New Asgard archaea capable of anaerobic hydrocarbon cycling

Kiley W. Seitz[1], Nina Dombrowski[1,3], Laura Eme[2], Anja Spang[2,3], Jonathan Lombard[2], Jessica R. Sieber[4], Andreas P. Teske[5], Thijs J.G. Ettema[2], and Brett J. Baker[1*]

1. Department of Marine Science, University of Texas Austin, Port Aransas, TX 78373 2. Uppsala University, Uppsala Sweden 3. NIOZ, Royal Netherlands Institute for Sea Research, and Utrecht University, The Netherlands 4. University Minnesota Duluth, MN 5. Department of Marine Sciences, University of North Carolina, Chapel Hill, NC *Corresponding author

Large reservoirs of natural gas in the oceanic subsurface sustain a complex biosphere of anaerobic microbes, including recently characterized archaeal lineages that extend the potential to mediate hydrocarbon oxidation (methane and butane) beyond the Methanomicrobia. Here we describe a new archaeal phylum, Helarchaeota, belonging to the Asgard superphylum with the potential for hydrocarbon oxidation. We reconstructed Helarchaeota genomes from hydrothermal deep-sea sediment metagenomes in hydrocarbon-rich Guaymas Basin, and show that these encode novel methyl-CoM reductase-like enzymes that are similar to those found in butane-oxidizing archaea. Based on these results as well as the presence of several alkyl-CoA oxidation and Wood-Ljungdahl pathway genes in the Helarchaeota genomes, we suggest that members of the Helarchaeota have the potential to activate and subsequently anaerobically oxidize short-chain hydrocarbons. These findings link a new phylum of Asgard archaea to the microbial utilization of hydrothermally generated hydrocarbons, and extend this genomic blueprint further through the archaeal domain.

25 Short-chain alkanes, such as methane and butane, are abundant in marine sediments and play
26 an important role in carbon cycling with methane concentrations of ~1 Gt being processed
27 globally through anoxic microbial communities[1–3]. Until recently, archaeal methane cycling was
28 thought to be limited to Euryarchaeota[4]. However, additional archaeal phyla, including
29 Bathyarchaeota[5] and Verstraetarchaeota[6], have been shown to contain proteins with homology
30 to the activating enzyme methyl-coenzyme M reductase (Mcr) and corresponding pathways for
31 methane utilization. Furthermore, new lineages within the Euryarchaeota belonging to
32 *Candidatus* Syntrophoarchaeum spp., have been shown to use methyl-CoM reductase-like
33 enzymes for anaerobic butane oxidation[7]. Similar to methane oxidation in many ANME-1 archaea,
34 butane oxidation in Syntrophoarchaeum is proposed to be enabled through a syntrophic
35 interaction with sulfur reducing bacteria[7]. Metagenomic reconstructions of genomes recovered
36 from deep-sea sediments from near 2000 m depth in Guaymas Basin (GB) in the Gulf of California
37 have revealed the presence of additional uncharacterized alkyl methyl-CoM reductase-like
38 enzymes in metagenome-assembled genomes within the Methanosarcinales (Gom-Arc1)[8]. GB is
39 characterized by hydrothermal alterations that transform large amounts of organic carbon into
40 methane, polycyclic aromatic hydrocarbons (PAHs), low-molecular weight alkanes and organic
41 acids allowing for diverse microbial communities to thrive (Supplementary Table 1)[8–11].

42 Recently, genomes of novel clade of uncultured archaea, referred to as the Asgard
43 superphylum that includes the closest archaeal relatives of eukaryotes, have been recovered
44 from anoxic environments around the world[12–14]. Diversity surveys in anoxic marine sediments
45 show that Asgard archaea appear to be globally distributed[9,11,12,13]. Based on phylogenomic
46 analyses, Asgard archaea have been divided into four distinct phyla: Lokiarchaeota,

