metatranscriptomic and metabolomic datasets, as has been shown by these previously mentioned studies, but it can go beyond that and include environmental data, remote sensing/satellite data or ocean/climate models, to name a few. This dissertation focused on the use of ‘omics data mining.

The studies mentioned above captured the diversity of novel taxa and phylogenetic groups, but data-integrative studies can also be applied to single taxonomic groups. Using available databases, such as the INSDC or GTDB, taxon-targeted data mining can easily be applied by retrieving the genomes of a group of interest for phylogenetic or phylogenomic comparison. For instance, a recent large study revealed five undescribed phylum-level groups by placing novel detected MAGs into a phylogenomic context using 4,000 reference genomes retrieved from such databases (Gong et al. 2022). In Chapters II, III and IV of this dissertation, we applied a similar approach by retrieving genome assemblies of the taxonomic classes of interest that were present in GTDB (Fig. 5). By using extracted 16S rRNA genes from these genomes, which included large numbers of uncultivated strains, we could improve the phylogenetic representation of these uncultivated taxa in our data set and identified microorganisms closely related to the active organisms of our enrichments. However, the information contained in retrieved genomes is not limited to the use in 16S rRNA gene phylogenetic studies. Using the same set of genomes, MAGs retrieved in Chapters III and IV were placed in a marker gene tree, allowing the taxonomic comparison of ASVs and those MAGs lacking 16S rRNA genes by identifying genomes that were closely related to both the ASVs and MAGs. Moreover, in studies lacking genomic data, such as studies based on 16S rRNA gene surveys, the novel retrieved taxonomic information can be used to screen publications in which the closest related genomes were described. By analyzing the metabolic capabilities of these strains, ASVs can be placed into a metabolic and functional context. The closest associated genomes could further be used for pathway reconstructions, thus elevating the study with additional genomic data. Despite the increase of novel taxonomic groups on INSDC databases and GTDB, only a fraction of microorganisms has been characterized. Many taxa remain undescribed and only a few or no published genomes are available. Data-integrative approaches that were applied to mine environmental genomes for their metabolic functionality detected high novel functionality, emphasizing the need to target the undescribed microorganisms (Pavlopoulos et al. 2023, Rodríguez del Río et al. 2024). The representation and characterization of undescribed taxa can be tackled by a data mining approach targeting not only group-specific genome assemblies present on GTDB and INSDC databases, such as the European Nucleotide Archive

(ENA), but additional genome assemblies that fall into the NCBI taxon ID's of ecological or unclassified metagenomes, including metagenomes of cold seeps, estuaries, glaciers or mine drainages and lack a taxonomic affiliation (Fig. 5). Furthermore, available metagenomic raw read data can be mapped against a set of MAGs of the target group to identify samples in which the target group is abundant. Such an approach was conducted in a study on a novel group containing high numbers of biosynthetic gene clusters, leading to the recovery of 19 MAGs of the formerly unrepresented group *Ca*. Eudoremicrobiaceae (Paoli et al. 2022). We applied this strategy in Chapter III and identified novel MAGs within the class EX4484-6, which were further used to mine an additional 57.8 TB of metagenomic short-read data and reconstruct novel EX4484-6 MAGs, extensively increasing the representation of this class from formerly two to four orders.

Besides the identification of short read data sets to reconstruct novel MAGs of the target group, mapping of short reads can be used to provide an overview of the biogeography and environmental distribution of the target group. In Chapter III, we used this mapping approach to identify the environmental abundance and biogeography of the class EX4484-6. Previous studies applied a similar strategy to retrieve the abundance and biogeography of novel bacterial phyla; however, using coverage information derived from mapping 16S rRNA gene sequences onto the IMG/M 16S rRNA public assembled metagenomes database instead of mapping short read data onto derived MAGs (Chen et al. 2016). More recently, novel resources derived from large-scale data mining focused on microbial abundance in diverse environments (Woodcroft et al. 2024). By computing the relative abundances of bacteria and archaea in a large set of short-read metagenomic data sets (248,559 metagenomic samples) without the need for MAG reconstruction, Sandpiper (https://sandpiper.qut.edu.au) was created. This novel tool now enables the quick targeting of specific taxa by removing the need to map large data sets of metagenomic short read data onto the target group (Woodcroft et al. 2024). While this tool offers excellent advances, metagenomic studies regularly detect novel organisms, and the description of these still requires targeted data mining to reconstruct novel MAGs from metagenomic short read data. For a broader ecological context, metadata supplied with genome assemblies and metagenomic short read data can be accessed, such as done in an analysis of physicochemical parameters and the correlation of this data with the abundance of *Ca*. Poseidoniales in the marine environment (Rinke et al. 2019). However, metadata is often missing. Therefore, not all data is suitable for data mining. The MIxS standards, which specify the minimum information about any sequence (Yilmaz et al. 2011), try to tackle such

inconsistencies. Still, large amounts of data available currently are missing important (and machine-readable) ecological information, affiliated with the environment the samples derive from (Hassenrück et al. 2021).

Combined with knowledge of the environmental context, the lifestyles and traits of the target group can be evaluated. Moreover, the target group can be set into a broader taxonomic scope to assess common and unique genetic features of the group. Such data mining is regularly applied to metagenomic comparative studies. Publicly available genomes are recruited and used to compare the metabolic potential of target groups with other members of the group or more distantly related phylogenetic groups (Fig. 5). On a small scale, this can include pangenomic analyses, with newly generated data and mined MAGs. On a larger scale, such approaches were used to, e.g. screen homolog extracellular peptidases in various archaeal groups (Yin et al. 2022), compare the metabolic potential of *Ca*. Poseidoniales MAGs (Rinke et al. 2019), or evaluate the evolution of key metabolic features in the phylum of Thermoplasmatota (Sheridan et al. 2022). We applied this strategy in two different extents. In Chapter III, we compared the whole metabolic potential of the EX4484-6 class with other Thermoplasmatota to detect differences in the distribution of known and unknown genes and, by this, detected increased novel functionality in the EX4484-6, indicating evolutionary distance from other known Thermoplasmatota. In Chapter IV this approach was applied to only genes of methyltransferases (MTI), comparing the metabolic potential of Clostridia MAGs reconstructed in this dissertation with Clostridia species representatives present on GTDB. By conducting this analysis, we detected similar metabolic functionality for members of the whole class and validated our findings. Furthermore, we identified more undescribed groups with the same metabolic potential, indicating a widespread yet mostly unexplored metabolic potential for methoxydotrophy in Clostridia.

In this dissertation, we combined the approaches that had been used in a variety of data-driven studies. We developed a data mining strategy to evaluate the taxonomic, metabolic and environmental context, needed to characterize yet undescribed taxa (Fig. 5). Overall, data mining has proven valuable in the three Chapters by (II) assessing the taxonomic and metabolic context to describe the potential activity of unclassified bacteria, (III) assessing the taxonomic, metabolic and environmental context to describe the lifestyles and traits of the class EX4484-6, along with their functional novelty and biogeography, and (IV) assessing the functional context and identifying common and unique features among the whole class of Clostridia. Thus, all studies demonstrated the potential to use data mining to characterize undescribed

microorganisms and address broader research objectives than possible without additional data resources. Considerable amounts of data can be obtained from databases, eliminating the need for additional costly and time-consuming sampling campaigns. Therefore, data mining can become a time- and cost-efficient way to answer ecological and open research questions.

