1

2 **Helarchaeota and Co-occurring Sulfate-Reducing Bacteria in**
3 **Subseafloor Sediments from the Costa Rica Margin**

4

5 Rui Zhao, Jennifer F. Biddle

6 School of Marine Science and Policy, University of Delaware, Lewes, DE 19958, USA

7

8 # Correspondence: Jennifer Biddle (jfbiddle@udel.edu)

9

10 Running title: New Helarchaeota and co-occurring SRB genomes

11

12 **Abstract**

13 Deep sediments host many archaeal lineages, including those of the Asgard superphylum that
 14 may depend on/require syntrophic partnerships. Our knowledge about sedimentary archaeal
 15 diversity and their metabolic pathways and syntrophic partners is still very limited. We
 16 present here new genomes of Helarchaeota and co-occurring sulfate-reducing bacteria (SRB)
 17 recovered from organic-rich sediments off Costa Rica Margin. Our molecular analyses
 18 revealed three new metagenome-assembled genomes (MAGs) affiliating with Helarchaeota,
 19 each of which has three variants of the methyl-CoM reductase-like (MCR-like) complex that
 20 may enable them to oxidize short-chain alkanes anaerobically. These Helarchaeota have no
 21 multi-heme cytochromes (MHCs) but have Group 3b and Group 3c [NiFe] hydrogenases, and
 22 formate dehydrogenase, and therefore could transfer the reducing equivalents generated from
 23 alkane oxidation to external partners via the transfer of these substances. We also recovered
 24 five MAGs of SRB affiliated with the class of Desulfobacteria, two of which showed relative
 25 abundances (represented by genome coverages) positively correlated with those of the three
 26 Helarchaeota. Genome analysis suggested that these SRB bacteria have the capacity of H₂ and
 27 formate utilizations and may facilitate electron transfers from other organisms by means of
 28 these reduced substances. Our findings suggest that Helarchaeota may metabolize
 29 synergistically with SRB in marine anoxic sediments, and exert an important influence on the
 30 carbon cycle by mitigating the hydrocarbon emission from sediments to the overlying ocean.

31

32 **Introduction**

33 Many of the total microbial cells in the marine realm are estimated to be present in marine
 34 sediments (Kallmeyer et al., 2012), in which a considerable fraction of the microbial cells are
 35 archaea (Lipp et al., 2008; Lloyd et al., 2013a; Hoshino and Inagaki, 2019). Although
 36 archaeal communities in oligotrophic and often oxic sediments are dominated by aerobic

37 ammonia-oxidizing Thaumarchaeota (Vuillemin et al., 2019; Zhao et al., 2019; Hiraoka et al.,
38 2020; Hoshino et al., 2020), those in coastal and often organic-rich sediments are more
39 diverse and complex (Durbin and Teske, 2012), and their metabolic activity exerts a critical
40 influence in the carbon and nutrient cycling on the global scale (e.g. (Biddle et al., 2006;
41 Lloyd et al., 2013b; Yu et al., 2018)).

42 The Asgard superphylum of archaea is of evolutionary importance because its present-
43 day members are thought to share a common ancestor with modern eukaryotes (Spang et al.,
44 2015; Zaremba-Niedzwiedzka et al., 2017). The archaeal phyla of this superphylum, including
45 but not limited to Lokiarchaeota, Thorarchaeota, Odinararchaeota, and Heimdallarchaeota, have
46 been established mainly by metagenome-assembled genomes (Spang et al., 2017; Zaremba-
47 Niedzwiedzka et al., 2017). A recent analysis of available Asgard 16S rRNA gene sequences
48 (Quast et al., 2013) showed that some sequences cannot be resolved to the existing lineages
49 and thus suggested the diversity of Asgard archaea is likely broader than currently recognized
50 (Manoharan et al., 2019). Marine sediments delivered the first cultured Asgard archaeon
51 *Candidatus* Prometheoarchaeum syntrophicum MK-D1 (Imachi et al., 2020) and first
52 identification of the superphylum via metagenomic analysis of DNA (Spang et al., 2015;
53 Zaremba-Niedzwiedzka et al., 2017), so further exploration of marine sediment is promising
54 to recover novel Asgards. This speculation is supported by the recent discovery of
55 Helarchaeota (Seitz et al., 2019), the fifth recognized Asgard phylum, from hydrothermal
56 sediments in the Guaymas Basin (GB).

57 Genome analysis and laboratory cultures suggest that Asgard archaea in marine
58 sediments will need external partner cells to consume the reducing equivalents (e.g., in the
59 forms of hydrogen and formate) generated during the organic matter degradation of Asgard
60 archaea (Spang et al., 2019; Imachi et al., 2020). Helarchaeota from GB encode the Methyl-
61 CoM reductase-like complex (MCR-like complex) and are proposed to be capable of

oxidizing short-chain alkanes to conserve energy (Seitz et al., 2019), using a pathway similar as other thermophilic alkane-oxidizing Euryarchaeota enriched from the same location (Laso-Perez et al., 2016; Chen et al., 2019). Most of these characterized alkane-oxidizing archaea form consortia with sulfate-reducing bacteria (SRB) that use the reducing equivalents released during alkane oxidation for sulfate reduction (Laso-Perez et al., 2016; Chen et al., 2019; Wang et al., 2019), although some alkane-oxidizing archaea can also channel the electrons to their own internal methanogenesis pathway (Laso-Pérez et al., 2019; Wang et al., 2019). As a prevalent SRB in hydrothermal sediments, *Candidatus Desulfofervidus auxilii* (Krukenberg et al., 2016) is the dominant partner bacteria of the characterized alkane-oxidizing archaea in heated marine sediments (Laso-Perez et al., 2016; Chen et al., 2019). However, the potential sinks of the reducing equivalents of Helarchaeota in natural, non-heated sediments still remain elusive.

Scientific drilling off the Costa Rica Margin during International Ocean Drilling Program (IODP) Expedition 334 provided another excellent avenue to explore the diversity of archaea, because sediments from this location were shown to harbor not only extraordinarily high abundances (Martino et al., 2019) but also phylogenetically novel lineages of archaea (Farag et al., 2020b; Farag et al., 2020a). The drilled sites are in the forearc basin of the subduction zone, where a considerable fraction of carbon exchange between Earth's surface and interior occurs (Barry et al., 2019). Abundant thermogenic alkanes (ethane, propane, and butane) were detected in the deep sediments, but not the surface sediments (Formolo et al., 2011; Expedition 334 Scientists, 2012). This prompted us to perform further metagenomic sequencing and analysis on shallow sediment samples from this expedition, to explore the mechanisms of alkane depletion in this unique setting. Here, we report three new Helarchaeota metagenome-assembled genomes (MAGs) in shallow sediments of Hole U1379B, each of which has three variants of the MCR-like complex and could be the major

alkane-oxidizing archaea. These Helarchaeota account for the majority of the *mcrA*-bearing archaea community, with the rest *mcrA*-bearing archaea affiliated to the family of ANME-1, and may engage in methane metabolism. We also recovered five novel MAGs of sulfate-reducing bacteria in the class of Desulfobacteria, some of which could be the syntrophic partners of Helarchaeota due to the abundance correlation and their matched metabolic capacities of H₂ and formate utilization. Our study expands our understanding about the diversity and metabolic functions, and interspecies dependency of Helarchaeota, which could exert a profound influence on the carbon exchange between marine sediments and the overlying seawater.

96

97 **Results and Discussion**

98 ***Geochemical context***

IODP Site U1379 is located on the upper slope of the erosive subduction zone off the Osa Peninsula, Costa Rica. The sediment pile at this site is estimated to be totally 890 meters thick, and characterized by high sedimentation rates of 1.60-10.35 cm/ky (Expedition 334 Scientists, 2012) and low porosities of ~0.6 (Expedition 334 Scientists, 2012). We focused this study on the microbial communities inhabiting 2-9 meters below seafloor (mbsf) of U1379B, including a total of 8 subsampled whole-round cores. Sediments in this depth interval are within the sulfate reduction zone with porewater sulfate concentration decreasing almost linearly with depth (Fig. 1a), while the sulfate-depletion depth was not examined in this study but was noted at ~30 mbsf (Expedition 334 Scientists, 2012). Dissolved Mn concentration in the porewater shows the same decreasing trend as sulfate (Fig. 1b), indicating that Mn²⁺ is mainly consumed rather than produced. The increasing porewater concentrations of ammonium (Fig. 1c) and alkalinity (Fig. 1d) with depth reflect the ongoing degradation of organic matter, likely with sulfate as the most prominent electron acceptor.