47 Thorarchaeota, Odinarchaeota and Heimdallarchaeota, with the latter possibly representing the
48 closest relatives of eukaryotes[12]. Supporting their close relationship to eukaryotes, Asgard
49 archaea possess a wide repertoire of proteins previously thought to be unique to eukaryotes
50 known as eukaryotic signature proteins (ESPs)[17]. These ESPs include homologs of eukaryotic
51 proteins, which in eukaryotes are involved in ubiquitin-mediated protein recycling, vesicle
52 formation and trafficking, endosomal sorting complexes required for transport (ESCRT)-mediated
53 multivesicular body formation as well as cytokinetic abscission and cytoskeleton formation[18].
54 Asgard archaea have been suggested to possess heterotrophic lifestyles and are proposed to play
55 a role in carbon degradation in sediments; however, several members of the Asgard archaea also
56 have genes that code for a complete Wood-Ljungdahl pathway and are therefore interesting with
57 regard to carbon cycling in sediments[14,19].

58 Here we present the first evidence of metagenome assembled genomes (MAGs),
59 recovered from Guaymas Basin deep-sea hydrothermal sediments, which represent a new
60 Asgard phylum with the metabolic potential to perform anaerobic hydrocarbon degradation
61 using a methyl-CoM reductase-like homolog.

62

63 # Results

64 **Identification of Helarchaeota genomes from Guaymas Basin sediments**. We recently obtained
65 over ~280 gigabases of sequencing data from 11 samples taken from various sites and depths at
66 Guaymas Basin hydrothermal vent sediments[20]. *De novo* assembly and binning of metagenomic
67 contigs resulted in the reconstruction of over 550 genomes (>50% complete)[20]. Within these
68 genomes we detected a surprising diversity of archaea, including >20 phyla, which appear to

69 represent up to 50% of the total microbial community in some of these samples[20]. Preliminary
70 phylogeny of the dataset using 37 concatenated ribosomal proteins revealed two draft genomic
71 bins representing a new lineage in the Asgard archaea. These draft genomes, referred to as
72 Hel_GB_A and Hel_GB_B, were re-assembled and re-binned resulting in final bins that were 82
73 and 87% complete and had a bin size of 3.54 and 3.84 Mbp, respectively (Table 1). An in-depth
74 phylogenetic analysis consisting of 56 concatenated ribosomal proteins was used to confirm the
75 placement of these final bins form a distant sister-group with the Lokiarchaeota (Figure 1a).
76 Hel_GB_A percent abundance ranged from 3.41×10^{-3}% to 8.59×10^{-5}% and relative abundance from 8.43
77 to 0.212. Hel_GB_B percent abundance ranged from 1.20×10^{-3}% to 7.99×10^{-5}% and relative abundance
78 from 3.41 to 0.22. For both Hel_GB_A and Hel_GB_B the highest abundance was seen at the site the
79 genomes bins were recovered from. These numbers are comparable to other Asgard archaea isolated
80 form these sites[20]. Hel_GB_A and Hel_GB_B had a mean GC content of 35.4% and 28%, respectively,
81 and were recovered from two distinct environmental samples, which share similar methane-
82 supersaturated and strongly reducing geochemical conditions (concentrations of methane
83 ranging from 2.3-3 mM, dissolved inorganic carbon ranging from 10.2-16.6 mM, sulfate near 21
84 mM and sulfide near 2 mM; Supplementary Table 1) but differed in temperature (28°C and 10°C,
85 respectively, Supplementary Table 1)[19].

86 Phylogenetic analyses of a 16S rRNA gene sequence (1058 bp in length) belonging to
87 Hel_GB_A confirmed that they are related to Lokiarchaeota and Thorarchaeota, but are
88 phylogenetically distinct from either of these lineages (Figure 1b). A comparison to published
89 Asgard archaeal 16S rRNA gene sequences indicate a phylum level division between the
90 Hel_GB_A sequence and other Asgard archaea[22] (Supplementary Table 2). A search for ESPs in