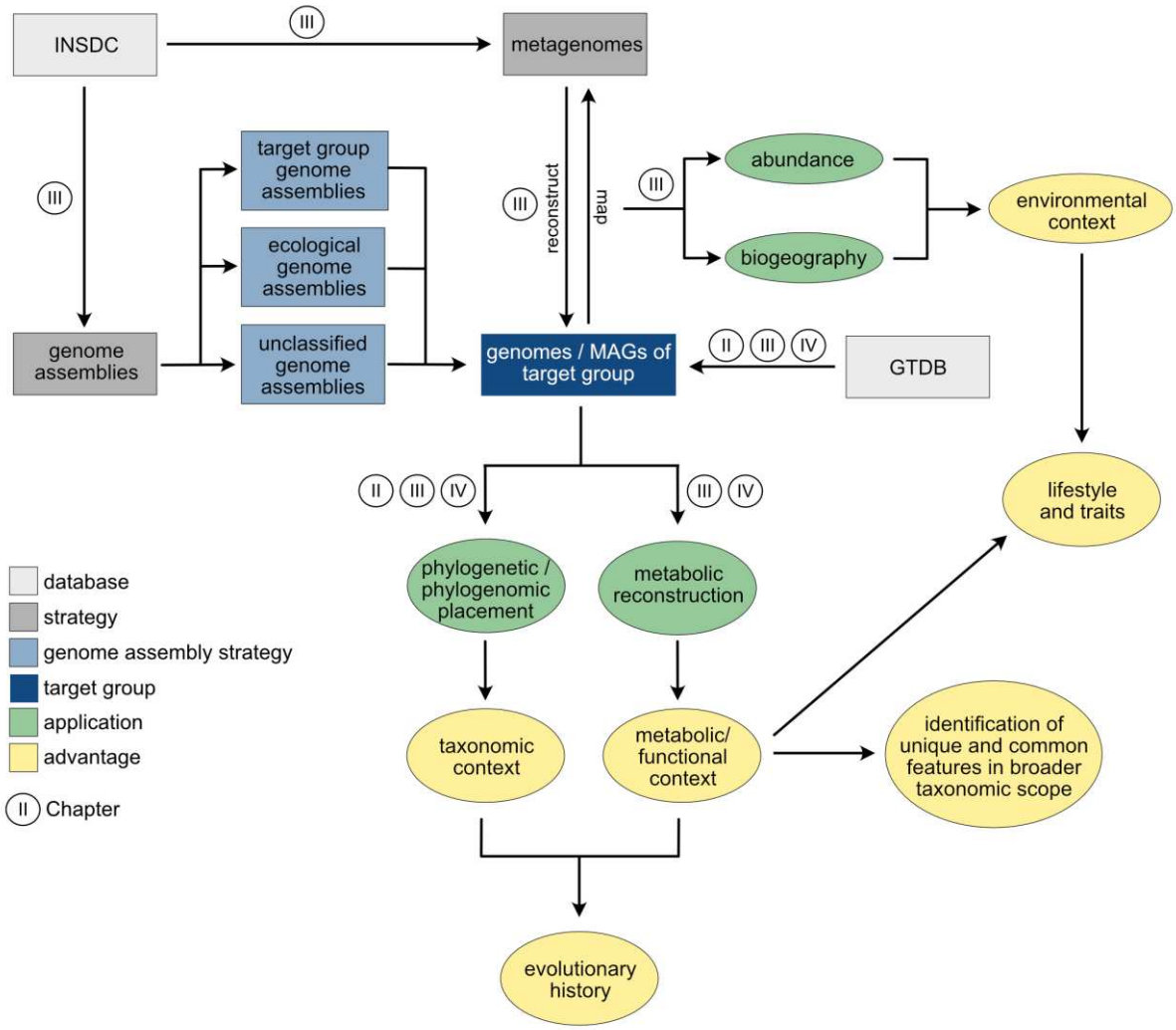

Figure 5 Schematic overview of the data mining strategy, applications and advantages. INSDC = International Nucleotide Sequence Database Collaboration, GTDB = Genome Taxonomy Database.

5.7 Concluding remarks

This dissertation aimed to identify and characterize yet uncultivated and undescribed microorganisms involved in the degradation of complex organic matter to understand microbial processes in anoxic environments better. By combining cultivation-independent methods, including enrichments, stable isotope probing, metagenomics and group-targeted data mining, undescribed carbon-cycling autotrophic and heterotrophic taxa were identified.

While 16S rRNA gene amplicon-based analyses to describe microbial communities provided the first insights into the microbial diversity in the Helgoland mud area, the application of a

novel strategy for anoxic small-scale RNA-SIP with $H_2{}^{18}O$ and RNA-SIP using ^{13}C-DIC as the sole carbon source shed more light onto the dominant sulfur-compound cycling organisms (Chapter II and Chapter IV). Most taxa were identified as autotrophic, challenging previous assumptions based on 16S rRNA gene sequencing. The study further detected different lifestyles in the MSBL7 cluster, providing first insights into the different ecological roles of members of this cluster. While these first attempts to utilize $H_2{}^{18}O$ RNA-SIP detected multiple members involved in cryptic sulfur cycling, this method did not allow to link taxa to processes, such as sulfur compound disproportionation. Further research is required to study these processes that are still hardly understood. By stimulating the anoxic marine sediments with different complex carbon substrates, undescribed microorganisms involved in the primary fermentation of these carbon substrates became abundant and active. Archaea and bacteria, undetected *in situ,* became predominant, indicating that low-abundant taxa in marine sediments can become essential for organic carbon degradation. We identified the previously undescribed class EX4484-6 as a protein degrader involved in the degradation of marine organic matter in globally distributed coastal marine areas (Chapter III). High functional novelty in this group gave first hints towards evolutionary distance to other Thermoplasmatota. While we did not further study the functionality of these genes in this dissertation, novel metabolic capabilities could be hidden in this class. Multiple so far undetected members of the Clostridia, usually common in terrestrial and host-associated environments, became predominant in the degradation of terrestrial-derived methoxylated aromatic compounds (Chapter IV). These findings elevate the knowledge of methoxydotrophic Clostridia in anoxic marine sediments and indicate that methoxydotrophy in marine sediments is a widespread trait in coastal sediments. The finding of the general metabolic potential for C_1-carbon compound metabolism in many members of the Clostridia hinted towards the metabolic potential that has yet to be explored.

Throughout this dissertation, we identified many novel phylogenetic groups, showing us first glimpses into the understudied microbial diversity hidden in marine sediments. While we are far from close to unraveling the microbial diversity, this dissertation offered first insights into some groups that were uncharacterized to date.

5.8 References

Abt, B., Foster, B., Lapidus, A., Clum, A., Sun, H., Pukall, R., Lucas, S., Glavina Del Rio, T., Nolan, M., Tice, H., et al. (2010). Complete genome sequence of *Cellulomonas flavigena* type strain (134^T). Stand. Genomic Sci. **3**:15-25. doi: 10.4056/sigs.1012662.

Adair, K. and Schwartz, E. (2011). Chapter seven - Stable isotope probing with ^{18}O-water to investigate growth and mortality of ammonia oxidizing bacteria and archaea in soil.*in* M. G. Klotz, editor. Methods in Enzymology. Academic Press. p. 155-169. doi: 10.1016/B978-0-12-381294-0.00007-9.

Ahn, Y.-B., Chae, J.-C., Zylstra Gerben, J. and Häggblom Max, M. (2009). Degradation of phenol via phenylphosphate and carboxylation to 4-hydroxybenzoate by a newly isolated strain of the sulfate-reducing bacterium *Desulfobacterium anilini*. Appl. Environ. Microbiol. **75**:4248-4253. doi: 10.1128/AEM.00203-09.

Allen, T. D., Caldwell, M. E., Lawson, P. A., Huhnke, R. L. and Tanner, R. S. (2010). *Alkalibaculum bacchi* gen. nov., sp. nov., a CO-oxidizing, ethanol-producing acetogen isolated from livestock-impacted soil. Int. J. Syst. Evol. Microbiol. **60**:2483-2489. doi: 10.1099/ijs.0.018507-0.

Aller, J. Y. and Kemp, P. F. (2008). Are archaea inherently less diverse than bacteria in the same environments? FEMS Microbiol. Ecol. **65**:74-87. doi: 10.1111/j.1574-6941.2008.00498.x.

Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., Thomas, B. C., Singh, A., Wilkins, M. J., Karaoz, U., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. **7**:13219. doi: 10.1038/ncomms13219.

Arndt, S., Jørgensen, B. B., LaRowe, D. E., Middelburg, J. J., Pancost, R. D. and Regnier, P. (2013). Quantifying the degradation of organic matter in marine sediments: a review and synthesis. Earth-Sci. Rev. **123**:53-86. doi: 10.1016/j.earscirev.2013.02.008.

Aromokeye, D. A., Oni, O. E., Tebben, J., Yin, X., Richter-Heitmann, T., Wendt, J., Nimzyk, R., Littmann, S., Tienken, D., Kulkarni, A. C., et al. (2021a). Crystalline iron oxides stimulate methanogenic benzoate degradation in marine sediment-derived enrichment cultures. ISME J. **15**:965-980. doi: 10.1038/s41396-020-00824-7.

Aromokeye, D. A., Willis-Poratti, G., Wunder, L. C., Yin, X., Wendt, J., Richter-Heitmann, T., Henkel, S., Vázquez, S., Elvert, M., Mac Cormack, W., et al. (2021b). Macroalgae degradation promotes microbial iron reduction via electron shuttling in coastal Antarctic sediments. Environ. Int. **156**:106602. doi: 10.1016/j.envint.2021.106602.

Bache, R. and Pfennig, N. (1981). Selective isolation of *Acetobacterium woodii* on methoxylated aromatic acids and determination of growth yields. Arch. Microbiol. **130**:255-261. doi: 10.1007/BF00459530.

Bak, F. and Cypionka, H. (1987). A novel type of energy metabolism involving fermentation of inorganic sulphur compounds. Nature **326**:891-892. doi: 10.1038/326891a0.

Baker, B. J. and Banfield, J. F. (2003). Microbial communities in acid mine drainage. FEMS Microbiol. Ecol. **44**:139-152. doi: 10.1016/s0168-6496(03)00028-x.

Baloza, M., Henkel, S., Kasten, S., Holtappels, M. and Molari, M. (2023). The impact of sea ice cover on microbial communities in Antarctic shelf sediments. Microorganisms **11**. doi: 10.3390/microorganisms11061572.

Barnum, T. P., Figueroa, I. A., Carlström, C. I., Lucas, L. N., Engelbrektson, A. L. and Coates, J. D. (2018). Genome-resolved metagenomics identifies genetic mobility, metabolic interactions, and unexpected diversity in perchlorate-reducing communities. ISME J. **12**:1568-1581. doi: 10.1038/s41396-018-0081-5.

Berner, R. A. (1982). Burial of organic carbon and pyrite sulfur in the modern ocean: its geochemical and environmental significance. Am. J. Sci. **282**:451-473. doi: 10.1016/0016-7037(83)90151-5.

Bienhold, C., Zinger, L., Boetius, A. and Ramette, A. (2016). Diversity and biogeography of bathyal and abyssal seafloor bacteria. PLoS One **11**:e0148016. doi: 10.1371/journal.pone.0148016.

Blazejak, A. and Schippers, A. (2010). High abundance of JS-1- and Chloroflexi-related bacteria in deeply buried marine sediments revealed by quantitative, real-time PCR. FEMS Microbiol. Ecol. **72**:198-207. doi: 10.1111/j.1574-6941.2010.00838.x.

Blazewicz, S. J., Barnard, R. L., Daly, R. A. and Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME J. **7**:2061-2068. doi: 10.1038/ismej.2013.102.

Braun, S., Mhatre, S. S., Jaussi, M., Røy, H., Kjeldsen, K. U., Pearce, C., Seidenkrantz, M. S., Jørgensen, B. B. and Lomstein, B. A. (2017). Microbial turnover times in the deep seabed studied by amino acid racemization modelling. Sci. Rep. **7**:5680. doi: 10.1038/s41598-017-05972-z.

Brüchert, V. (1998). Early diagenesis of sulfur in estuarine sediments: the role of sedimentary humic and fulvic acids. Geochim. Cosmochim. Acta **62**:1567-1586. doi: 10.1016/S0016-7037(98)00089-1.