Microbial abundances and overall community structures

We used domain-specific primers (Zhao et al., 2019) to quantify the abundance of archaeal and bacterial 16S rRNA genes in the nine sediment horizons from 2-9 mbsf of Hole U1379B. Both archaeal and bacterial 16S rRNA gene abundances are stable throughout the examined sediments (Fig. 1e). However, archaea are about one order of magnitude lower than bacteria, resulting in the bacterial dominance in most of the studied samples except the shallowest sample (~2 mbsf). A previous study also reported that sediments around ~2 mbsf harbor bacteria and archaea of equal abundances, based on the amplicon and metagenome sequencing data (Martino et al., 2019).

To recover as many archaeal genomes as possible, we selected four sediment horizons (2.45, 3.20, 7.70, and 8.23 mbsf) to subject to shotgun metagenome sequencing, due to the high proportions of archaea in these sediment horizons assessed by qPCR (Fig. 1e). We examined the microbial community structures in these four samples by 16S rRNA gene analyses based on both the un-assembled metagenome reads and the 16S rRNA gene amplicon sequencing data (Fig. 2). Based on the metagenome data, in the archaeal domain, Bathyarchaeota is the major archaeal phylum and accounts for on average 44.7% (37.7-51.2% in individual samples) of the total communities, while Asgard archaea especially Lokiarchaeota were also detected to account for on average 5.6% (2.6-8.7% in individual samples) of the total communities (Fig. 2a). In the bacterial domain, Chloroflexi is the most dominant phylum, which account for on average 19.0% of the total communities in different depths. Other major bacterial phyla include Acidobacteria, Actinobacteria, Elusimicrobia, Planctomycetes and Proteobacteria (mainly Deltaproteobacteria) (Fig. 2a). Similar results were also obtained from the 16S rRNA gene amplicon sequencing and analysis (Fig. 2b).

These results are consistent with the previous assessments of microbial community structures of marine sediments off Costa Rica (Martino et al., 2019; Farag et al., 2020b).

The presence and diversity of mcrA-bearing microbes

As common products of thermocatalytic degradation of organic matter in subduction zones of high temperature and pressure, C₂-C₅ alkanes (ethane, propane, butane, and pentane) were measured in sediments deeper than 50 mbsf (Expedition 334 Scientists, 2012) (Fig. S1). Alkanes in U1379 were generally not detected in the sediments shallower than 50 mbsf and no flux toward the overlying water column can be quantified, suggesting that most of the upward diffusing fluxes of alkanes are consumed in the shallow sediments. To examine the presence and overall diversity of *mcrA*-bearing, potentially alkane-consuming archaea in the CR sediments, we used the GraftM program (Boyd et al., 2018) to detect and classify the *mcrA* gene (encoding the methyl-CoM reductase alpha subunit, the key enzyme of methane/alkane metabolism) in the metagenome datasets. Our results revealed that *mcrA* gene was present in all four metagenome-sequenced sediment samples. Except for 3.20 mbsf, the *mcrA*-containing archaeal communities in all the other three horizons were not dominated by well-known archaea in the Euryarchaeota phylum (Fig. S2 and S3), indicating that uncharacterized archaea may be the dominant methane/alkane metabolizing microbes in the CR sediments, which supports the recently expanded phylogenetic breadth of methane/alkane-metabolizing archaea (Borrel et al., 2019; Hua et al., 2019; Wang et al., 2019).

Description of the Helarchaeota MAGs

Through metagenome assembly and binning, we obtained 12 archaeal MAGs that could be taxonomically assigned to the Asgard superphylum, three of which are affiliated with the newly proposed Helarchaeota phylum (Seitz et al., 2019), four Lokiarchaeota, and three

Heimdallarchaeota (Fig. 3a). The three Helarchaeota MAGs (CR_097, CR_143, and CR_291) recovered from CR have greatly expanded the diversity of this newly proposed phylum, doubling the number of available Helarchaeota genomes (Seitz et al., 2019; Cai et al., 2020). The genome sizes of the three CR Helarchaeota MAGs are similar to SZ_4_Bin10.384 (4.7 Mbp) recovered from coastal sediments (Cai et al., 2020), but are larger than two Helarchaeota genomes recovered from GB (4.7-5.4 Mbp versus 3.54-3.84 Mbp for GB Helarchaeota; Table 1). The CR Helarchaeota MAGs have 4,592 – 4,957 coding sequences distributed on 212 - 589 scaffolds (Table 1). Based on the calculated average amino acid identity (AAI; Table S1), the six available Helarchaeota genomes could be resolved to four families (i.e., with <65% intra-family AAIs; (Konstantinidis et al., 2017)): Hel_GB_A, CR_Bin_097, and CR_Bin_291 belong to a family, while each of the rest three genomes represents a family.

Among the three CR Helarchaeota MAGs, only CR_Bin_097 has a reconstructed partial (1191 bp) 16S rRNA gene sequence. This is notable since none of the three previously reported Helarchaeota MAGs has a 16S rRNA gene sequence (Seitz et al., 2019; Cai et al., 2020). An alignment of the 16S rRNA gene sequence of Helarchaeota CR_Bin_097 and other archaea revealed three insertions of >20 bp in this sequence, and such insertions are also present in other environmental clones of Asgard archaea. This observation suggests that insertions in 16S rRNA genes are common not only in bacteria of candidate phyla radiation (Brown et al., 2015), but also in Asgard archaea. Such insertions unfortunately prevent a perfect match between Helarchaeota CR_Bin_097 and the amplicon operational taxonomy units (OTUs), making it challenging to accurately estimate the distribution of these Helarchaeota in the shallow sediment horizons of Hole U1379. However, based on the read recruiting and genome coverage calculation, the three Helarchaeota seem to have similar distribution patterns and be highest in the depth of 7.70 mbsf sample (Fig. S4). After

manually removal of the insertions, the 16S rRNA gene of CR_Bin_097 showed nucleotide identities of >90% with diverse uncultured archaea sequences from marine anoxic sediments and formed a monophyletic clade with them in the phylogenetic tree of archaeal 16S rRNA gene (Fig. 3b), supporting that Helarchaeota is a separate phylum within the Asgard superphylum.

Alkane oxidation potential of CR Archaea

Similar to the GB Helarchaeota (Seitz et al., 2019), Helarchaeota in CR sediments have the potential of alkane oxidation. Each of the three CR Helarchaeota MAGs has three variants of methyl-CoM reductase-like enzymes (McrABG operon, or MCR) (Fig. 4), whereas the previously reported Helarchaeota MAGs (Seitz et al., 2019; Cai et al., 2020) have only 1 - 2 MCR variants. Phylogenetic analyses of the alpha subunit of methyl-CoM reductase-like enzymes (McrA) revealed that McrA sequences of the six available Helarchaeota MAGs form three clusters (Hel Clusters I, II, and III) within a monophyletic cluster together with homologs of Hadesarchaea, Bathyarchaeota, and *Ca. Methanoliparia* (Fig. 4). These lineages also form a monophyletic clade with McrA homologs of Syntrophoarchaea, *Argoarchaeum ethanivorans*, and *Ethanoperedens thermophilum* (Hahn et al., 2020), all are alkane-oxidizing Euryarchaeota confirmed by laboratory incubations (Laso-Perez et al., 2016; Chen et al., 2019; Wang et al., 2019). Because *mcr*-like genes so far have not been found in any Asgard phylum other than Helarchaeota, *mcr* genes in Helarchaeota, as recently suggested (Hua et al., 2019), may have resulted from horizontal gene transfer events from alkane-oxidizing Euryarchaeota. The three Hel clusters of McrA are divergent, with intra-cluster similarities in the range of 44%-63%. Multiple divergent *mcr*-like genes were also reported in “Ca. Syntrophoarchaeum” genomes, in which the duplicated *mcr*-like genes have evolved to use substrates other than methane, such as butane and propane (Laso-Perez et al., 2016). These CR Helarchaeota

MAGs also possess genes associated with the archaeal Wood-Ljungdahl, fatty acids beta-oxidation, and other pathways similar to those found in the characterized alkane-oxidizing genomes (Laso-Perez et al., 2016; Hahn et al., 2020) that have been recently identified (Seitz et al., 2019). Therefore, the gene duplication in Helarchaeota may lead to the neofunctionalization (Rastogi and Liberles, 2005) of the MCR-like complex, and enable them to metabolize short chain alkanes of different lengths anaerobically.