91 both bins revealed that they contained a similar suite compared to those previously identified in
92 Lokiarchaeota, which is consistent with their distant phylogenomic relationship (Figure 2). These
93 lineages are relatively distantly related as evidenced by their difference in GC content and
94 relatively low pairwise sequence identity of proteins. An analysis of the average amino acid
95 identity (AAI) showed that Hel_GB_A and Hel_GB_B shared 1477 genes with and AAI of 51.96%.
96 When compared to Lokiarcheota_CR4, Hel_GB_A share 634 out of orthologous genes 3595 and
97 Hel_GB_B had 624 orthologous genes out of 3157. Helarchaeota bins showed the highest AAI
98 similarity to Odinarchaeota LCB_4 (45.9%); however, it contained fewer orthologous genes (574
99 out of 3595 and 555 out of 3157 for Hel_GB_A and Hel_GB_B, respectively). Additionally, the
100 Hel_GB bins differed from Lokiarchaeota in their total gene number, for example Hel_GB_A
101 possessed 3595 genes and CR_4 possessed 4218; this difference is consistent with the larger
102 estimated genome size for Lokiarchaeum CR_4 compared to Hel_GB_A (~5.2 Mbp to ~4.6 Mbp)
103 (Supplementary Table 3, Supplementary Methods). These results add support to the phylum level
104 distinction observed for Hel_GB_A and Hel_GB_B in both the ribosomal protein and 16s rRNA
105 phylogenetic trees. We propose the name Helarchaeota after Hel, the Norse goddess of the
106 underworld and Loki's daughter for this lineage.