Burdige, D. J. (2007). Preservation of organic matter in marine sediments: controls, mechanisms, and an imbalance in sediment organic carbon budgets? Chem. Rev. **107**:467-485. doi: 10.1021/cr050347q.

Cai, M., Liu, Y., Yin, X., Zhou, Z., Friedrich, M. W., Richter-Heitmann, T., Nimzyk, R., Kulkarni, A., Wang, X., Li, W., et al. (2020). Diverse Asgard archaea including the novel phylum Gerdarchaeota participate in organic matter degradation. Sci. China Life Sci. **63**:886-897. doi: 10.1007/s11427-020-1679-1.

Carmona, M., Zamarro, M. T., Blázquez, B., Durante-Rodríguez, G., Juárez, J. F., Valderrama, J. A., Barragán, M. J., García, J. L. and Díaz, E. (2009). Anaerobic catabolism of aromatic compounds: a genetic and genomic view. Microbiol. Mol. Biol. Rev. **73**:71-133. doi: 10.1128/mmbr.00021-08.

Chen, I.-M. A., Markowitz, V. M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Andersen, E., Huntemann, M., et al. (2016). IMG/M: integrated genome and metagenome comparative data analysis system. Nucleic Acids Res. **45**:D507-D516. doi: 10.1093/nar/gkw929.

Compte-Port, S., Subirats, J., Fillol, M., Sànchez-Melsió, A., Marcé, R., Rivas-Ruiz, P., Rosell-Melé, A. and Borrego, C. M. (2017). Abundance and co-distribution of widespread marine archaeal lineages in surface sediments of freshwater qater bodies across the Iberian Peninsula. Microb. Ecol. **74**:776-787. doi: 10.1007/s00248-017-0989-8.

Dai, W., Sun, W., Fu, T., Jia, C., Cui, H., Han, Y., Shi, X. and Zhang, X. H. (2022). *Marinifilum caeruleilacunae* sp. nov., isolated from Yongle Blue Hole in the South China Sea. Int. J. Syst. Evol. Microbiol **72**. doi: 10.1099/ijsem.0.005358.

Davidova, I. A., Gieg, L. M., Duncan, K. E. and Suflita, J. M. (2007). Anaerobic phenanthrene mineralization by a carboxylating sulfate-reducing bacterial enrichment. ISME J. **1**:436-442. doi: 10.1038/ismej.2007.48.

Davis, T. A., Volesky, B. and Mucci, A. (2003). A review of the biochemistry of heavy metal biosorption by brown algae. Water Res. **37**:4311-4330. doi: 10.1016/S0043-1354(03)00293-8.

Deja-Sikora, E., Gołębiewski, M., Kalwasińska, A., Krawiec, A., Kosobucki, P. and Walczak, M. (2019). Comamonadaceae OTU as a remnant of an ancient microbial community in sulfidic waters. Microb. Ecol. **78**:85-101. doi: 10.1007/s00248-018-1270-5.

DeWeerd, K. A., Mandelco, L., Tanner, R. S., Woese, C. R. and Suflita, J. M. (1990). *Desulfomonile tiedjei* gen. nov. and sp. nov., a novel anaerobic, dehalogenating, sulfate-reducing bacterium. Arch. Microbiol. **154**:23-30. doi:

Diamond, S., Lavy, A., Crits-Christoph, A., Matheus Carnevali, P. B., Sharrar, A., Williams, K. H. and Banfield, J. F. (2022). Soils and sediments host Thermoplasmata archaea encoding novel copper membrane monooxygenases (CuMMOs). ISME J. **16**:1348-1362. doi: 10.1038/s41396-021-01177-5.

Dombrowski, N., Teske, A. P. and Baker, B. J. (2018). Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. Nat. Commun. **9**:4999. doi: 10.1038/s41467-018-07418-0.

Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B. and Terrapon, N. (2021). The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. **50**:D571-D577. doi: 10.1093/nar/gkab1045.

Durkin, C. A., Mock, T. and Armbrust, E. V. (2009). Chitin in diatoms and its association with the cell wall. Eukaryot. Cell **8**:1038-1050. doi: 10.1128/ec.00079-09.

Dyksma, S., Bischof, K., Fuchs, B. M., Hoffmann, K., Meier, D., Meyerdierks, A., Pjevac, P., Probandt, D., Richter, M., Stepanauskas, R., et al. (2016). Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. ISME J. **10**:1939-1953. doi: 10.1038/ismej.2015.257.

Evans, P. N., Parks, D. H., Chadwick, G. L., Robbins, S. J., Orphan, V. J., Golding, S. D. and Tyson, G. W. (2015). Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. Science **350**:434. doi: 10.1126/science.aac7745.

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S. G., González, J. M. and Pedrós-Alió, C. (2013). Ecology of marine Bacteroidetes: a comparative genomics approach. ISME J. **7**:1026-1037. doi: 10.1038/ismej.2012.169.

Finster, K. (2008). Microbiological disproportionation of inorganic sulfur compounds. J. Sulfur Chem. **29**:281-292. doi: 10.1080/17415990802105770.

Fraser, M., Gareth, S. K., Hon Lun, W., Ray, C. and Brendan, P. B. (2019). Asgard archaea: diversity, function, and evolutionary implications in a range of microbiomes. AIMS Microbiol. **5**:48-61. doi: 10.3934/microbiol.2019.1.48.

Fu, T., Jia, C., Fu, L., Zhou, S., Yao, P., Du, R., Sun, H., Yang, Z., Shi, X. and Zhang, X.-H. (2018). *Marinifilum breve* sp. nov., a marine bacterium isolated from the Yongle Blue Hole in the South China Sea and emended description of the genus *Marinifilum*. Int. J. Syst. Evol. Microbiol. **68**:3540-3545. doi: 10.1099/ijsem.0.003027.

Gibson, J. and Harwood, C. S. (2002). Metabolic diversity in aromatic compound utilization by anaerobic microbes. Annu. Rev. Microbiol. **56**:345-369. doi: 10.1146/annurev.micro.56.012302.160749.

Gittel, A., Seidel, M., Kuever, J., Galushko, A. S., Cypionka, H. and Könneke, M. (2010). *Desulfopila inferna* sp. nov., a sulfate-reducing bacterium isolated from the subsurface of a tidal sand-flat. Int. J. Syst. Evol. Microbiol. **60**:1626-1630. doi: 10.1099/ijs.0.015644-0.

Goksøyr, J. (1988). Cellulases from *Sporocytophaga myxococcoides*. *In*: Methods in Enzymology. Academic Press. p. 338-342. doi: 10.1016/0076-6879(88)60136-4.

Gong, X., del Río, Á. R., Xu, L., Chen, Z., Langwig, M. V., Su, L., Sun, M., Huerta-Cepas, J., De Anda, V. and Baker, B. J. (2022). New globally distributed bacterial phyla within the FCB superphylum. Nat. Commun. **13**:7516. doi: 10.1038/s41467-022-34388-1.

Gooday, G. W. (1990). The ecology of chitin degradation.*in* K. C. Marshall, editor. Advances in Microbial Ecology. Springer US, Boston, MA. p. 387-430. doi: 10.1007/978-1-4684-7612-5_10.

Grech-Mora, I., Fardeau, M.-L., Patel, B., Ollivier, B., Rimbault, A., Prensier, G., Garcia, J.-L. and Garnier-Sillam, E. (1996). Isolation and characterization of *Sporobacter termitidis* gen. nov., sp. nov., from the digestive tract of the wood-feeding termite *Nasutitermes lujae*. Int. J. Syst. Evol. Microbiol. **46**:512-518. doi: 10.1099/00207713-46-2-512.

Greening, R. C. and Leedle, J. A. (1989). Enrichment and isolation of *Acetitomaculum ruminis*, gen. nov., sp. nov.: acetogenic bacteria from the bovine rumen. Arch. Microbiol. **151**:399-406. doi: 10.1007/bf00416597.

Han, Y., Zhang, C., Zhao, Z., Peng, Y., Liao, J., Jiang, Q., Liu, Q., Shao, Z. and Dong, X. (2023). A comprehensive genomic catalog from global cold seeps. Sci. Data **10**:596. doi: 10.1038/s41597-023-02521-4.

Hashimoto, Y., Tame, A., Sawayama, S., Miyazaki, J., Takai, K. and Nakagawa, S. (2021). *Desulfomarina profundi* gen. nov., sp. nov., a novel mesophilic, hydrogen-oxidizing, sulphate-reducing chemolithoautotroph isolated from a deep-sea hydrothermal vent chimney. Int J Syst Evol Microbiol **71**. doi: 10.1099/ijsem.0.005083.

Hassenrück, C., Poprick, T., Helfer, V., Molari, M., Meyer, R. and Kostadinov, I. (2021). FAIR enough? A perspective on the status of nucleotide sequence data and metadata on public archives. PREPRINT bioRxiv. doi: 10.1101/2021.09.23.461561.

He, Y., Li, M., Perumal, V., Feng, X., Fang, J., Xie, J., Sievert, S. M. and Wang, F. (2016). Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. Nat. Microbiol. **1**:16035. doi: 10.1038/nmicrobiol.2016.35.

Hoshino, T., Doi, H., Uramoto, G.-I., Wörmer, L., Adhikari, R. R., Xiao, N., Morono, Y., D'Hondt, S., Hinrichs, K.-U. and Inagaki, F. (2020). Global diversity of microbial communities in marine sediment. Proc. Natl. Acad. Sci. U. S. A. **117**:27587-27597. doi: 10.1073/pnas.1919139117.

Hoshino, T. and Inagaki, F. (2018). Abundance and distribution of archaea in the subseafloor sedimentary biosphere. ISME J. **13**:227-231. doi: 10.1038/s41396-018-0253-3.