Among the MCR-containing MAGs from CR sediments, in addition to the three Helarchaeota MAGs, we also recovered an MAG (CR_Bin_179) affiliated with ANME-1. Based on the phylogenetic analysis of the 15 concatenated ribosomal proteins (Hug et al., 2016), this MAG should represent a new genus within the family of ANME-1 (Fig. S5A), which is also supported by the classification of GTDB-tk (Chaumeil et al., 2020) using the 122 single-copy genes of archaea, and the calculated 69%-76% AAI between this MAG and the others in the family of ANME-1 (Supplementary Table S2). CR_ANME1_Bin_179 has a near full-length 16S rRNA gene sequence (1 487 bp), which shows an identity of 94.9% with that of ANME1_CONS3730B06UFb1 (a MAG recovered from Hydrate Ridge methane seep sediments (Skenner et al., 2017)) and even lower identities with the other genomes in the family of ANME-1, supporting that this genome should represent a new genus (Konstantinidis et al., 2017). Its genomic close relatives include ANME1_CONS3730B06UFb1 recovered from Hydrate Ridge methane seep sediments, M5.MMPM from Aarhus Bay sediments, and ANME-1-THS recovered from Tibetan hot spring sediments, for all of which methane metabolism pathways have been proposed (Skenner et al., 2017; Beulig et al., 2019; Borrel et al., 2019). CR_ANME1_Bin_179 may be involved in methane metabolisms, which was supported by the partial MCR operon (constituted by McrB and McrG but the McrA is unfortunately missing) detected in this genome.

To assess the representability of these McrA-bearing genomes in the CR sediments, we performed a phylogenetic analysis of McrA amino acid sequences in the bulk metagenome assemblies of the four CR sediment layers. Our results indicate that all McrA sequences from CR metagenomes were affiliated with either the three Helarchaeota clusters or the ANME-1 cluster (Fig. S3). Because CR_ ANME1_Bin_179 may only be involved in methane but not alkane metabolism, CR Helarchaeota could be responsible for most of the presumed alkane depletion in the surface sediments at U1379.

Alkane-oxidizing archaea need to employ various means to channel the reducing equivalents generated during the oxidation of alkanes to either another internal metabolic pathway (Laso-Pérez et al., 2019) or to the external partner cells (Laso-Perez et al., 2016; Chen et al., 2019; Wang et al., 2019). Helarchaeota MAGs recovered from CR likely have to depend on bacterial partners as external electron sinks because they do not have (1) the internal pathways for the reduction of sulfate or nitrate/nitrite, (2) multi-heme cytochrome (Table 1), or (3) Type IV electric pilins for extracellular electron transfer to solid electron acceptors (e.g. metal oxides). Instead, these CR Helarchaeota MAGs contain Group 3 [NiFe] hydrogenases and formate dehydrogenases that could facilitate the transfer of the reducing equivalents in the forms of H₂ and formate. Phylogenetic analysis of the hydrogenase subunit alpha indicate that all Helarchaeot MAGs have one Group 3b [NiFe] hydrogenase and one Group 3c [NiFe] hydrogenase (Fig. 5), except for CR_Bin_291 in which the missing Group 3b hydrogenase was probably unrecovered. Helarchaeota Group 3b hydrogenases formed a monophyletic cluster with homologs of diverse Lokiarchaeota genomes (Spang et al., 2019), but were divergent from other Group 3b lineages (Fig. 5). Similarly, Helarchaeota Group 3c hydrogenases also formed a monophyletic cluster with homologs of diverse Lokiarchaeota and Thorarchaeota genomes (Spang et al., 2019) (Fig. 5). Interestingly, the only hydrogenase in the closed genome of the cultured *Ca. Prometheoarchaeum syntrophicum* MK-D1 (Imachi

et al., 2020) is also affiliated with this cluster. *Ca. Prometheoarchaeum syntrophicum* was revealed to depend on partnering cells to consume the reduced equivalents (in the forms of H₂ (Spang et al., 2019; Imachi et al., 2020). The congruence between the phylogenies of Asgard Group 3 hydrogenases and their genomes suggested that these hydrogenases may be vertically inherited from their ancestor rather than horizontal transferred from other organisms. Even though these Asgard-specific Group 3 [NiFe] hydrogenases are probably not membrane-bound, they could still play an important role in hydrogen production and transfer between organisms, given that the genome of the first cultured Lokiarchaeota has a single Group 3c [NiFe] hydrogenase (Fig. 5).

Co-occurrence of sulfate-reducing bacteria and Helarchaeota

The potential partner bacteria of Helarchaeota have not been identified. From the same metagenome sequencing dataset, we recovered five bacterial MAGs affiliated to the phylum of Desulfobacterota (Fig. 6a and Fig. 7). Four of them have the complete dissimilatory sulfate reduction pathway (constituted by sulfate adenylyltransferase (sat), adenylyl-sulfate reductase (AprAB), and dissimilatory sulfite reductase (DsrAB)), while the incompleteness of this pathway in CR_bin_60 could be due to the low genome completion (71% based on bacterial single-copy genes, Table 1). The presence of SRB genomes in these sediments is not surprising, considering that the examined sediments were from the sulfate reduction zone (Fig. 1). Phylogenomic analysis based on the 15 concatenated ribosomal proteins suggested that all the five SRB MAGs from Costa Rica margin sediments are affiliated to the class of Desulfobacteria in the phylum Desulfobacterota, according to the genome-based taxonomy classification frame (Parks et al., 2018), and are different from *Ca. Desulfofervidus auxilii* (Krukenberg et al., 2016), the partner bacterium of thermophilic alkane-oxidizing archaea, which is affiliated to the class of Desulfofervidia (Fig. 6a). CR_bin_073, CR_bin_276, and

CR_bin_277 are members of the order of “C00003060” (Fig. 6a), which corresponds to the lineage of SEEP-SRB1c (Skenner et al., 2017). The close relative genomes in the order of C00003060 are exclusively recovered from hydrocarbon-rich marine sediments, such as cold seep sites on Hydrate Ridge off the Pacific coast (Skenner et al., 2017), and B60_G16 recovered from GB hydrothermal sediments (Dombrowski et al., 2018). The other two MAGs, CR_bin_129 and CR_bin_060, formed a new lineage within the order of Desulfatiglandales, together with MAGs recovered from hydrocarbon-rich sediments in GB (Dombrowski et al., 2018) and Gulf of Mexico (Dong et al., 2019) (Fig. 6a).

Among the five SRB MAGs, only CR_Bin_276 has an identifiable partial (743 bp) 16S rRNA gene. It shows >99% identity (with two mismatches) with OTU_8 of the amplicon sequencing data. Due to the difficulties to link the Helarchaeota and SRB MAGs with short amplicon OTUs (i.e. Helarchaeota_097 does not match with any OTUs, while only one SRB has a 16S rRNA gene and matches with OTU_8), we use genome coverage as a proxy to represent the relative abundance of MAGs in each of the four metagenome samples. We found that all CR Helarchaeota MAG coverages show positive linear correlations with two of the five SRB MAGs (CR_bin_277 and CR_bin_277) (Fig. 6b-6d), representing the both lineages of SRB recovered from CR. CR_bin_277 and CR_bin_277 co-occur with Helarchaeota in CR sediments.

In addition to the abundance correlations, metabolic pathway analysis of the MAGs also support that the co-occurring Helarchaeota and SRB bacteria in CR sediments may engage in synergy metabolism, in which SRB are capable of utilizing the products of the alkane oxidation activity of Helarchaeota, H₂ and formate (Fig. 7). Three of the five SRB MAGs from CR contain a periplasmic hydrogenase affiliated to a novel lineage (Group 3e) in the Group 3 [NiFe] hydrogenases (Fig. 5), which was not well resolved in previous phylogenetic analysis (Greening et al., 2016) probably due to the small number sequences

then available. This hydrogenase may enable these SRBs to conserve energy from the oxidation of H₂, similar as *D. auxilii*, which contains a periplasmic hydrogenase and is capable of growing using H₂ in the absence of an ANME partner (Krukenberg et al., 2016). In addition, all the five SRB MAGs have two copies of respiratory formate dehydrogenases, indicating that they also have the capacity of oxidizing formate that potential released by the Helarchaeota as an form of reducing equivalents (Fig. 7). These SBR genomes from CR sediments also the full gene set for the synthesis of flagella, which have previously been reported to be an important feature of syntrophy establishment (Shimoyama et al., 2009) and could be important in initiating the contact between SRB and Helarchaeota.