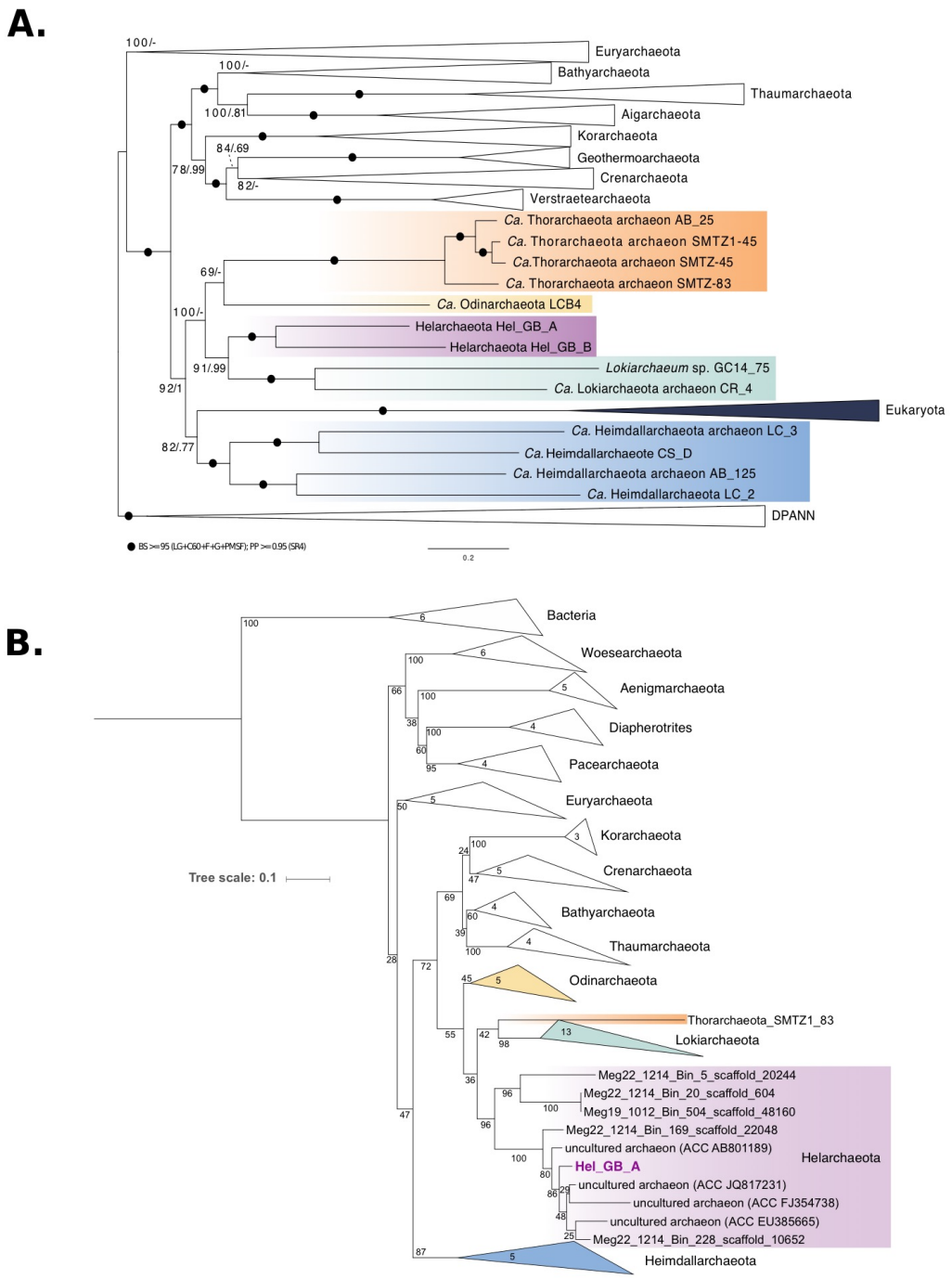
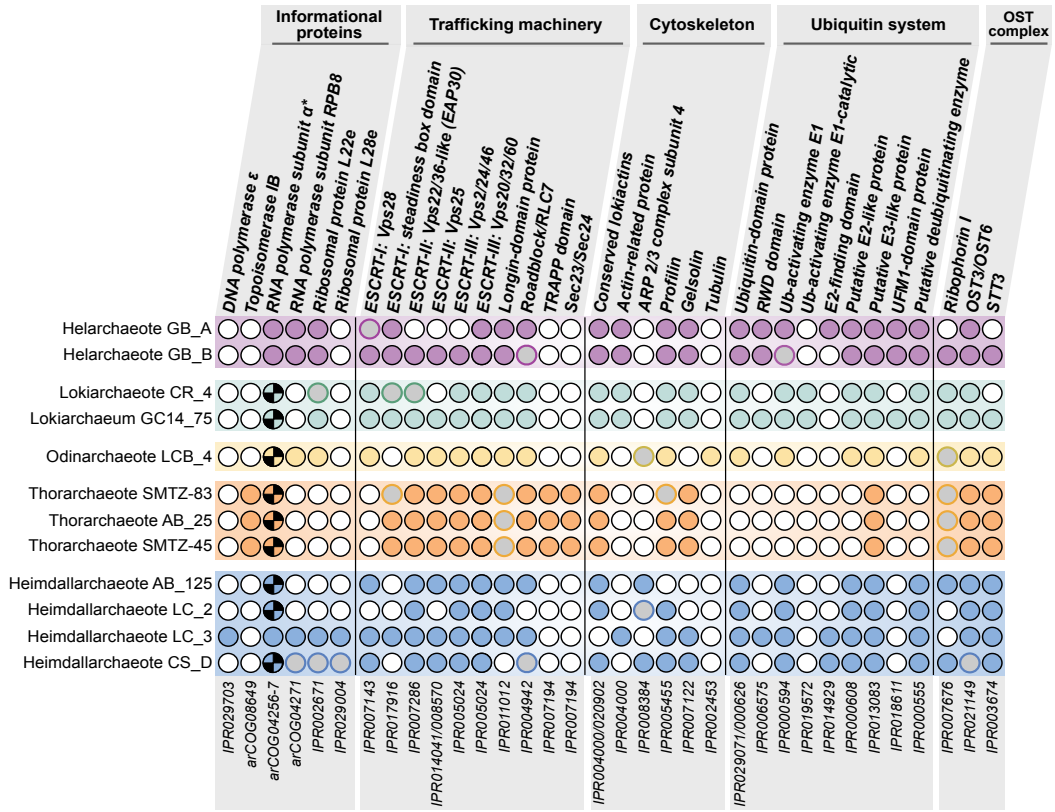


Figure 1. Phylogenomic position of Helarchaeota within the Asgard archaea superphylum (A) Phylogenomic analysis of 56 concatenated ribosomal proteins identified in Helarchaeota bins. Black circles indicate Bootstrap values greater than 95 (LG+C60+F+G+PMSF); Posterior Probability >= 0.95 (SR4). (B) Maximum-likelihood phylogenetic tree of 16S rRNA gene sequences thought to belong to Helarchaeota.