Hou, J., Wang, Y., Zhu, P., Yang, N., Liang, L., Yu, T., Niu, M., Konhauser, K., Woodcroft, B. J. and Wang, F. (2023). Taxonomic and carbon metabolic diversification of Bathyarchaeia during its coevolution history with early Earth surface environment. Sci. Adv. **9**:eadf5069. doi: 10.1126/sciadv.adf5069.

Hu, W., Pan, J., Wang, B., Guo, J., Li, M. and Xu, M. (2021). Metagenomic insights into the metabolism and evolution of a new Thermoplasmata order (*Candidatus* Gimiplasmatales). Environ. Microbiol. **23**:3695-3709. doi: 10.1111/1462-2920.15349.

Huber, H. and Stetter, K. O. (2006). Thermoplasmatales.*in* M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt, editors. The Prokaryotes: Volume 3: Archaea. Bacteria: Firmicutes, Actinomycetes. Springer New York, New York, NY. p. 101-112. doi: 10.1007/0-387-30743-5_7.

Hungate, R. E. (1944). Studies on cellulose fermentation: I. The culture and physiology of an anaerobic cellulose-digesting bacterium. J. Bacteriol. **48**:499-513. doi: 10.1128/jb.48.5.499-513.1944.

Iino, T., Tamaki, H., Tamazawa, S., Ueno, Y., Ohkuma, M., Suzuki, K., Igarashi, Y. and Haruta, S. (2013). *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. doi: 10.1264/jsme2.me12189.

Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., et al. (2020). Isolation of an archaeon at the prokaryote–eukaryote interface. Nature **577**:519-525. doi: 10.1038/s41586-019-1916-6.

Isaksen, M. F. and Teske, A. (1996). *Desulforhopalus vacuolatus* gen. nov., sp. nov., a new moderately psychrophilic sulfate-reducing bacterium with gas vacuoles isolated from a temperate estuary. Arch. Microbiol. **166**:160-168. doi: 10.1007/s002030050371.

Jain, A. and Krishnan, K. P. (2017). A glimpse of the diversity of complex polysaccharide-degrading culturable bacteria from Kongsfjorden, Arctic Ocean. Ann. Microbiol. **67**:203-214. doi: 10.1007/s13213-016-1252-0.

Janssen, P. H. and Liesack, W. (1995). Succinate decarboxylation by *Propionigenium maris* sp. nov., a new anaerobic bacterium from an estuarine sediment. Arch. Microbiol. **164**:29-35. doi: 10.1007/bf02568731.

Jørgensen, B. B. (1982). Mineralization of organic matter in the sea bed - the role of sulphate reduction. Nature **296**:643-645. doi: 10.1038/296643a0.

Jørgensen, B. B., Findlay, A. J. and Pellerin, A. (2019). The biogeochemical sulfur cycle of marine sediments. Front. Microbiol. **10**:849. doi: 10.3389/fmicb.2019.00849.

Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., Küsel, K., Rillig, M. C., Rivett, D. W., Salles, J. F., et al. (2017). Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J. **11**:853-862. doi: 10.1038/ismej.2016.174.

Junghare, M. and Schink, B. (2015). *Desulfoprunum benzoelyticum* gen. nov., sp. nov., a Gram-stain-negative, benzoate-degrading, sulfate-reducing bacterium isolated from a wastewater treatment plant. Int J Syst Evol Microbiol **65**:77-84. doi: 10.1099/ijs.0.066761-0.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. **44**:D457-462. doi: 10.1093/nar/gkv1070.

Kato, S., Haruta, S., Cui, Z. J., Ishii, M., Yokota, A. and Igarashi, Y. (2004). *Clostridium straminisolvens* sp. nov., a moderately thermophilic, aerotolerant and cellulolytic bacterium isolated from a cellulose-degrading bacterial community. Int. J. Syst. Evol. Microbiol. **54**:2043-2047. doi: 10.1099/ijs.0.63148-0.

Kaufmann, F., Wohlfarth, G. and Diekert, G. (1998). *O*-demethylase from *Acetobacterium dehalogenans* - substrate specificity and function of the participating proteins. Eur. J. Biochem. **253**:706-711. doi: 10.1046/j.1432-1327.1998.2530706.x.

Khan, A. W., Meek, E., Sowden, L. C. and Colvin, J. R. (1984). Emendation of the genus Acetivibrio and description of *Acetivibrio cellulosolvens* sp. nov., a nonmotile cellulolytic mesophile. Int. J. Syst. Evol. Microbiol. **34**:419-422. doi: 10.1099/00207713-34-4-419.

Khomyakova, M., Merkel, A., Novikov, A., Klyukina, A. and Slobodkin, A. (2021). *Alkalibacter mobilis* sp. nov., an anaerobic bacterium isolated from a coastal lake. Int. J. Syst. Evol. Microbiol. **71**:005174. doi: 10.1099/ijsem.0.005174.

Khomyakova, M., Merkel, A., Petrova, D., Bonch-Osmolovskaya, E. and Slobodkin, A. (2020). *Alkalibaculum sporogenes* sp. nov., isolated from a terrestrial mud volcano and emended description of the genus *Alkalibaculum*. Int. J. Syst. Evol. Microbiol. **70**:4914-4919. doi: 10.1099/ijsem.0.004361.

Khomyakova, M. A., Merkel, A. Y., Mamiy, D. D., Klyukina, A. A. and Slobodkin, A. I. (2023). Phenotypic and genomic characterization of *Bathyarchaeum tardum* gen. nov., sp. nov., a cultivated representative of the archaeal class Bathyarchaeia. Front. Microbiol. **14**. doi: 10.3389/fmicb.2023.1214631.

Khomyakova, M. A. and Slobodkin, A. I. (2023). Transformation of methoxylated aromatic compounds by anaerobic microorganisms. Microbiol. **92**:97-118. doi: 10.1134/S0026261722603128.

Konstantinidis, K. T., Rosselló-Móra, R. and Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. ISME J. **11**:2399-2406. doi: 10.1038/ismej.2017.113.

Kreft, J.-U. and Schink, B. (1993). Demethylation and degradation of phenylmethylethers by the sulfide-methylating homoacetogenic bacterium strain TMBS 4. Arch. Microbiol. **159**:308-315. doi: 10.1007/BF00290912.

Kubo, K., Lloyd, K. G., J, F. B., Amann, R., Teske, A. and Knittel, K. (2012). Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. ISME J. **6**:1949-1965. doi: 10.1038/ismej.2012.37.

Kurth, J. M., Nobu, M. K., Tamaki, H., de Jonge, N., Berger, S., Jetten, M. S. M., Yamamoto, K., Mayumi, D., Sakata, S., Bai, L., et al. (2021). Methanogenic archaea use a bacteria-like methyltransferase system to demethoxylate aromatic compounds. ISME J. **15**:3549-3565. doi: 10.1038/s41396-021-01025-6.

Kurth, J. M., Op den Camp, H. J. M. and Welte, C. U. (2020). Several ways one goal - methanogenesis from unconventional substrates. Appl. Microbiol. Biotechnol. **104**:6839-6854. doi: 10.1007/s00253-020-10724-7.

Lazar, C. S., Baker, B. J., Seitz, K., Hyde, A. S., Dick, G. J., Hinrichs, K. U. and Teske, A. P. (2016). Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. Environ. Microbiol. **18**:1200-1211. doi: 10.1111/1462-2920.13142.

Lazar, C. S., Baker, B. J., Seitz, K. W. and Teske, A. P. (2017). Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches. ISME J. **11**:1118-1129. doi: 10.1038/ismej.2016.189.

Lazar, C. S., Biddle, J. F., Meador, T. B., Blair, N., Hinrichs, K. U. and Teske, A. P. (2015). Environmental controls on intragroup diversity of the uncultured benthic archaea of the Miscellaneous

Crenarchaeotal group lineage naturally enriched in anoxic sediments of the White Oak River estuary (North Carolina, USA). Environ. Microbiol. **17**:2228-2238. doi: 10.1111/1462-2920.12659.

Lechtenfeld, M., Heine, J., Sameith, J., Kremp, F. and Müller, V. (2018). Glycine betaine metabolism in the acetogenic bacterium *Acetobacterium woodii*. Environ. Microbiol. **20**:4512-4525. doi: 10.1111/1462-2920.14389.

Lengeler, J. W., Drews, G. and Schlegel, H. G., editors. (1999). Biology of the Prokaryotes. Georg Thieme Verlag, Germany.

Lenk, S., Arnds, J., Zerjatke, K., Musat, N., Amann, R. and Mussmann, M. (2011). Novel groups of Gammaproteobacteria catalyse sulfur oxidation and carbon fixation in a coastal, intertidal sediment. Environ. Microbiol. **13**:758-774. doi: 10.1111/j.1462-2920.2010.02380.x.

Li, M., Baker, B. J., Anantharaman, K., Jain, S., Breier, J. A. and Dick, G. J. (2015). Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. Nat. Commun. **6**:8933. doi: 10.1038/ncomms9933.

Li, X. and Gao, P. (1997). Isolation and partial properties of cellulose-decomposing strain of *Cytophaga* sp. LX-7 from soil. J. Appl. Microbiol. **82**:73-80. doi: 10.1111/j.1365-2672.1997.tb03299.x.

Lie, T. J., Clawson, M. L., Godchaux, W. and Leadbetter, E. R. (1999). Sulfidogenesis from 2-aminoethanesulfonate (Taurine) fermentation by a morphologically unusual sulfate-reducing bacterium, *Desulforhopalus singaporensis* sp. nov. Appl. Environ. Microbiol. **65**:3328-3334. doi: 10.1128/AEM.65.8.3328-3334.1999.

Liu, Y., Ji, M., Yu, T., Zaugg, J., Anesio, A. M., Zhang, Z., Hu, S., Hugenholtz, P., Liu, K., Liu, P., et al. (2022). A genome and gene catalog of glacier microbiomes. Nat. Biotechnol. **40**:1341-1348. doi: 10.1038/s41587-022-01367-2.