Sulfate-dependent alkane oxidation is thermodynamically favorable

To explore if the oxidation of short-chain alkanes (ethane, propane, and butane) are thermodynamically favorable when coupled to sulfate reduction, we calculated the Gibbs free energy of these reactions (Supplementary Table S3) under the near *in situ* conditions in the shallow sediments at CR. Our results showed that the Gibbs free energy of the sulfate-dependent alkane oxidation (in the unit of kJ m⁻¹ alkane) is feasible over a wide range of alkane concentrations (10⁻⁶ – 1 μM), and is highest for butane, followed by propane and ethane (Fig. S6). Importantly, these short-chain alkanes can provide more energy than methane per mole of oxidized alkane (Fig. S5), suggesting that these alkanes are more energy-rich than methane and could be more easily consumed by microbes such as Helarchaeota and SRB detected in CR.

Although more data are required to properly assess the ecological roles of Helarchaeota in the environment, all known Helarchaeota so far are found in hydrocarbon-rich marine sediments (Seitz et al., 2019) and seem to be intrinsically related to the oxidation of alkanes. If they have a sulfate-reducing bacterial partner, their syntrophic metabolic

reactions, the sulfate-dependent alkane oxidations, are also high energy-yielding processes. Analogous to the role anaerobic methane-oxidizing archaea have on reducing methane emission from marine sediments (Valentine and Reeburgh, 2000), the activities of Helarchaeota and SRB in marine sediments may convert alkane to CO₂ at the expense of sulfate reduction and, therefore, reduce the emission of alkanes into the overlying water column (Fig. S1).

Conclusion

This study revealed that the majority of Asgard archaeal cells in sulfate-reducing organic-rich shallow sediments co-occur with SRB in the subduction zone off Costa Rica. Among these are three new Helarchaeota identified as MAGs, each of which have three variants of the MCR operon, and may anaerobically oxidize the steadily available alkane that is thermogenically produced from organic matter degradation in deep sediments. Like other characterized thermophilic alkane-oxidizing archaea, these Helarchaeota may engage in a syntrophic relationship with the co-occurring sulfate-reducing bacteria from the class of Desulfobacteria. These findings expand the diversity of Helarchaeota, and suggest that geological products such as hydrocarbons released from thermogenic processes in subduction zones exemplified by the Costa Rica Margin could play an important role in fueling the catabolism of microbial life in the marine deep biosphere. Experimental approaches, such as laboratory incubation complemented with cell visualization, further microscopic observations and transcriptome sequencing (Laso-Perez et al., 2016) are required to confirm these proposed syntrophic interactions.

Materials and Methods

Study sites and sample collection

Sediments used in this study were collected from Hole U1379B at the Costa Rica margin during the IODP Expedition 334. This site is located on the Caribbean plate shelf into the upper slope, with the water depth of this site is 127 m. Sediments in this Hole was drilled using the advanced piston corer system with the core recovery of 100%. Detail of drilling at this site was provided in (Vannucchi et al., 2013). Thorough descriptions of the sediment properties were published elsewhere (Expedition 334 Scientists, 2012; Martino et al., 2019). Whole-round cores of 10-cm long for microbiology analyses were capped upon drilling using blue plastic caps at both ends, and stored in anoxic plastic bags. Microbiology samples were stored at the Gulf Coast Repository of IODP at -80 °C and then were shipped to the University of Delaware and stored in -80°C freezer until further analyses.

We in this study focused on the shallow silty clay sediments of 2-9 mbsf from Hole U1379B, which is 20 m south to Hole U1379C where the most comprehensive geochemical profiles at this station were measured. The geochemical profiles of sediment in this interval of U1379C were reported in previous publications: porewater concentrations of sulfate, ammonium, and alkalinity in Expedition 334 Scientists (2012), and dissolved manganese in (Torres et al., 2014). Procedures for hydrocarbon composition determined by headspace analysis were described in (Riedinger et al., 2019), and the data can be found in the IODP Expedition 334 report (<http://publications.iodp.org/proceedings/334/334toc.htm>).

DNA extraction, quantitative PCR, and amplicon analysis

DNA for amplicon sequencing was extracted from 0.5 g of sediment using the PowerSoil DNA extraction kit (MOBIO, CA). A parallel extraction blank without adding sample material in the beginning was also perform to track the potential contamination introduced during the experimental process. DNA for metagenomic sequencing was extracted from about 5 g of sediment (0.5 g was added into each of the 10 lysis tubes) using the PowerSoil DNA

extraction kit (MOBIO, CA). For each sample, DNA extracts from the 10 parallel extractions were eluted with 100 μ L double-distilled H₂O into a single 1.5-mL Eppendorf tube and preserved at -20°C for further analyses.

The bacterial and archaeal 16S rRNA genes were quantified using the primer sets Uni341/Uni519 and Uni515F/Arc908r, respectively, combining with the thermal conditions described in (Zhao et al., 2019). For the amplicon preparation, a two-step strategy was employed to prepare the amplicons of the V4 region of the 16S rRNA gene with the primer pair Uni519F/806R and thermal cycling condition described elsewhere (Zhao et al., 2019). Details about the qPCR standard preparation, qPCR reaction chemistry, amplicon sequencing data analysis were presented in the Supplementary Information.

Metagenome assembly, binning, and genome refinement

Metagenomic libraries (2x150 paired-end) were prepared and sequenced on an Illumina NextSeq 500 sequencing platform (Illumina Inc., San Diego, CA, USA). Quality of the reads and presence of adaptor sequences were checked using FastQC v.0.11.5 (Andrews, 2010) and then trimmed using Trimmomatic v.0.36 (Bolger et al., 2014). Putative 16S rRNA genes in the trimmed metagenome reads were used to assess the microbial community structure using phyloFlash v3.3 beta 2 (Gruber-Vodicka et al., 2020), in which the short reads were identified and classified against the SILVA 132 release.

The quality-controlled paired-end reads were *de novo* assembled into contigs using Megahit v.1.1.2 (Li et al., 2015) with the k-mer length varying from 27 to 117. Contigs longer than 1000 bp were automatically binned with MaxBin2 v2.2.6 (Wu et al., 2016) using the default parameters. The quality of the obtained genome bins was assessed using CheckM v.1.0.7 (Parks et al., 2015) with the option “lineage_wf”, which uses lineage-specific sets of single-copy genes to estimate completeness and contamination and assigns contamination to

strain heterogeneity if amino acid identity is >90%. Genome bins of >50% completeness were manually refined using the *R* package *gbtools* (Seah and Gruber-Vodicka, 2015), based on the GC content, taxonomic assignments, and differential coverages in different samples. To improve the quality of MAGs, metagenome reads of the sample with the highest coverage was detected were mapped onto the MAG contigs using BBmap (Bushnell, 2014), and the aligned reads were re-assembled using SPAdes v.3.12.0 (Bankevich et al., 2012) with the default parameters and minimum contig length of 1000 bp. The resulting scaffolds were visualized and re-binned using *gbtools* (Seah and Gruber-Vodicka, 2015) as described above. The qualities of the resulting MAGs were checked using the CheckM v.1.0.7 “lineage_wf” command again. Details were presented in the Supplementary Information.

Overall diversity of *mcrA*-bearing archaea

Reads of *mcrA* in the unassembled metagenome sequencing data were identified and classified using GraftM v0.13.1 (Boyd et al., 2018), with the pre-curated *mcrA* package (including HMM profiles and pre-constructed phylogenetic tree) described in (Singleton et al., 2018) as the reference database (downloaded from <https://data.ace.uq.edu.au/public/grafm/7/>).

Genome annotation

Genome annotation was conducted using Prokka v.1.13 (Seemann, 2014), eggNOG (Huerta-Cepas et al., 2016), and BlastKoala (Kanehisa et al., 2016) using the KEGG database. The functional assignments of genes of interest were also confirmed using BLASTp against the NCBI RefSeq database. The metabolic pathways were reconstructed using KEGG Mapper (Kanehisa et al., 2011).

For the multi-heme cytochromes (MHCs) detection, the heme-binding sites of a protein, the CXXCH motif, were counted using the python script “cytochrome_stats.py”

described in (Badalamenti et al., 2016) (available at <https://github.com/bondlab/scripts>) with the amino acid sequences predicted with Prodigal (Hyatt et al., 2010) as the input. Proteins with >3 CXXCH motifs were counted as MHCs following the criteria described elsewhere (Badalamenti et al., 2016; Hernsdorf et al., 2017).