112 The phylogeny was generated using RAxML (GTRGAMMA model and number of bootstraps determined
113 using the extended majority-rule consensus tree criterion). The purple box shows possible Helarchaeota
114 sequences from GB data, as well as closely related published sequences and sequences form newly
115 identified Helarchaeota bins (identified as Megxx_xxxx_Bin_xxx_scaffold_xxxxx). Number of sequences is
116 depicted in the closed branches.
117
118 **Metabolic analysis of Helarchaeota**. To reconstruct the metabolic potential of these archaea,
119 the Helarchaeota proteomes were compared to several functional protein databases[20] (Figure
120 3a). Like many archaea in marine sediments[23], Helarchaeota may be able to utilize organic carbon
121 as they possess a variety of extracellular peptidases and carbohydrate degradation enzymes that
122 include the β-glucosidase, α-L-arabinofuranosidase and putative rhamnosidase, among others
123 (Supplementary Table 4 and 5). Degraded organic substrates can then be metabolized via
124 glycolysis and an incomplete TCA cycle from citrate to malate and a partial gamma-aminobutyric
125 acid shunt (Figure 3a, Supplementary Table 4). Both Helarchaeota bins are missing fructose-1,6-
126 bisphosphatase and have few genes coding for the pentose phosphate pathway. Genes encoding
127 for the bifunctional enzyme 3-hexulose-6-phosphate synthase/6-phospho-3-hexuloisomerase
128 (hps-phi) were identified in Hel_GB_B suggesting they may be using the ribulose monophosphate
129 (RuMP) pathway for formaldehyde anabolism. Genes coding for acetate-CoA ligase (both APM
130 and ADP-forming) and an alcohol dehydrogenase (*adhE*) were identified in both genomes
131 suggesting that the organisms may be capable of both fermentation and production of acetyl-
132 CoA using acetate and alcohols (Supplementary Table 4). Like in Thorarchaeota and
133 Lokiarchaeota, these genomes possess the large subunit of type IV Ribulose bisphosphate
134 carboxylase[19,24]. Additionally, the Helarchaeota genomes encode for the catalytic subunit of the

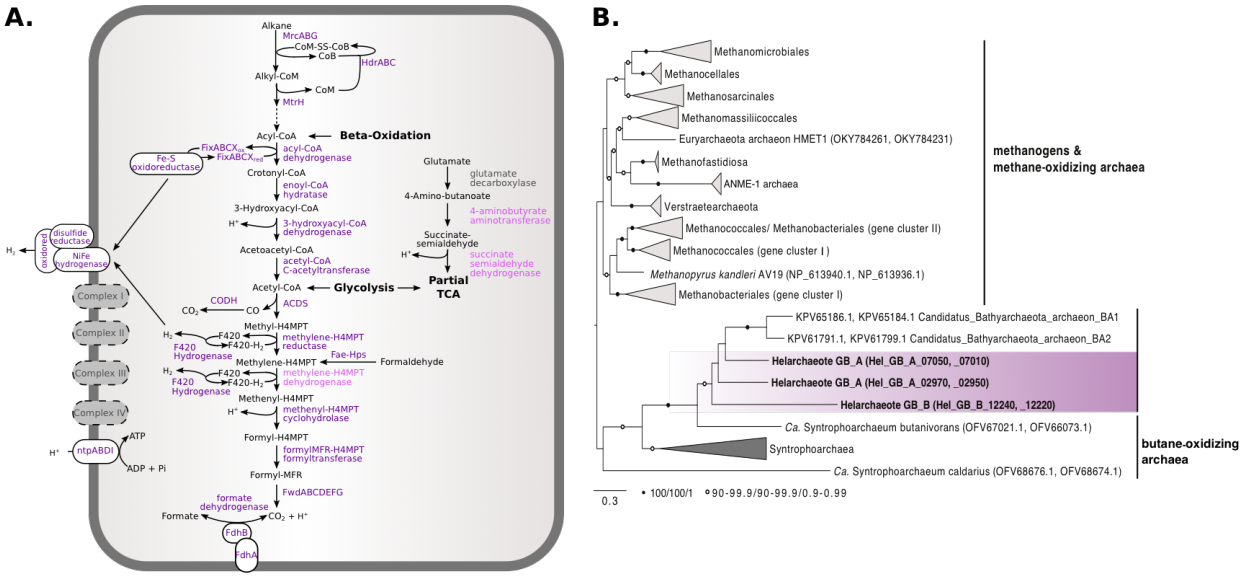
135 methanogenic type III ribulose bisphosphate carboxylase used for C-fixation[24]. Helarchaeota are
136 metabolically distinct from Lokiarchaeota as both Hel_GB draft genomes appear to lack a
137 complete TCA cycle as genes coding for citrate synthase and malate/lactate dehydrogenase are
138 absent. Both genomes also likely produce acetyl-CoA using glyceraldehyde 3-phosphate
139 dehydrogenase which is absent in Lokiarchaeota[19] (Supplementary Table 4). Helarchaeota
140 genomes lack genes that code for enzymes involved in dissimilatory nitrogen and sulfur
141 metabolism. Assimilatory genes including *sat, cysN and cysC* were found in Hel_GB_B however
142 these genes were not identified in Hel_GB_A. This absence may be indicative of species-specific
143 characteristics of their genomes or could be a results of genome incompleteness. Additional
144 genomes of members of the Helarchaeota will help to fully understand the diversity of these
145 pathways across the whole phylum.