Liu, Y., Zhou, Z., Pan, J., Baker, B. J., Gu, J.-D. and Li, M. (2018). Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. ISME J. **12**:1021-1031. doi: 10.1038/s41396-018-0060-x.

Lloyd, K. G., Schreiber, L., Petersen, D. G., Kjeldsen, K. U., Lever, M. A., Steen, A. D., Stepanauskas, R., Richter, M., Kleindienst, S., Lenk, S., et al. (2013). Predominant archaea in marine sediments degrade detrital proteins. Nature **496**:215-218. doi: 10.1038/nature12033.

Lomstein, B. A., Langerhuus, A. T., D'Hondt, S., Jørgensen, B. B. and Spivack, A. J. (2012). Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. Nature **484**:101-104. doi: 10.1038/nature10905.

Louca, S., Mazel, F., Doebeli, M. and Parfrey, L. W. (2019). A census-based estimate of Earth's bacterial and archaeal diversity. PLoS Biol. **17**:e3000106. doi: 10.1371/journal.pbio.3000106.

Lux, M. F. and Drake, H. L. (1992). Re-examination of the metabolic potentials of the acetogens *Clostridium aceticum* and *Clostridium formicoaceticum*: chemolithoautotrophic and aromatic-dependent growth. FEMS Microbiol. Lett. **95**:49-56. doi: 10.1016/0378-1097(92)90735-7.

Lykidis, A., Mavromatis, K., Ivanova, N., Anderson, I., Land, M., DiBartolo, G., Martinez, M., Lapidus, A., Lucas, S., Copeland, A., et al. (2007). Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX. J. Bacteriol. **189**:2477-2486. doi: 10.1128/jb.01899-06.

Lynch, M. D. J. and Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. Nat. Rev. Microbiol. **13**:217-229. doi: 10.1038/nrmicro3400.

Lynd, L. R., Weimer, P. J., van Zyl, W. H. and Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. **66**:506-577. doi: 10.1128/mmbr.66.3.506-577.2002.

Mann, A. J., Hahnke, R. L., Huang, S., Werner, J., Xing, P., Barbeyron, T., Huettel, B., Stüber, K., Reinhardt, R., Harder, J., et al. (2013). The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides. Appl. Environ. Microbiol. **79**:6813-6822. doi: 10.1128/aem.01937-13.

Martin-Cuadrado, A. B., Garcia-Heredia, I., Moltó, A. G., López-Úbeda, R., Kimes, N., López-García, P., Moreira, D. and Rodriguez-Valera, F. (2015). A new class of marine Euryarchaeota group II from the Mediterranean deep chlorophyll maximum. ISME J. **9**:1619-1634. doi: 10.1038/ismej.2014.249.

Massana, R., DeLong Edward, F. and Pedrós-Alió, C. (2000). A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. Appl. Environ. Microbiol. **66**:1777-1787. doi: 10.1128/AEM.66.5.1777-1787.2000.

Massana, R., Murray, A. E., Preston, C. M. and DeLong, E. F. (1997). Vertical distribution and phylogenetic characterization of marine planktonic archaea in the Santa Barbara Channel. Appl. Environ. Microbiol. **63**:50-56. doi: 10.1128/aem.63.1.50-56.1997.

Mayumi, D., Mochimaru, H., Tamaki, H., Yamamoto, K., Yoshioka, H., Suzuki, Y., Kamagata, Y. and Sakata, S. (2016). Methane production from coal by a single methanogen. Science **354**:222-225. doi: 10.1126/science.aaf8821.

McInerney, M., Bryant, M., Hespell, R. and Costerton, J. (1981). *Syntrophomonas wolfei* gen. nov. sp. nov., an anaerobic, syntrophic, fatty acid-oxidizing bacterium. Appl. Environ. Microbiol. **41**:1029-1039. doi: 10.1128/aem.41.4.1029-1039.1981.

McKee, L. S., La Rosa, S. L., Westereng, B., Eijsink, V. G., Pope, P. B. and Larsbrink, J. (2021). Polysaccharide degradation by the Bacteroidetes: mechanisms and nomenclature. Environ. Microbiol. Rep. **13**:559-581. doi: 10.1111/1758-2229.12980.

Mechichi, T., Labat, M., Garcia, J. L., Thomas, P. and Patel, B. K. (1999). *Sporobacterium olearium* gen. nov., sp. nov., a new methanethiol-producing bacterium that degrades aromatic compounds, isolated from an olive mill wastewater treatment digester. Int. J. Syst. Bacteriol. **49 Pt 4**:1741-1748. doi: 10.1099/00207713-49-4-1741.

Meng, J., Xu, J., Qin, D., He, Y., Xiao, X. and Wang, F. (2014). Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. ISME J. **8**:650-659. doi: 10.1038/ismej.2013.174.

Müller, D., Liu, B., Geibert, W., Holtappels, M., Sander, L., Miramontes, E., Taubner, H., Henkel, S., Hinrichs, K. U., Bethke, D., et al. (2024). Depositional controls and budget of organic carbon burial in fine-grained sediments of the North Sea - the Helgoland mud area as a test field. PREPRINT EGUsphere. doi: 10.5194/egusphere-2024-1632.

Musat, F., Galushko, A., Jacob, J., Widdel, F., Kube, M., Reinhardt, R., Wilkes, H., Schink, B. and Rabus, R. (2009). Anaerobic degradation of naphthalene and 2-methylnaphthalene by strains of marine sulfate-reducing bacteria. Environ. Microbiol. **11**:209-219. doi: 10.1111/j.1462-2920.2008.01756.x.

Musat, F. and Widdel, F. (2008). Anaerobic degradation of benzene by a marine sulfate-reducing enrichment culture, and cell hybridization of the dominant phylotype. Environ. Microbiol. **10**:10-19. doi: 10.1111/j.1462-2920.2007.01425.x.

Na, H., Kim, S., Moon, E. Y. and Chun, J. (2009). *Marinifilum fragile* gen. nov., sp. nov., isolated from tidal flat sediment. Int. J. Syst. Evol. Microbiol. **59**:2241-2246. doi: 10.1099/ijs.0.009027-0.

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I. M., Huntemann, M., et al. (2021). A genomic catalog of Earth's microbiomes. Nat. Biotechnol. **39**:499-509. doi: 10.1038/s41587-020-0718-6.

Neidhardt, F. C., Ingraham, J. L. and Schaechter, M., editors. (1990). Physiology of the bacterial cell. A molecular approach. Sinauer Associates Inc., U.S., Sunderland, MA.

Nelson, K. E., Weinstock, G. M., Highlander, S. K., Worley, K. C., Creasy, H. H., Wortman, J. R., Rusch, D. B., Mitreva, M., Sodergren, E., Chinwalla, A. T., et al. (2010). A catalog of reference genomes from the human microbiome. Science **328**:994-999. doi: 10.1126/science.1183605.

Oni, O. E., Schmidt, F., Miyatake, T., Kasten, S., Witt, M., Hinrichs, K.-U. and Friedrich, M. W. (2015). Microbial communities and organic matter composition in surface and subsurface sediments of the Helgoland mud area, North Sea. Front. Microbiol. **6**. doi: 10.3389/fmicb.2015.01290.

Orsi, W. D., Schink, B., Buckel, W. and Martin, W. F. (2020). Physiological limits to life in anoxic subseafloor sediment. FEMS Microbiol. Rev. **44**:219-231. doi: 10.1093/femsre/fuaa004.

Orsi, W. D., Smith, J. M., Liu, S., Liu, Z., Sakamoto, C. M., Wilken, S., Poirier, C., Richards, T. A., Keeling, P. J., Worden, A. Z., et al. (2016). Diverse, uncultivated bacteria and archaea underlying the cycling of dissolved protein in the ocean. ISME J. **10**:2158-2173. doi: 10.1038/ismej.2016.20.

Paarup, M., Friedrich, M. W., Tindall, B. J. and Finster, K. (2006). Characterization of the psychrotolerant acetogen strain SyrA5 and the emended description of the species *Acetobacterium carbinolicum*. Antonie van Leeuwenhoek **89**:55-69. doi: 10.1007/s10482-005-9009-y.

Pachiadaki, M. G., Yakimov, M. M., LaCono, V., Leadbetter, E. and Edgcomb, V. (2014). Unveiling microbial activities along the halocline of Thetis, a deep-sea hypersaline anoxic basin. ISME J. **8**:2478-2489. doi: 10.1038/ismej.2014.100.

Paoli, L., Ruscheweyh, H.-J., Forneris, C. C., Hubrich, F., Kautsar, S., Bhushan, A., Lotti, A., Clayssen, Q., Salazar, G., Milanese, A., et al. (2022). Biosynthetic potential of the global ocean microbiome. Nature **607**:111-118. doi: 10.1038/s41586-022-04862-3.

Parekh, M., Keith, E. S., Daniel, S. L. and Drake, H. L. (1992). Comparative evaluation of the metabolic potentials of different strains of *Peptostreptococcus productus*: utilization and transformation of aromatic compounds. FEMS Microbiol. Lett. **73**:69-74. doi: 10.1016/0378-1097(92)90585-c.

Parkes, R. J., Cragg, B., Roussel, E., Webster, G., Weightman, A. and Sass, H. (2014). A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere-geosphere interactions. Mar. Geol. **352**:409-425. doi: 10.1016/j.margeo.2014.02.009.

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A. and Hugenholtz, P. (2021). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. **50**:D785-D794. doi: 10.1093/nar/gkab776.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A. and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. **36**:996-1004. doi: 10.1038/nbt.4229.

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P. and Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. **2**:1533-1542. doi: 10.1038/s41564-017-0012-7.