Average nucleotide identify (ANI) between different genomes were calculated using FastANI with default parameters (Jain et al., 2018). Average amino acid identify (AAI) were calculated using CompareM (<https://github.com/dparks1134/CompareM>) with the “aai_wf” option, in which the protein coding sequences (CDS) predicted by Prodigal (Hyatt et al., 2010) were taken as the input to identify orthologous proteins by all-vs-all reciprocal sequence similarity search with Diamond (Buchfink et al., 2015). The average similarity of the orthologous proteins between the two genomes were taken as the pairwise AAI.

Phylogenetic analysis

For the phylogenomic analysis of Helarchaeota, representative genomes of all major archaeal lineages described in the GTDB database (Parks et al., 2018) were downloaded and included. For sulfate-reducing bacteria, representative genomes of the candidate bacteria phyla of Desulfobacterota, Desulfobacterota_A, and Desulfuromonadota (phyla per nomenclatures of GTDB; <https://gtdb.ecogenomic.org/>) were included. The phylogenomic analyses were based on markers consisting of 15 concatenated ribosomal proteins (rpL2, 3, 4, 5, 6, 14, 16, 18, 22, 24 and rpS3, 8, 10, 17, 19) that have been demonstrated to undergo limited lateral gene transfer (Sorek et al., 2007). These selected proteins, among the conservative single-copy ribosomal proteins included in (Campbell et al., 2013), were identified in Anvi'o v.5.5 (Eren et al., 2015) using Hidden Markov Model (HMM) profiles, following the procedure outlined at <http://merenlab.org/2017/06/07/phylogenomics/>. Sequences were aligned individually using MUSCLE (Edgar, 2004), and alignment gaps were removed using trimAl (Capella-Gutierrez

et al., 2009) with the mode of “automated”. Individual alignments of ribosomal proteins were concatenated. The maximum-likelihood phylogenetic tree was reconstructed based on the concatenated alignment using IQ-TREE v1.6.10 (Nguyen et al., 2015) with LG+F+R6 as the best-fit evolutionary model selected by ModelFinder (Kalyaanamoorthy et al., 2017) and 1000 ultrafast bootstraps using UFBoot2 (Hoang et al., 2018).

A maximum likelihood phylogenetic tree based on 16S rRNA genes was also constructed based on an alignment of 16S rRNA gene sequences of the genomes that included in the above-mentioned phylogenomic analysis. In addition, amplicon OTUs (287 bp) of Asgard archaea and their close relatives (environmental clones of >1300 bp) identified via BLASTn (Altschul et al., 1997) in the NCBI database. Sequences were aligned using MAFFT-LINSi (Katoh and Standley, 2013), and putative insertions were manually trimmed in Unipro UGENE (Okonechnikov et al., 2012). The maximum-likelihood phylogenetic tree was inferred based on the insertion-free alignment using IQ-TREE v1.6.10 (Nguyen et al., 2015) with SYM+R5 as the best-fit evolutionary model determined by ModelFinder (Kalyaanamoorthy et al., 2017) and 1,000 ultrafast bootstrap replicates using UFBoot2 (Hoang et al., 2018).

For the phylogeny of McrA, in addition of Helarchaeota, the genomes of known MCR-bearing archaea genomes were downloaded from NCBI, annotated using Prokka v1.13 (Seemann, 2014), and the McrA amino acid sequences were extracted. Phylogenetic analysis was also performed for the McrA sequences in the bulk assemblies of the four CR sediment horizons, in which the McrA sequences were extracted from the Prokka annotation outputs. In both analyses, all sequences were aligned using MAFFT-LINSi (Katoh and Standley, 2013), trimmed using trimAl (Capella-Gutierrez et al., 2009) with the mode of “automated”, and the maximum likelihood phylogenetic tree was inferred using IQ-TREE v1.6.10 (Nguyen

et al., 2015) with LG+F+R6 as the best-fit evolutionary model and 1,000 fast bootstrap replicates.

For the phylogeny of [NiFe] hydrogenases, reference sequences were mainly extracted from (Carnevali et al., 2019) and (Kessler et al., 2019). [NiFe] hydrogenases of the genomes reported in this study were extracted from the Prokka annotations, and used as queries in BLASTp search (Altschul et al., 1997) in the NCBI database to identify their close relatives. All retrieved sequences were aligned using MAFFT-LINSi (Katoh and Standley, 2013), trimmed using trimAl (Capella-Gutierrez et al., 2009) with the mode of “automated”, and the phylogenetic tree was reconstructed using IQ-TREE v1.6.10 (Nguyen et al., 2015), with the procedure described above.

Genome coverage calculation and the linear correlation

As an proxy of the relative abundance, genome coverages of MAGs in each of the four metagenome-sequenced sediment depths were determined by recruiting reads from the individual metagenome datasets using BBmap (Bushnell, 2014) with the read identity threshold of 98%. Relative abundances of the Helarchaeota MAGs in the total communities were also calculated as the product of the average coverage and genome size, divided by the total trimmed reads. Linear correlations between the five Desulfobacteria MAGs and the three Helarchaeota MAGs were determined in R (R Development Core Team, 2011), using Pearson correlation test.

Thermodynamic calculation

We assessed the feasibility of alkane oxidation processes by calculating the Gibbs free energy for reactions of sulfate-dependent oxidation of ethane, propane, butane as well as methane in

the shallow sediments (<10 mbsf) (See Table S3 for the chemical equations). Details were presented in the Supplementary Information.

Data availability

All sequencing data used in this study are available in NCBI Short Reads Archive under the project number PRJNA599172. In particular, the amplicon sequencing data can be accessed through the BioSample number SAMN13740702 - SAMN13740710. The raw metagenomic sequencing data are available in NCBI under the BioSample numbers SAMN13740741-SAMN13740744. The MAGs described in this study are available in NCBI with the accession numbers JABXJT0000000000 - JABXKA0000000000.

Acknowledgements

This study uses samples and data provided by the IODP. We are grateful to all scientists and crew members of IODP Expedition 334, whose efforts made the samples available. We acknowledge computational support from the University of Delaware Center for Bioinformatics and Computational Biology Core Facility and the use of the BIOMIX computing cluster was made possible through funding from Delaware INBRE (NIGMS P20GM103446), the State of Delaware, and the Delaware Biotechnology Institute. This work was funded by the WM Keck Foundation (to J.F.B.).

Competing Interests

The authors declare no competing interests.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

535 Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data.
536 <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

537 Badalamenti, J.P., Summers, Z.M., Chan, C.H., Gralnick, J.A., and Bond, D.R. (2016)
538 Isolation and genomic characterization of ‘Desulfuromonas soudanensis WTL’, a metal-and
539 electrode-respiring bacterium from anoxic deep subsurface brine. *Frontiers in microbiology* **7**:
540 913.

541 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. et al.
542 (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell
543 sequencing. *Journal of Computational Biology* **19**: 455-477.

544 Barry, P.H., de Moor, J.M., Giovannelli, D., Schrenk, M., Hummer, D.R., Lopez, T. et al.
545 (2019) Forearc carbon sink reduces long-term volatile recycling into the mantle. *Nature* **568**:
546 487-492.

547 Beulig, F., Røy, H., McGlynn, S.E., and Jørgensen, B.B. (2019) Cryptic CH₄ cycling in the
548 sulfate–methane transition of marine sediments apparently mediated by ANME-1 archaea.
549 *The ISME Journal* **13**: 250-262.

550 Biddle, J.F., Lipp, J.S., Lever, M.A., Lloyd, K.G., Sorensen, K.B., Anderson, R. et al. (2006)
551 Heterotrophic archaea dominate sedimentary subsurface ecosystems off Peru. *Proceedings of*
552 *the National Academy of Sciences of the United States of America* **103**: 3846-3851.

553 Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina
554 sequence data. *Bioinformatics* **30**: 2114-2120.

555 Borrel, G., Adam, P.S., McKay, L.J., Chen, L.-X., Sierra-García, I.N., Sieber, C.M. et al.
556 (2019) Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea.
557 *Nature Microbiology* **4**: 603.

558 Boyd, J.A., Woodcroft, B.J., and Tyson, G.W. (2018) GraftM: a tool for scalable,
559 phylogenetically informed classification of genes within metagenomes. *Nucleic Acids*
560 *Research* **46**: e59.

561 Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A. et al. (2015)
562 Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**:
563 208-211.

564 Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using
565 DIAMOND. *Nature Methods* **12**: 59-60.

566 Bushnell, B. (2014) BBMap: a fast, accurate, splice-aware aligner. In: Ernest Orlando
567 Lawrence Berkeley National Laboratory, Berkeley, CA (US).