146

147 **Figure 2. Distribution of eukaryotic signature proteins (ESPs) in Helarchaeota and other Asgard archaea.**
148 Numbers under each column correspond to the InterPro accession number (IPR) and Archaeal Clusters of
149 Orthologous Genes (arcCOG) IDs that were searched for. Full circles refer to cases in which a homologue
150 was found in the respective genomes. Empty circles with black outlines represent the absence of the ESP.
151 The checkered pattern in the RNA polymerase subunit alpha represents the fact that the proteins were
152 split, while the fused proteins are represented by the full circles. Grey circles with borders in any other
153 color represent cases where the standard profiles were not found but potential homologs where detected.
154 In the Roadblock proteins, potential homologs were detected but the phylogeny could not support the
155 close relationship of any of these copies to the Asgard archaea group closest to eukaryotes. In the Ub-
156 activating enzyme E1 represents homologs found clustered appropriately with its potential orthologs in
157 the phylogeny but the synteny of this gene with other ubiquitin-related proteins in the genome is
158 uncertain.
159
160 Interestingly, both Helarchaeota genomes have *mcrABG*-containing gene clusters
161 encoding putative methyl-CoM reductase-like enzymes (Figure 3b, Supplementary Figure 2)[4,5,7].
162 Phylogenetic analyses of both the A subunit of methyl-CoM reductase-like enzymes
163 (Supplementary Figure 2) as well as the concatenated A and B subunits (Figure 3b) revealed that
164 the Helarchaeota sequences are distinct from those involved in methanogenesis and methane
165 oxidation but cluster with homologs from butane oxidizing Syntrophoarchaea[7] and
166 Bathyarchaeota with high statistical support (rapid bootstrap support/single branch test
167 bootstrap support/posterior probability of 99.8/100/1; Figure 3b) excluding the distant homolog
168 of *Ca.* Syntrophoarchaeum caldarius (OFV68676). Analysis of the Helarchaeota mcrA alignment
169 confirmed that amino acids present at their active sites are similar to those identified on

170 Bathyarchaeota and Syntrophoarchaeum methyl-CoM reductase-like enzymes (Supplementary
171 Figure 3). In Syntrophoarchaeum, the methyl-CoM reductase-like enzymes have been suggested
172 to activate butane to butyl-CoM[7]. It is proposed that this process is then followed by the
173 conversion of butyl-CoM to butyryl-CoA; however, the mechanism of this reaction is still
174 unknown. Butyryl-CoA can then be oxidized to acetyl-CoA that can be further feed into the Wood-
175 Ljungdahl pathway to produce CO_2[7]. While some n-butane is detected in Guaymas Basin
176 sediments (usually below 10 micromolar), methane is the most abundant hydrocarbon
177 (Supplementary Table 1) followed by ethane and propane (often reaching the 100 micromolar
178 range); thus, a spectrum of short-chain alkanes could potentially be metabolized by
179 Helarchaeota[26].