Patel, G. B., Khan, A. W., Agnew, B. J. and Colvin, J. R. (1980). Isolation and characterization of an anaerobic, cellulolytic microorganism, *Acetivibrio cellulolyticus* gen. nov., sp. nov.†. Int. J. Syst. Evol. Microbiol. **30**:179-185. doi: 10.1099/00207713-30-1-179.

Pavlopoulos, G. A., Baltoumas, F. A., Liu, S., Selvitopi, O., Camargo, A. P., Nayfach, S., Azad, A., Roux, S., Call, L., Ivanova, N. N., et al. (2023). Unraveling the functional dark matter through global metagenomics. Nature **622**:594-602. doi: 10.1038/s41586-023-06583-7.

Pelikan, C., Wasmund, K., Glombitza, C., Hausmann, B., Herbold, C. W., Flieder, M. and Loy, A. (2021). Anaerobic bacterial degradation of protein and lipid macromolecules in subarctic marine sediment. ISME J. **15**:833-847. doi: 10.1038/s41396-020-00817-6.

Pierce, E., Xie, G., Barabote, R. D., Saunders, E., Han, C. S., Detter, J. C., Richardson, P., Brettin, T. S., Das, A. and Ljungdahl, L. G. (2008). The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). Environ. Microbiol. **10**:2550-2573. doi: 10.1111/j.1462-2920.2008.01679.x.

Plugge, C. M., Balk, M. and Stams, A. J. (2002). *Desulfotomaculum thermobenzoicum* subsp. *thermosyntrophicum* subsp. nov., a thermophilic, syntrophic, propionate-oxidizing, spore-forming bacterium. Int. J. Syst. Evol. Microbiol. **52**:391-399. doi: 10.1099/00207713-52-2-391.

Qiu, D., Zeng, X., Zeng, L., Li, G. and Shao, Z. (2021). *Fusibacter ferrireducens* sp. nov., an anaerobic, Fe(III)- and sulphur-reducing bacterium isolated from mangrove sediment. Int. J. Syst. Evol. Microbiol. **71**. doi: 10.1099/ijsem.0.004952.

Qiu, Y. L., Sekiguchi, Y., Imachi, H., Kamagata, Y., Tseng, I. C., Cheng, S. S., Ohashi, A. and Harada, H. (2003). *Sporotomaculum syntrophicum* sp. nov., a novel anaerobic, syntrophic benzoate-degrading bacterium isolated from methanogenic sludge treating wastewater from terephthalate manufacturing. Arch. Microbiol. **179**:242-249. doi: 10.1007/s00203-003-0521-z.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. **41**:D590-D596. doi: 10.1093/nar/gks1219.

Ravenschlag, K., Sahm, K., Pernthaler, J. and Amann, R. (1999). High bacterial diversity in permanently cold marine sediments. Appl. Environ. Microbiol. **65**:3982-3989. doi: 10.1128/aem.65.9.3982-3989.1999.

Ravot, G., Magot, M., Fardeau, M. L., Patel, B. K., Thomas, P., Garcia, J. L. and Ollivier, B. (1999). *Fusibacter paucivorans* gen. nov., sp. nov., an anaerobic, thiosulfate-reducing bacterium from an oil-producing well. Int. J. Syst. Bacteriol. **49 Pt 3**:1141-1147. doi: 10.1099/00207713-49-3-1141.

Repeta, D. J. (2015). Chapter 2 - Chemical characterization and cycling of dissolved organic matter.*in* D. A. Hansell and C. A. Carlson, editors. Biogeochemistry of Marine Dissolved Organic Matter (Second Edition). Academic Press, Boston. p. 21-63. doi: 10.1016/B978-0-12-405940-5.00002-9.

Rettenmaier, R., Gerbaulet, M., Liebl, W. and Zverlov, V. V. (2019). *Hungateiclostridium mesophilum* sp. nov., a mesophilic, cellulolytic and spore-forming bacterium isolated from a biogas fermenter fed with maize silage. Int. J. Syst. Evol. Microbiol. **69**:3567-3573. doi: 10.1099/ijsem.0.003663.

Rettenmaier, R., Kowollik, M.-L., Klingl, A., Liebl, W. and Zverlov, V. (2021). *Ruminiclostridium herbifermentans* sp. nov., a mesophilic and moderately thermophilic cellulolytic and xylanolytic bacterium isolated from a lab-scale biogas fermenter fed with maize silage. Int. J. Syst. Evol. Microbiol. **71**. doi: 10.1099/ijsem.0.004692.

Reysenbach, A.-L., Liu, Y., Banta, A. B., Beveridge, T. J., Kirshtein, J. D., Schouten, S., Tivey, M. K., Von Damm, K. L. and Voytek, M. A. (2006). A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. Nature **442**:444-447. doi: 10.1038/nature04921.

Rinke, C., Rubino, F., Messer, L. F., Youssef, N., Parks, D. H., Chuvochina, M., Brown, M., Jeffries, T., Tyson, G. W., Seymour, J. R., et al. (2019). A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (*Ca*. Poseidoniales ord. nov.). ISME J. **13**:663-675. doi: 10.1038/s41396-018-0282-y.

Rodríguez del Río, Á., Giner-Lamia, J., Cantalapiedra, C. P., Botas, J., Deng, Z., Hernández-Plaza, A., Munar-Palmer, M., Santamaría-Hernando, S., Rodríguez-Herva, J. J., Ruscheweyh, H.-J., et al. (2024). Functional and evolutionary significance of unknown genes from uncultivated taxa. Nature **626**:377-384. doi: 10.1038/s41586-023-06955-z.

Roy, F., Samain, E., Dubourguier, H. C. and Albagnac, G. (1986). *Synthrophomonas sapovorans* sp. nov., a new obligately proton reducing anaerobe oxidizing saturated and unsaturated long chain fatty acids. Arch. Microbiol. **145**:142-147. doi: 10.1007/BF00446771.

Ruff, S. E., de Angelis, I. H., Mullis, M., Payet, J. P., Magnabosco, C., Lloyd, K. G., Sheik, C. S., Steen, A. D., Shipunova, A., Morozov, A., et al. (2024). A global comparison of surface and subsurface microbiomes reveals large-scale biodiversity gradients, and a marine-terrestrial divide. Sci. Adv. **10**:eadq0645. doi: 10.1126/sciadv.adq0645.

Ruvira, M. A., Lucena, T., Pujalte, M. J., Arahal, D. R. and Macián, M. C. (2013). *Marinifilum flexuosum* sp. nov., a new Bacteroidetes isolated from coastal Mediterranean Sea water and emended description of the genus *Marinifilum* Na et al., 2009. Syst. Appl. Microbiol. **36**:155-159. doi: 10.1016/j.syapm.2012.12.003.

Sajeela, V., Thajudeen, J., B.M, A., Kattatheyil, H., Kabeer. S, S., K.P, K., C.K, R. and Abdulla, M. H. (2023). Diversity of complex polysaccharide degrading bacteria from the sediments of interlinked high Arctic fjords, Svalbard. Reg. Stud. Mar. Sci. **63**:102989. doi: 10.1016/j.rsma.2023.102989.

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., et al. (2022). Database resources of the national center for biotechnology information. Nucleic Acids Res. **50**:d20-d26. doi: 10.1093/nar/gkab1112.

Schink, B. and Pfennig, N. (1982). *Propionigenium modestum* gen. nov. sp. nov. a new strictly anaerobic, nonsporing bacterium growing on succinate. Arch. Microbiol. **133**:209-216. doi: 10.1007/BF00415003.

Schink, B., Philipp, B. and Müller, J. A. (2000). Anaerobic degradation of phenolic compounds. Naturwissenschaften. doi: 10.1007/s001140050002.

Schmidt, T. S. B., Fullam, A., Ferretti, P., Orakov, A., Maistrenko, O. M., Ruscheweyh, H.-J., Letunic, I., Duan, Y., Van Rossum, T., Sunagawa, S., et al. (2023). SPIRE: a Searchable, Planetary-scale mIcrobiome REsource. Nucleic Acids Res. **52**:D777-D783. doi: 10.1093/nar/gkad943.

Schnell, S., Bak, F. and Pfennig, N. (1989). Anaerobic degradation of aniline and dihydroxybenzenes by newly isolated sulfate-reducing bacteria and description of *Desulfobacterium anilini*. Arch. Microbiol. **152**:556-563. doi: 10.1007/bf00425486.

Schulz, F., Eloe-Fadrosh, E. A., Bowers, R. M., Jarett, J., Nielsen, T., Ivanova, N. N., Kyrpides, N. C. and Woyke, T. (2017). Towards a balanced view of the bacterial tree of life. Microbiome **5**:140. doi: 10.1186/s40168-017-0360-9.

Schwarz, W. (2001). The cellulosome and cellulose degradation by anaerobic bacteria. Appl. Microbiol. Biotechnol. **56**:634-649. doi: 10.1007/s002530100710.

Seitz, K. W., Lazar, C. S., Hinrichs, K.-U., Teske, A. P. and Baker, B. J. (2016). Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. ISME J. **10**:1696-1705. doi: 10.1038/ismej.2015.233.

Sekiguchi, Y., Yamada, T., Hanada, S., Ohashi, A., Harada, H. and Kamagata, Y. (2003). *Anaerolinea thermophila* gen. nov., sp. nov. and *Caldilinea aerophila* gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain Bacteria at the subphylum level. Int. J. Syst. Evol. Microbiol. **53**:1843-1851. doi: 10.1099/ijs.0.02699-0.

Shade, A., Jones, S. E., Caporaso, J. G., Handelsman, J., Knight, R., Fierer, N. and Gilbert, J. A. (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. mBio **5**:e01371-01314. doi: 10.1128/mBio.01371-14.

Sheridan, P. O., Meng, Y., Williams, T. A. and Gubry-Rangin, C. (2022). Recovery of Lutacidiplasmatales archaeal order genomes suggests convergent evolution in Thermoplasmatota. Nat. Commun. **13**:4110. doi: 10.1038/s41467-022-31847-7.