568 Cai, M., Liu, Y., Yin, X., Zhou, Z., Friedrich, M.W., Richter-Heitmann, T. et al. (2020)
569 Diverse Asgard archaea including the novel phylum Gerdarchaeota participate in organic
570 matter degradation. *Science China Life Sciences*.

571 Campbell, J.H., O'Donoghue, P., Campbell, A.G., Schwientek, P., Sczyrba, A., Woyke, T. et
572 al. (2013) UGA is an additional glycine codon in uncultured SR1 bacteria from the human

573 microbiota. *Proceedings of the National Academy of Sciences of the United States of America*
574 **110**: 5540-5545.

575 Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009) trimAl: a tool for
576 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-
577 1973.

578 Carnevali, P.B.M., Schulz, F., Castelle, C.J., Kantor, R.S., Shih, P.M., Sharon, I. et al. (2019)
579 Hydrogen-based metabolism as an ancestral trait in lineages sibling to the Cyanobacteria.
580 *Nature Communications* **10**: 463.

581 Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2020) GTDB-Tk: a toolkit to
582 classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**: 1925-1927.

583 Chen, S.-C., Musat, N., Lechtenfeld, O.J., Paschke, H., Schmidt, M., Said, N. et al. (2019)
584 Anaerobic oxidation of ethane by archaea from a marine hydrocarbon seep. *Nature* **568**: 108-
585 111.

586 Dombrowski, N., Teske, A.P., and Baker, B.J. (2018) Expansive microbial metabolic
587 versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nature*
588 *Communications* **9**: 4999.

589 Dong, X., Greening, C., Rattray, J.E., Chakraborty, A., Chuvochina, M., Mayumi, D. et al.
590 (2019) Metabolic potential of uncultured bacteria and archaea associated with petroleum
591 seepage in deep-sea sediments. *Nature communications* **10**: 1816.

592 Durbin, A.M., and Teske, A. (2012) Archaea in organic-lean and organic-rich marine
593 subsurface sediments: an environmental gradient reflected in distinct phylogenetic lineages.
594 *Frontiers in Microbiology* **3**: 168.

595 Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high
596 throughput. *Nucleic Acids Research* **32**: 1792-1797.

597 Eren, A.M., Esen, O.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont,
598 T.O. (2015) Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**:
599 e1319.

600 Expedition 334 Scientists (2012) Site U1379. In Vannucchi, P., Ujiie, K., Stroncik, N,
601 Malinverno, A, and the Expedition 334 Scientists, *Proc IODP, 334: Tokyo (Integrated Ocean*
602 *Drilling Program Management International, Inc)*

603 Farag, I.F., Zhao, R., and Biddle, J.F. (2020a) "Sifarchaeota" a novel Asgard phylum capable
604 of polysaccharide degradation and anaerobic methylotrophy. *bioRxiv*:
605 2020.2010.2014.339440.

606 Farag, I.F., Biddle, J.F., Zhao, R., Martino, A.J., House, C.H., and León-Zayas, R.I. (2020b)
607 Metabolic potentials of archaeal lineages resolved from metagenomes of deep Costa Rica
608 sediments. *ISME Journal* **14**: 1345-1358.

609 Formolo, M., Nuzzo, M., Torres, M., Solomon, E., and Expedition, I. (2011) Gas geochemical
610 results from IODP Expedition 334: Influence of subsurface structure and fluid flow on gas
611 composition. In *AGU Fall Meeting Abstracts*.

612 Greening, C., Biswas, A., Carere, C.R., Jackson, C.J., Taylor, M.C., Stott, M.B. et al. (2016)
613 Genomic and metagenomic surveys of hydrogenase distribution indicate H₂ is a widely
614 utilised energy source for microbial growth and survival. *ISME Journal* **10**: 761-777.

615 Gruber-Vodicka, H.R., Seah, B.K.B., and Pruesse, E. (2020) phyloFlash: rapid small-subunit
616 rRNA profiling and targeted assembly from metagenomes. *mSystems* **5**: e00920-00920.

617 Hahn, C.J., Laso-Pérez, R., Vulcano, F., Vaziourakis, K.-M., Stokke, R., Steen, I.H. et al.
618 (2020) “Candidatus Ethanoperedens,” a thermophilic genus of Archaea mediating the
619 anaerobic oxidation of ethane. *mBio* **11**: e00600-00620.

620 Hernsdorf, A.W., Amano, Y., Miyakawa, K., Ise, K., Suzuki, Y., Anantharaman, K. et al.
621 (2017) Potential for microbial H₂ and metal transformations associated with novel bacteria
622 and archaea in deep terrestrial subsurface sediments. *ISME Journal* **11**: 1915-1929.

623 Hiraoka, S., Hirai, M., Matsui, Y., Makabe, A., Minegishi, H., Tsuda, M. et al. (2020)
624 Microbial community and geochemical analyses of trans-trench sediments for understanding
625 the roles of hadal environments. *ISME Journal* **14**: 740-756.

626 Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018) UFBoot2:
627 Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**: 518-
628 522.

629 Hoshino, T., and Inagaki, F. (2019) Abundance and distribution of Archaea in the subseafloor
630 sedimentary biosphere. *ISME Journal* **13**: 227-231.

631 Hoshino, T., Doi, H., Uramoto, G.-I., Wörmer, L., Adhikari, R.R., Xiao, N. et al. (2020)
632 Global diversity of microbial communities in marine sediment. *Proceedings of the National*
633 *Academy of Sciences* **117**: 27587-27597.

634 Hua, Z.-S., Wang, Y.-L., Evans, P.N., Qu, Y.-N., Goh, K.M., Rao, Y.-Z. et al. (2019) Insights
635 into the ecological roles and evolution of methyl-coenzyme M reductase-containing hot spring
636 Archaea. *Nature Communications* **10**: 1-11.

637 Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C. et al. (2016)
638 eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for
639 eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research* **44**: D286-D293.

640 Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J. et al.
641 (2016) A new view of the tree of life. *Nature Microbiology* **1**: 16048.

642 Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010)
643 Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC*
644 *Bioinformatics* **11**: 119.

645 Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y. et al. (2020)
646 Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**: 519-525.

647 Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018) High
648 throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.
649 *Nature Communications* **9**: 5114.

650 Kallmeyer, J., Pockalny, R., Adhikari, R.R., Smith, D.C., and D'Hondt, S. (2012) Global
651 distribution of microbial abundance and biomass in subseafloor sediment. *Proceedings of the*
652 *National Academy of Sciences of the United States of America* **109**: 16213-16216.

653 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermin, L.S. (2017)
654 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**:
655 587-589.

656 Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG
657 Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of*
658 *Molecular Biology* **428**: 726-731.

659 Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2011) KEGG for
660 integration and interpretation of large-scale molecular data sets. *Nucleic acids research* **40**:
661 D109-D114.

662 Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version
663 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772-780.

664 Kessler, A.J., Chen, Y.-J., Waite, D.W., Hutchinson, T., Koh, S., Popa, M.E. et al. (2019)
665 Bacterial fermentation and respiration processes are uncoupled in anoxic permeable sediments.
666 *Nature Microbiology* **4**: 1014.

667 Konstantinidis, K.T., Rosselló-Móra, R., and Amann, R. (2017) Uncultivated microbes in
668 need of their own taxonomy. *ISME Journal* **11**: 2399-2406.

669 Krukenberg, V., Harding, K., Richter, M., Glöckner, F.O., Gruber-Vodicka, H.R., Adam, B.
670 et al. (2016) *Candidatus Desulfoferriplus auxilii*, a hydrogenotrophic sulfate-reducing
671 bacterium involved in the thermophilic anaerobic oxidation of methane. *Environmental*
672 *microbiology* **18**: 3073-3091.

673 Laso-Pérez, R., Wegener, G., Knittel, K., Widdel, F., Harding, K.J., Krukenberg, V. et al.
674 (2016) Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* **539**:
675 396-401.

676 Laso-Pérez, R., Hahn, C., van Vliet, D.M., Tegetmeyer, H.E., Schubotz, F., Smit, N.T. et al.
677 (2019) Anaerobic degradation of non-methane alkanes by “*Candidatus Methanoliparia*” in
678 hydrocarbon seeps of the Gulf of Mexico. *mBio* **10**: e01814-01819.

679 Li, D.H., Liu, C.M., Luo, R.B., Sadakane, K., and Lam, T.W. (2015) MEGAHIT: an ultra-fast
680 single-node solution for large and complex metagenomics assembly via succinct de Bruijn
681 graph. *Bioinformatics* **31**: 1674-1676.