180

181 **Figure 3. Metabolic inference of Helarchaeota and phylogenetic analyses of concatenated McrAB**
182 **proteins.** (A) Enzymes shown in dark purple are present in both genomes, those shown in light purple are
183 present in a single genome and ones in grey are absent. (B) The tree was generated using IQ-tree with
184 1000 ultrafast bootstraps, single branch test bootstraps and posterior predictive values from the Bayesian
185 phylogeny. White circles indicate bootstrap values of 90-99.9/90-99.9/0.9-0.99 and black filled circles

186 indicate values of 100/100/1. The tree was rooted arbitrarily between the cluster comprising canonical
187 McrAB homologs and divergent McrAB homologs, respectively. Scale bars indicate the average number
188 of substitutions per site.
189
190 **Proposed hydrocarbon degradation pathway for Helarchaeota.** Next, we searched for genes
191 encoding enzymes potentially involved in hydrocarbon utilization pathways including propane
192 and butane oxidation. Along with the methyl-CoM reductase-like enzyme that could convert
193 alkane to alkyl-CoM, Helarchaeota possess heterodisulfide reductase subunits ABC (*hdrABC*)
194 which is needed to recycle the CoM and CoB heterodisulfides after this reaction occurs (Figure 3
195 and 4)[7,8]. The conversion of alkyl-CoM to acyl-CoA is currently not understood in archaea capable
196 of butane oxidation. Novel alkyl-binding versions of methyltransferases would be required to
197 convert alkyl-CoM to butyl-CoA or other acyl-CoAs, as discussed for *Ca*. S. butanivorans[7]. Genes
198 coding for methyltransferases were identified in both Helarchaeota genomes, including a likely
199 tetrahydromethanopterin S-methyltransferase subunit H (MtrH) homolog (Figure 4;
200 Supplementary Table 4). Short-chain acyl-CoA could be oxidized to acetyl-CoA using the beta-
201 oxidation pathway via a short-chain acyl-CoA dehydrogenase, enoyl-CoA hydratase, 3-
202 hydroxyacyl-CoA dehydrogenase and acetyl-CoA acetyltransferase, candidate enzymes for all of
203 which are present in the Helarchaeota genomes and are also found in genomes of other Asgard
204 archaea (Figure 4)[19]. Along with these enzymes, genes coding for the associated electron transfer
205 systems, including an Fe-S oxidoreductase and all subunits of the electron transfer flavoprotein
206 (ETF) complex were identified in Helarchaeota (Figure 4). Acetyl-CoA produced by beta-oxidation
207 might be further oxidized to CO_2 via the Wood-Ljungdahl pathway, using among others the

classical 5,10-methylene-tetrahydromethanopterin reductase (Figure 3a and 4).

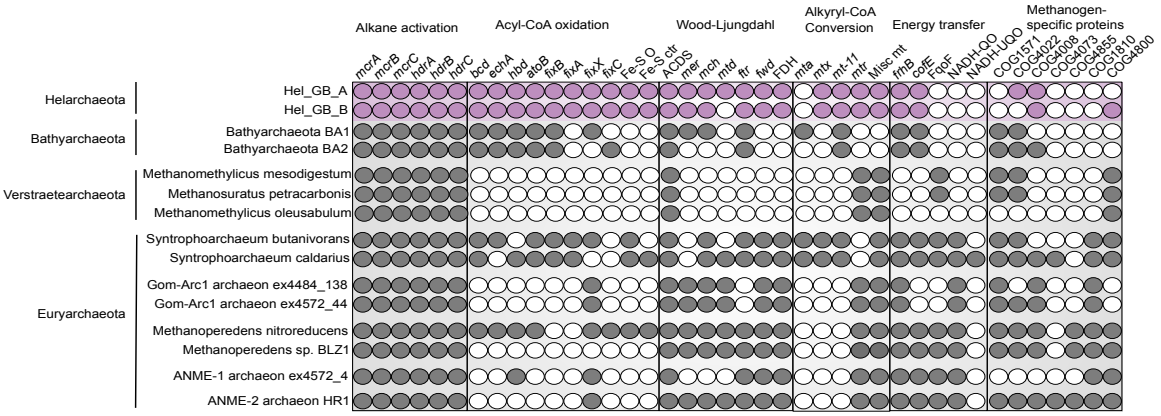


Figure 4. Comparison of Helarchaeota alkane metabolism to other alkane oxidizing and methanogenic archaea. Alkane metabolism of Helarchaeota compared to Bathyarchaeota and *Ca*. Syntrophoarchaeum sp., Verstraetearchaeota, GoM-Arc1 sp., ANME-1 sp. and ANME-2 sp. A list of genes and corresponding contig identifiers can be found in Supplementary Table 4.