Shiratori, H., Sasaya, K., Ohiwa, H., Ikeno, H., Ayame, S., Kataoka, N., Miya, A., Beppu, T. and Ueda, K. (2009). *Clostridium clariflavum* sp. nov. and *Clostridium caenicola* sp. nov., moderately thermophilic, cellulose-/cellobiose-digesting bacteria isolated from methanogenic sludge. Int. J. Syst. Evol. Microbiol. **59**:1764-1770. doi: 10.1099/ijs.0.003483-0.

Sichert, A., Corzett, C. H., Schechter, M. S., Unfried, F., Markert, S., Becher, D., Fernandez-Guerra, A., Liebeke, M., Schweder, T., Polz, M. F., et al. (2020). Verrucomicrobia use hundreds of enzymes to digest the algal polysaccharide fucoidan. Nat. Microbiol. **5**:1026-1039. doi: 10.1038/s41564-020-0720-2.

Sleat, R., Mah, R. A. and Robinson, R. (1984). Isolation and characterization of an anaerobic, cellulolytic bacterium, *Clostridium cellulovorans* sp. nov. Appl. Environ. Microbiol. **48**:88-93. doi: 10.1128/aem.48.1.88-93.1984.

Slobodkin, A. I. and Slobodkina, G. B. (2019). Diversity of sulfur-disproportionating microorganisms. Microbiol. **88**:509-522. doi: 10.1134/S0026261719050138.

Smii, L., Ben Hania, W., Cayol, J. L., Joseph, M., Hamdi, M., Ollivier, B. and Fardeau, M. L. (2015). *Fusibacter bizertensis* sp. nov., isolated from a corroded kerosene storage tank. Int. J. Syst. Evol. Microbiol. **65**:117-121. doi: 10.1099/ijs.0.066183-0.

Song, J., Hwang, J., Kang, I. and Cho, J. C. (2021). A sulfate-reducing bacterial genus, *Desulfosediminicola* gen. nov., comprising two novel species cultivated from tidal-flat sediments. Sci. Rep. **11**:19978. doi: 10.1038/s41598-021-99469-5.

Spain, A. M., Krumholz, L. R. and Elshahed, M. S. (2009). Abundance, composition, diversity and novelty of soil proteobacteria. ISME J. **3**:992-1000. doi: 10.1038/ismej.2009.43.

Stieb, M. and Schink, B. (1985). Anaerobic oxidation of fatty acids by *Clostridium bryantii* sp. nov., a sporeforming, obligately syntrophic bacterium. Arch. Microbiol. **140**:387-390. doi: 10.1007/BF00446983.

Suzuki, D., Ueki, A., Amaishi, A. and Ueki, K. (2007). *Desulfopila aestuarii* gen. nov., sp. nov., a gram-negative, rod-like, sulfate-reducing bacterium isolated from an estuarine sediment in Japan. Int. J. Syst. Evol. Microbiol. **57**:520-526. doi: 10.1099/ijs.0.64600-0.

Tanaka, K. and Pfennig, N. (1988). Fermentation of 2-methoxyethanol by *Acetobacterium malicum* sp. nov. and *Pelobacter venetianus*. Arch. Microbiol. **149**:181-187. doi: 10.1007/BF00422003.

Taylor, B. F. (1983). Aerobic and anaerobic catabolism of vanillic acid and some other methoxy-aromatic compounds by *Pseudomonas* sp. Strain PN-1. Appl. Environ. Microbiol. **46**:1286-1292. doi: 10.1128/aem.46.6.1286-1292.1983.

Thamdrup, B., Finster, K., Hansen, J. W. and Bak, F. (1993). Bacterial disproportionation of elemental sulfur coupled to chemical reduction of iron or manganese. Appl. Environ. Microbiol. **59**:101-108. doi: 10.1128/aem.59.1.101-108.1993.

Tschech, A. and Schink, B. (1985). Fermentative degradation of resorcinol and resorcylic acids. Arch. Microbiol. **143**:52-59. doi: 10.1007/BF00414768.

Tully, B. J. (2019). Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. Nat. Commun. **10**:271. doi: 10.1038/s41467-018-07840-4.

Wang, Y., Thompson, K. N., Yan, Y., Short, M. I., Zhang, Y., Franzosa, E. A., Shen, J., Hartmann, E. M. and Huttenhower, C. (2023). RNA-based amplicon sequencing is ineffective in measuring metabolic activity in environmental microbial communities. Microbiome **11**:131. doi: 10.1186/s40168-022-01449-y.

Wasmund, K., Cooper, M., Schreiber, L., Lloyd, K. G., Baker, B. J., Petersen, D. G., Jørgensen, B. B., Stepanauskas, R., Reinhardt, R., Schramm, A., et al. (2016). Single-cell genome and group-specific dsrAB sequencing implicate marine members of the class Dehalococcoidia (phylum Chloroflexi) in sulfur cycling. mBio **7**. doi: 10.1128/mBio.00266-16.

Wasmund, K., Schreiber, L., Lloyd, K. G., Petersen, D. G., Schramm, A., Stepanauskas, R., Jørgensen, B. B. and Adrian, L. (2014). Genome sequencing of a single cell of the widely distributed marine subsurface Dehalococcoidia, phylum Chloroflexi. ISME J. **8**:383-397. doi: 10.1038/ismej.2013.143.

Wegener, G., Krukenberg, V., Ruff, S. E., Kellermann, M. Y. and Knittel, K. (2016). Metabolic capabilities of microorganisms involved in and associated with the anaerobic oxidation of methane. Front. Microbiol. **7**. doi: 10.3389/fmicb.2016.00046.

Welte, C. U., de Graaf, R., Dalcin Martins, P., Jansen, R. S., Jetten, M. S. M. and Kurth, J. M. (2021). A novel methoxydotrophic metabolism discovered in the hyperthermophilic archaeon *Archaeoglobus fulgidus*. Environ. Microbiol. **23**:4017-4033. doi: 10.1111/1462-2920.15546.

Woodcroft, B. J., Aroney, S. T. N., Zhao, R., Cunningham, M., Mitchell, J. A. M., Blackall, L. and Tyson, G. W. (2024). SingleM and Sandpiper: robust microbial taxonomic profiles from metagenomic data. PREPRINT bioRxiv. doi: 10.1101/2024.01.30.578060.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., Kunin, V., Goodwin, L., Wu, M., Tindall, B. J., et al. (2009). A phylogeny-driven genomic encyclopaedia of bacteria and archaea. Nature **462**:1056-1060. doi: 10.1038/nature08656.

Wunder, L. C., Aromokeye, D. A., Yin, X., Richter-Heitmann, T., Willis-Poratti, G., Schnakenberg, A., Otersen, C., Dohrmann, I., Römer, M., Bohrmann, G., et al. (2021). Iron and sulfate reduction structure microbial communities in (sub-)Antarctic sediments. ISME J. **15**:3587-3604. doi: 10.1038/s41396-021-01014-9.

Yamada, T., Sekiguchi, Y., Hanada, S., Imachi, H., Ohashi, A., Harada, H. and Kamagata, Y. (2006). *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the bacterial phylum Chloroflexi. Int. J. Syst. Evol. Microbiol. **56**:1331-1340. doi: 10.1099/ijs.0.64169-0.

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., Gilbert, J. A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., et al. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat. Biotechnol. **29**:415-420. doi: 10.1038/nbt.1823.

Yin, X., Cai, M., Liu, Y., Zhou, G., Richter-Heitmann, T., Aromokeye, D. A., Kulkarni, A. C., Nimzyk, R., Cullhed, H., Zhou, Z., et al. (2020). Subgroup level differences of physiological activities in marine Lokiarchaeota. ISME J. **15**:848-861. doi: 10.1038/s41396-020-00818-5.

Yin, X., Zhou, G., Cai, M., Zhu, Q.-Z., Richter-Heitmann, T., Aromokeye, D. A., Liu, Y., Nimzyk, R., Zheng, Q., Tang, X., et al. (2022). Catabolic protein degradation in marine sediments confined to distinct archaea. ISME J. **16**:1617-1626. doi: 10.1038/s41396-022-01210-1.

Yin, X., Zhou, G., Wang, H., Han, D., Maeke, M., Richter-Heitmann, T., Wunder, L. C., Aromokeye, D. A., Zhu, Q.-Z., Nimzyk, R., et al. (2024). Unexpected carbon utilization activity of sulfate-reducing microorganisms in temperate and permanently cold marine sediments. ISME J. **18**. doi: 10.1093/ismejo/wrad014.

Yu, T., Hu, H., Zeng, X., Wang, Y., Pan, D., Deng, L., Liang, L., Hou, J. and Wang, F. (2023a). Widespread Bathyarchaeia encode a novel methyltransferase utilizing lignin-derived aromatics. mLife **2**:272-282. doi: 10.1002/mlf2.12082.

Yu, T., Wu, W., Liang, W., Wang, Y., Hou, J., Chen, Y., Elvert, M., Hinrichs, K. U. and Wang, F. (2023b). Anaerobic degradation of organic carbon supports uncultured microbial populations in estuarine sediments. Microbiome **11**:81. doi: 10.1186/s40168-023-01531-z.

Yuan, Y., Liu, J., Yang, T. T., Gao, S. M., Liao, B. and Huang, L. N. (2021). Genomic insights into the ecological role and evolution of a novel Thermoplasmata order, "*Candidatus* Sysuiplasmatales". Appl. Environ. Microbiol. **87**:e0106521. doi: 10.1128/aem.01065-21.

Zhang, C., Liu, X. and Dong, X. (2004). *Syntrophomonas curvata* sp. nov., an anaerobe that degrades fatty acids in co-culture with methanogens. Int. J. Syst. Evol. Microbiol. **54**:969-973. doi: 10.1099/ijs.0.02903-0.