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

684 Lloyd, K.G., May, M.K., Kevorkian, R.T., and Steen, A.D. (2013a) Meta-analysis of
685 quantification methods shows that Archaea and Bacteria have similar abundances in the
686 subseafloor. *Applied and Environmental Microbiology* **79**: 7790-7799.

687 Lloyd, K.G., Schreiber, L., Petersen, D.G., Kjeldsen, K.U., Lever, M.A., Steen, A.D. et al.
688 (2013b) Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**: 215-
689 218.

690 Manoharan, L., Kozłowski, J.A., Murdoch, R.W., Löffler, F.E., Sousa, F.L., and Schleper, C.
691 (2019) Metagenomes from coastal marine sediments give insights into the ecological role and
692 cellular features of Loki-and Thorarchaeota. *mBio* **10**: e02039-02019.

693 Martino, A., Rhodes, M.E., León-Zayas, R., Valente, I.E., Biddle, J.F., and House, C.H. (2019)
694 Microbial diversity in sub-seafloor sediments from the Costa Rica Margin. *Geosciences* **9**:
695 218.

696 Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015) IQ-TREE: A fast and
697 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular*
698 *Biology and Evolution* **32**: 268-274.

699 Okonechnikov, K., Golosova, O., Fursov, M., and Team, U. (2012) Unipro UGENE: a unified
700 bioinformatics toolkit. *Bioinformatics* **28**: 1166-1167.

701 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015)
702 CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and
703 metagenomes. *Genome Research* **25**: 1043-1055.

704 Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., and
705 Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny
706 substantially revises the tree of life. *Nature Biotechnology* **36**: 996-1004.

707 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA
708 ribosomal RNA gene database project: improved data processing and web-based tools.
709 *Nucleic Acids Research* **41**: D590-D596.

710 R Development Core Team (2011) R: A language and environment for statistical computing.
711 In: R foundation for statistical computing, Vienna, Austria.

712 Rastogi, S., and Liberles, D.A. (2005) Subfunctionalization of duplicated genes as a transition
713 state to neofunctionalization. *BMC Evolutionary Biology* **5**: 28.

714 Riedinger, N., Torres, M.E., Sreaton, E., Solomon, E.A., Kutterolf, S., Schindlbeck-Belo, J.
715 et al. (2019) Interplay of subduction tectonics, sedimentation, and carbon cycling.
716 *Geochemistry, Geophysics, Geosystems* **20**: 4939-4955.

717 Seah, B.K., and Gruber-Vodicka, H.R. (2015) gbtools: interactive visualization of
718 metagenome bins in R. *Frontiers in Microbiology* **6**: 1451.

719 Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-
720 2069.

721 Seitz, K.W., Dombrowski, N., Eme, L., Spang, A., Lombard, J., Sieber, J.R. et al. (2019)
722 Asgard archaea capable of anaerobic hydrocarbon cycling. *Nature Communications* **10**: 1822.

723 Shimoyama, T., Kato, S., Ishii, S.i., and Watanabe, K. (2009) Flagellum mediates symbiosis.
724 *Science* **323**: 1574-1574.

725 Singleton, C.M., McCalley, C.K., Woodcroft, B.J., Boyd, J.A., Evans, P.N., Hodgkins, S.B. et
726 al. (2018) Methanotrophy across a natural permafrost thaw environment. *ISME Journal* **12**:
727 2544-2558.

728 Skennerton, C.T., Chourey, K., Iyer, R., Hettich, R.L., Tyson, G.W., and Orphan, V.J. (2017)
729 Methane-fueled syntrophy through extracellular electron transfer: uncovering the genomic
730 traits conserved within diverse bacterial partners of anaerobic methanotrophic archaea. *mBio*
731 **8**: e00530-00517.

732 Sorek, R., Zhu, Y.W., Creevey, C.J., Francino, M.P., Bork, P., and Rubin, E.M. (2007)
733 Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**:
734 1449-1452.

735 Spang, A., Caceres, E.F., and Ettema, T.J.G. (2017) Genomic exploration of the diversity,
736 ecology, and evolution of the archaeal domain of life. *Science* **357**: eaaf3883.

737 Spang, A., Stairs, C.W., Dombrowski, N., Eme, L., Lombard, J., Caceres, E.F. et al. (2019)
738 Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative
739 analyses of Asgard archaeal metabolism. *Nature Microbiology* **4**: 1138-1148.

740 Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E. et
741 al. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*
742 **521**: 173-179.

743 Torres, M.E., Muratli, J.M., and Solomon, E.A. (2014) Data report: minor element
744 concentrations in pore fluids from the CRISP-A transect drilled during Expedition 334. In
745 *Proc IODP/ Volume*, p. 2.

746 Valentine, D.L., and Reeburgh, W.S. (2000) New perspectives on anaerobic methane
747 oxidation: minireview. *Environmental Microbiology* **2**: 477-484.

748 Vannucchi, P., Ujiie, K., and Stroncik, N. (2013) IODP expedition 334: An investigation of
749 the sedimentary record, fluid flow and state of stress on top of the seismogenic zone of an
750 erosive subduction margin. *Scientific Drilling*.

751 Vuillemin, A., Wankel, S.D., Coskun, Ö.K., Magritsch, T., Vargas, S., Estes, E.R. et al. (2019)
752 Archaea dominate oxic seafloor communities over multimillion-year time scales. *Science*
753 *Advances* **5**: eaaw4108.

754 Wang, Y., Wegener, G., Hou, J., Wang, F., and Xiao, X. (2019) Expanding anaerobic alkane
755 metabolism in the domain of Archaea. *Nature Microbiology* **4**: 595-602.

756 Wu, Y.W., Simmons, B.A., and Singer, S.W. (2016) MaxBin 2.0: an automated binning
757 algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**: 605-
758 607.

759 Yu, T., Wu, W., Liang, W., Lever, M.A., Hinrichs, K.-U., and Wang, F. (2018) Growth of
760 sedimentary Bathyarchaeota on lignin as an energy source. *Proceedings of the National*
761 *Academy of Sciences* **115**: 6022-6027.

Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E. et al. (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**: 353-358.

Zhao, R., Hannisdal, B., Mogollon, J.M., and Jørgensen, S.L. (2019) Nitrifier abundance and diversity peak at deep redox transition zones. *Scientific Reports* **9**: 8633.

Figure legends

Fig. 1. Porewater profiles and microbial abundances in the sediments at IODP Site 1379 on the Costa Rica margin. Porewater profiles were compiled from previous reports/publications of IODP Expedition 334 (Expedition 334 Scientists, 2012; Torres et al., 2014). (a-d) Porewater profiles of sulfate (a), dissolved manganese (b), ammonium (c), and alkalinity (d) in the upper 10 m sediments in Hole U1379. (e) Abundances of archaeal and bacterial 16S rRNA genes quantified using qPCR with domain-specific primers. Stars denote the four sediment horizons where metagenome sequencing data were generated in this study.

Fig. 2. Microbial community structures assessed by 16S rRNA gene reads in raw metagenome sequencing data (a) and amplicon sequencing (b). For (a), unassembled reads were mapped to the SILVA 132 release and classified using phyloFlash (Gruber-Vodicka et al., 2020).

Fig. 3. Maximum-likelihood phylogenetic tree of 15 concatenated ribosomal proteins (a) and 16S rRNA gene (b) of archaea. (a) This tree was inferred using IQ-TREE v1.6.10 with the LG+R7 model and 1000 ultrafast bootstraps. The Helarchaeota MAGs recovered in this study is highlighted in red. Asgard archaea MAGs other than Helarchaeota recovered in this study are shown in orange. (b) This tree was inferred using IQ-TREE v1.6.10 with the SYM+R5 model and 1000 ultrafast bootstraps. The Helarchaeota MAG is highlighted in red, while the short OTUs (287 bp) affiliated with Asgard archaea are shown in blue. Lineages were collapsed at the phylum level if possible, except for those in the Asgard superphylum. The scale bars show estimated sequence substitutions per residue.

Fig. 4. Maximum-likelihood phylogeny of methyl-coenzyme M reductase subunit alpha (McrA) of archaea. This tree was inferred using IQ-TREE v1.6.10 with LG+F+I+G4 as the best-fit evolutionary model and 1000 ultrafast bootstraps based on an alignment of 450 amino acid positions of 167 sequences from diverse MCR-containing archaea. The McrA sequences of Helarchaeota were placed into three distinct clusters: Hel Cluster I, II, and III. The scale bar shows estimated sequence substitutions per residue.