Zhang, C., Liu, X. and Dong, X. (2005). *Syntrophomonas erecta* sp. nov., a novel anaerobe that syntrophically degrades short-chain fatty acids. Int. J. Syst. Evol. Microbiol. **55**:799-803. doi: 10.1099/ijs.0.63372-0.

Zhang, C. L., Xie, W., Martin-Cuadrado, A.-B. and Rodriguez-Valera, F. (2015). Marine Group II archaea, potentially important players in the global ocean carbon cycle. Front. Microbiol. **6**. doi: 10.3389/fmicb.2015.01108.

Zhang, Y. S., Zhang, Y. Q., Zhao, X. M., Liu, X. L., Qin, Q. L., Liu, N. H., Xu, F., Chen, X. L., Zhang, Y. Z. and Li, P. Y. (2024). Metagenomic insights into the dynamic degradation of brown algal polysaccharides by kelp-associated microbiota. Appl. Environ. Microbiol. **90**:e0202523. doi: 10.1128/aem.02025-23.

Zheng, P.-F., Wei, Z., Zhou, Y., Li, Q., Qi, Z., Diao, X. and Wang, Y. (2022). Genomic evidence for the recycling of complex organic carbon by novel Thermoplasmatota clades in deep-sea sediments. mSystems **7**:e00077-00022. doi: 10.1128/msystems.00077-22.

Zhong, Y.-W., Zhou, P., Cheng, H., Zhou, Y.-D., Pan, J., Xu, L., Li, M., Tao, C.-H., Wu, Y.-H. and Xu, X.-W. (2022). Metagenomic features characterized with microbial iron oxidoreduction and mineral interaction in Southwest Indian Ridge. Microbiol. Spectr. **10**:e00614-00622. doi: 10.1128/spectrum.00614-22.

Zhou, H., Zhao, D., Zhang, S., Xue, Q., Zhang, M., Yu, H., Zhou, J., Li, M., Kumar, S. and Xiang, H. (2022). Metagenomic insights into the environmental adaptation and metabolism of *Candidatus* Haloplasmatales, one archaeal order thriving in saline lakes. Environ. Microbiol. **24**:2239-2258. doi: 10.1111/1462-2920.15899.

Zhou, H., Zhao, D., Zhang, S., Xue, Q., Zhang, M., Yu, H., Zhou, J., Li, M., Kumar, S. and Xiang, H. (2023a). Metagenomic insights into the environmental adaptation and metabolism of *Candidatus* Halarchaeoplasmatales, one archaeal order thriving in saline lakes. Environ. Microbiol. **25**:592. doi: 10.1111/1462-2920.16338.

Zhou, Y., Wei, Y., Jiang, L., Jiao, X. and Zhang, Y. (2023b). Anaerobic phloroglucinol degradation by *Clostridium scatologenes*. mBio **14**:e01099-01023. doi: 10.1128/mbio.01099-23.

Zhou, Z., Liu, Y., Lloyd, K. G., Pan, J., Yang, Y., Gu, J.-D. and Li, M. (2019). Genomic and transcriptomic insights into the ecology and metabolism of benthic archaeal cosmopolitan, Thermoprofundales (MBG-D archaea). ISME J. **13**:885-901. doi: 10.1038/s41396-018-0321-8.

Zhou, Z., Pan, J., Wang, F., Gu, J.-D. and Li, M. (2018). Bathyarchaeota: globally distributed metabolic generalists in anoxic environments. FEMS Microbiol. Rev. **42**:639-655. doi: 10.1093/femsre/fuy023.

Zhu, Q.-Z., Yin, X., Taubner, H., Wendt, J., Friedrich, M. W., Elvert, M., Hinrichs, K.-U. and Middelburg, J. J. (2024). Secondary production and priming reshape the organic matter composition in marine sediments. Sci. Adv. **10**:eadm8096. doi: 10.1126/sciadv.adm8096.

Zhuang, L., Tang, Z., Ma, J., Yu, Z., Wang, Y. and Tang, J. (2019). Enhanced anaerobic biodegradation of benzoate under sulfate-reducing conditions with conductive iron-oxides in sediment of Pearl River Estuary. Front. Microbiol. **10**. doi: 10.3389/fmicb.2019.00374.

Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B. M., Martiny, J. B. H., Sogin, M., Boetius, A. and Ramette, A. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. PloS One **6**:e24570. doi: 10.1371/journal.pone.0024570.

Acknowledgments

Firstly, I would like to thank Prof. Michael W. Friedrich, Dr. Tristan Wagner and Prof. William Orsi for agreeing to review this thesis. I would also like to thank Prof. Michael W. Friedrich, Dr. Tristan Wagner, Dr. Christiane Hassenrück, Prof. Kai Bischof, Dr. Karen Krüger and Charlotte Recke for agreeing to be members of my examination board.

Pursuing this PhD would not have been possible without Michael Friedrich, giving me the opportunity to do my PhD in his group and allowing me to explore my scientific path. Thank you for letting me follow a route that included many bioinformatic methods not conducted in your working group prior. I am very thankful for your support, supervision and advice.

I want to express my deepest gratitude to Xiuran Yin. After gaining first insights into the world of anaerobic microbial ecophysiology during the lab rotation you offered, you supervised me in the lab during my PhD and taught me every method I know. From setting up far too many enrichments, cloning, qPCR to teaching me the precious method of SIP, you dedicated your time to teaching me everything you know to become the researcher I am now. I really enjoyed working with you and am very grateful for the time you dedicated to guiding me in the projects of my PhD. Thank you very much!

I would also like to express my deepest appreciation to Christiane Hassenrück. Being supervised by you in the bioinformatic work conducted during my thesis was one of the biggest wins throughout the five years. I really enjoyed our hour-long discussions, late evening coding sessions in my R, and early morning problem-solving sessions. You introduced me to the world of data-mining and I am grateful for all the passion and dedication you had throughout all this time, helping me to overcome more than one low, when I was frustrated. I am really looking forward to continuing to work with you!

Many thanks to my thesis advisory committee members, including Prof. Michael Friedrich, Dr. Xiuran Yin, Dr. Christiane Hassenrück, Dr. Tristan Wagner and Dr. Tim Richter-Heitmann. Thank you for your valuable insights and ideas to further evolve my research.

Thank you also to the co-authors I have worked with throughout the last years. Your valuable input has made the projects in this dissertation possible.

An important thank you to the MPI for Marine Microbiology and the MarMic program for supporting me throughout my Master and PhD. Christiane Glöckner, I am very grateful for all

the effort you put into helping me throughout the years and the time you took to chat about life together.

I could not have undertaken this journey without the Microbial Ecophysiology working group. Special thanks to Charlotte, Nazila and Sabine for keeping our lab afloat even in the most chaotic times. Also, a big thank you goes to my coworkers Lea, Yunru, Wenxuan, Karen, Tim and Darjan for helping out in the lab, sitting together for a coffee or cake and chatting about science and life. Also, a big thank you to the former group members. I would also like to thank the Azubis and Hiwis, who helped immensely during my work. Especially Lea, Charlotte and Vivian, you three worked magic in the lab and allowed me to focus on my bioinformatic work while you continued to do everything that needed to be done in the lab. Thank you as well to my lab rotation and Bachelor students, Mitali, Anna, Nadja, Charlotte, and Jutta, for their great work.

My friends, especially Johanna, Lea, Henny, Murmel, Jenny and Isabell, I could always count on you for mental support, distractions, or just some fun times. Thank you for always being there!

Alistair, my Love, you have been there in the good and hard times. With you by my side, I started the journey of my PhD. I could always count on your support and love, you listened to me explaining my research numerous times and were there whenever I needed to sort my thoughts. And even though you understood only parts of it or nothing at all, you helped me solve my mental blockages. Apart from science, no matter what has come my way, you have supported me in any way possible. You were my rock and helped me to grow throughout these last years. I am beyond grateful to have you by my side. I will also send huge thanks to Loki and Nala, who are the best emotional support cats, bringing me happiness.

Finally, I have been beyond lucky to have the endless support and love of my parents and siblings. Ohne eure Unterstützung würde ich nicht sein, wo ich heute bin. Ihr habt mir vertraut meinen eigenen Weg zu finden und wart immer für mich da, wenn ich ins Stolpern geraten bin. Besonders Mama und Papa, ihr seid Wege mit mir gegangen, die sich keine Eltern jemals hätten ausmalen können. Danke für eure unendliche Liebe!

Universität Bremen

Universität Bremen | Fachbereich 02 | Postfach 33 04 40, 28334 Bremen

Universität Bremen
Fachbereich 2
Prüfungsamt Chemie
z. Hd. Frau Frauke Ernst
Leobener Straße

28359 Bremen
Deutschland

Prüfungsamt
Chemie

Frauke Ernst
Geschäftsstelle
Fachbereich 02
Leobener Str. / NW2
D-28359 Bremen

Verwaltungspavillon 06
Tel. 0421 218-62802
Fax 0421 218-9862802
frauke.ernst@uni-bremen.de
www.fb2.uni-bremen.de

Versicherung an Eides Statt

Name, Vorname	Maeke, Mara
Matrikel-Nr.	4052492
Straße	
Ort, PLZ	

Ich, _____ (Vorname, Name)

versichere an Eides Statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.
Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.
Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

_____
Ort, Datum / Unterschrift

Universität Bremen

Ort, Datum: _____

.

Erklärungen zur elektronischen Version und zur Überprüfung einer Dissertation

Hiermit betätige ich gemäß §7, Abs. 7, Punkt 4, dass die zu Prüfungszwecken beigelegte elektronische Version meiner Dissertation identisch ist mit der abgegebenen gedruckten Version.

Ich bin mit der Überprüfung meiner Dissertation gemäß §6 Abs. 2, Punkt 5 mit qualifizierter Software im Rahmen der Untersuchung von Plagiatsvorwürfen einverstanden.

_____

Unterschrift