Fig. 5 Maximum-likelihood phylogenetic tree of Group 3 [NiFe] hydrogenase alpha subunit. This tree was inferred using IQ-TREE v1.6.10 with LG+R6 as the best-fit evolutionary model and 1000 ultrafast bootstraps based on an alignment of 394 positions of protein sequences from diverse Bacteria and Archaea. Sequences of Helarchaeota recovered from CR are shown in red, while those from SRB MAGs are highlighted in blue. The tree is

rooted to Group 4 [NiFe] hydrogenase sequences. The scale bar shows estimated sequence substitutions per residue.

Fig. 6. Phylogeny of Desulfobacteria genomes and their genome coverage correlations with Helarchaeota genomes in CR sediments. (a) Maximum-likelihood phylogenetic tree of Desulfobacteria genomes recovered from CR sediments. This tree was inferred using IQ-TREE v1.6.10 with LG+F+R6 as the best-fit evolutionary model and 1000 ultrafast bootstraps, based on an alignment of the 15 concatenated ribosomal proteins sequences. This tree is rooted to genomes of the Desulfurmonadota phylum. Desulfobacteria MAGs recovered from Costa Rica margin sediments are color-coded, while the reference genomes are shown in black. Sulfate-reducing bacterial partner of thermophilic alkane-oxidizing archaea, *Candidatus Desulfoferriplus auxilli* in the class of Desulfoferriplus, is highlighted in purple. The scale bar shows estimated sequence substitutions per residue. (b-d) Correlations of the genome coverages between Helarchaeota and Desulfobacteria recovered from CR margin sediments. The blue and black dashed lines show the linear correlations of CR_Bin_277 and CR_Bin_129, respectively, with the three Helarchaeota genomes.

Fig. 7. Proposed metabolic scheme and interaction between Helarchaeota (red) and sulfate-reducing bacteria (blue) in Costa Rica Margin sediments. Missing genes/pathways are shown in dashed lines. Cyt, (multi-heme) cytochromes. Absent membrane complexes are shown in grey.

Figures and Tables

Fig. 1. Porewater profiles and microbial abundances in the sediments at IODP Site 1379 on the Costa Rica margin. Porewater profiles in meters below seafloor (meters bsf) were compiled from previous reports/publications of IODP Expedition 334 (Expedition 334 Scientists, 2012; Torres et al., 2014). (a-d) Porewater profiles of sulfate (a), dissolved manganese (b), ammonium (c), and alkalinity (d) in the upper 10 meters sediments in Hole U1379. (e) Abundances of archaeal and bacterial 16S rRNA genes quantified using qPCR with domain-specific primers. Stars denote the four sediment horizons where metagenome sequencing data was generated for this study.

Fig. 2. Microbial community structures assessed by 16S rRNA gene reads in raw metagenome sequencing data (a) and amplicon sequencing (b). For (a), unassembled reads in the metagenomes were mapped to the SILVA 132 release and classified using phyloFlash. In both (a) and (b), taxa <5% relative abundance were grouped into “Others”.

Fig. 3. Maximum-likelihood phylogenetic tree of 15 concatenated ribosomal proteins (a) and 16S rRNA gene (b) of archaea. (a) Phylogenetic tree of 15 concatenated ribosomal proteins inferred using IQ-TREE v1.6.10 with the LG+R7 model and 1000 ultrafast bootstraps. The Helarchaeota MAGs recovered in this study are highlighted in red. Asgard archaea MAGs other than Helarchaeota recovered in this study are shown in orange, phyla are labeled with abbreviated names. (b) Tree of 16S rRNA genes inferred using IQ-TREE v1.6.10 with the SYM+R5 model and 1000 ultrafast bootstraps. The Helarchaeota

MAG is highlighted in red, while the short OTUs (287 bp) affiliated with Asgard archaea are shown in blue. Lineages were collapsed at the phylum level, except for those in the Asgard superphylum. The scale bars show estimated sequence substitutions per residue.

Fig. 4. Maximum-likelihood phylogeny of methyl-coenzyme M reductase subunit alpha (McrA) of archaea. This tree was inferred using IQ-TREE v1.6.10 with LG+F+I+G4 as the best-fit evolutionary model and 1000 ultrafast bootstraps based on an alignment of 450 amino acid positions of 167 sequences from diverse MCR-containing archaea. The McrA sequences of Helarchaeota (Hel) were placed into three distinct clusters: Hel Cluster I, II, and III. The scale bar shows estimated sequence substitutions per residue.

Fig. 5 Maximum-likelihood phylogenetic tree of Group 3 [NiFe] hydrogenase alpha subunit. This tree was inferred using IQ-TREE v1.6.10 with LG+R6 as the best-fit evolutionary model and 1000 ultrafast bootstraps based on an alignment of 394 positions of protein sequences from diverse Bacteria and Archaea. Sequences of Helarchaeota recovered from CR are shown in red, while those from SRB MAGs are highlighted in blue. The tree is rooted to Group 4 [NiFe] hydrogenase sequences. The scale bar shows estimated sequence substitutions per residue.

Fig. 6. Phylogeny of Desulfobacteria genomes and their genome coverage correlations with Helarchaeota genomes in CR sediments. (a) Maximum-likelihood phylogenetic tree of Desulfobacteria genomes recovered from CR sediments. This tree was inferred using IQ-TREE v1.6.10 with LG+F+R6 as the best-fit evolutionary model and 1000 ultrafast bootstraps, based on an alignment of the concatenated 15 ribosomal proteins sequences. This tree is rooted to genomes of the Desulfurmonadota phylum. Desulfobacteria MAGs recovered from Costa Rica margin sediments are color-coded, while the reference genomes are shown in black. The sulfate-reducing bacterial partner of thermophilic alkane-oxidizing archaea, *Candidatus Desulfoferidus auxilli* in the class of Desulfoferidia, is highlighted in purple. The scale bar shows estimated sequence substitutions per residue. (b-d) Correlations of the genome coverages between Helarchaeota and Desulfobacteria recovered from CR margin sediments. The blue and black dashed lines show the linear correlations of CR_Bin_277 and CR_Bin_129, respectively, with the three CR Helarchaeota genomes.

Table 1 Summary of the Helarchaeota and Deltaproteobacteria MAGs recovered from Hole U1379B

		Helarchaeota			ANME-1 archaea		Sulfate-Reducing Bacteria			
MAG ID		Helarchaeota Bin_097	Helarchaeota Bin_291	Helarchaeota Bin_143	CR_Bin_179	CR_Bin_276	CR_Bin_073	CR_Bin_129	CR_Bin_060	CR_Bin_277
Genome size	(Mbp)	5.4	4.7	4.7	2.50	4.3	2.2	3.0	2.2	1.5
# Scaffolds		212	475	589	136	591	197	509	532	495
16S rRNA gene	1191 bp	--	--	--	1483 bp	743 bp	--	--	--	--
% GC		38.6%	36.4%	30.7%	41.9%	49.5%	48.2%	48.0%	49.2%	51.5%
Completeness (%)		91.6%	83.2%	92.2%	90.3%	92.3%	90.2%	77.6%	71.2%	64.6%
Redundancy (%)		3.3%	5.6%	4.7%	2.6%	3.4%	0	3.4%	3.2%	1.9%
Strain heterogeneity (%)		0	0	25%	20%	0	0	16.7%	20.0%	33.3%
N50 of contigs		51,451	16,908	13,188	29,614	13,703	19,399	9,247	5,293	3,516
# Coding sequences		4,957	4,592	4,805	2,675	4,605	2,132	3,131	2,441	1,685
Coding density (%)		89.7%	88.7%	87.5%	88.5%	84.7%	87.5%	85.4%	87.6%	86.6%
rRNAs		1	1	0	3	1	0	0	0	0
tRNAs		11	13	25	42	43	46	32	30	15
MHCs		0	0	0	10	15	24	4	5	1

*MHCs, multi-heme cytochromes.

Expedition 334 Scientists (2012) Site U1379. In Vannucchi, P, Ujiie, K, Stroncik, N, Malinverno, A, and the Expedition 334 Scientists, *Proc IODP, 334: Tokyo (Integrated Ocean Drilling Program Management International, Inc)*

Torres, M.E., Muratli, J.M., and Solomon, E.A. (2014) Data report: minor element concentrations in pore fluids from the CRISP-A transect drilled during Expedition 334. In *Proc IODP/ Volume*, p. 2.