Three possible energy-transferring mechanisms for Helarchaeota. To make anaerobic alkane oxidation energetically favorable, it must be coupled to the reduction of an internal electron acceptor or transferred to a syntrophic partner that can perform this reaction[7,26,27]. We could not identify an internal electron sink or any canonical terminal reductases used by ANME archaea (such as iron, sulfur or nitrogen), leading to the conclusion that a syntrophic partner organism would be necessary to enable growth on short-chain hydrocarbons. However, we could not identify any obvious syntrophic partner organisms based on co-occurrence analyses of abundance profiles of metagenomic datasets generated in this study[20].

224 An evaluation of traditional energy transferring mechanisms showed that our
225 Helarchaeota bins lack genes coding for NADH:ubiquinone oxidoreductase, F_{420}-dependent
226 oxidoreductase, $F_{420}H_2$:quinone oxidoreductase and NADH:quinone oxidoreductase that were
227 identified in *Ca.* S. butanivorans (Figure 4)[7]. These electron-carrying proteins are important for
228 energy transfer across the cell membrane and are common among syntrophic organisms[2,28,29].
229 Helarchaeota also lack genes coding for pili or cytochromes that are generally associated with
230 electron transfer to a bacterial partner, as demonstrated for different ANME archaea[26,30].
231 Therefore, Helarchaeota may use a thus far unknown approach for energy conservation. Below
232 we analyzed potential energy-transferring mechanisms that might be involved in syntrophic
233 interactions between Helarchaeota and potential partner organisms.

234 A possible candidate for energy transfer to a partner may be formate dehydrogenase
235 because substrate exchange in form of formate has previously been described to occur between
236 methanogens and sulfur-reducing bacteria[27]. Helarchaeota genomes code for the alpha and beta
237 subunits of a membrane-bound formate dehydrogenase (EC. 1.2.1.2) that could facilitate this
238 transfer (Figure 2, Supplementary Table 4). However, to our knowledge formate transfer has not
239 been shown to mediate methane oxidation. Alternatively, Helarchaeota may possess a novel
240 redox-active complex. In both Helarchaeota bins, a gene cluster was found encoding three
241 proteins that were identified as members of the HydB/Nqo4-like superfamily, Oxidored_q6
242 superfamily and a Fe-S disulfide reductase with a FlpD domain (mvhD) (Figure 5a). An analysis of
243 these three proteins showed that each possessed transmembrane motifs (Figure 5b, and
244 Supplementary Methods). While the membrane association of the disulfide reductase/FlpD

245 needs to be confirmed, interactions with the other two membrane-associated subunits may
246 allow for the bifurcated electrons to be transferred across the membrane.

247 Finally, hydrogen production and release was also considered as possible electron sink for
248 Helarchaeota. We identified several hydrogenases and putative Fe-S disulfide reductase-
249 encoding genes in the Helarchaeota genomes. Subsequent phylogenetic analyses revealed that
250 the majority of these hydrogenases represent small and large subunits of group IIIC hydrogenases
251 (methanogenic F_{420}-non-reducing hydrogenase (*mvh*)) that are usually involved in bifurcating
252 electrons from hydrogen (Supplementary Figure 4, Supplementary Table 4). In contrast, while
253 homologs belonging to the above mentioned Oxidored_q6 superfamily protein family are often
254 found to be associated with group IV hydrogenases, canonical membrane-bound group IV-
255 hydrogenases could not be identified in the genomes of the Helarchaeota. Altogether, this
256 indicates that hydrogen could play a central role in energy metabolism of Helarcharota, but the
257 absence of a classical membrane-bound hydrogenase makes it unlikely that hydrogen is the
258 major syntrophic electron carrier.

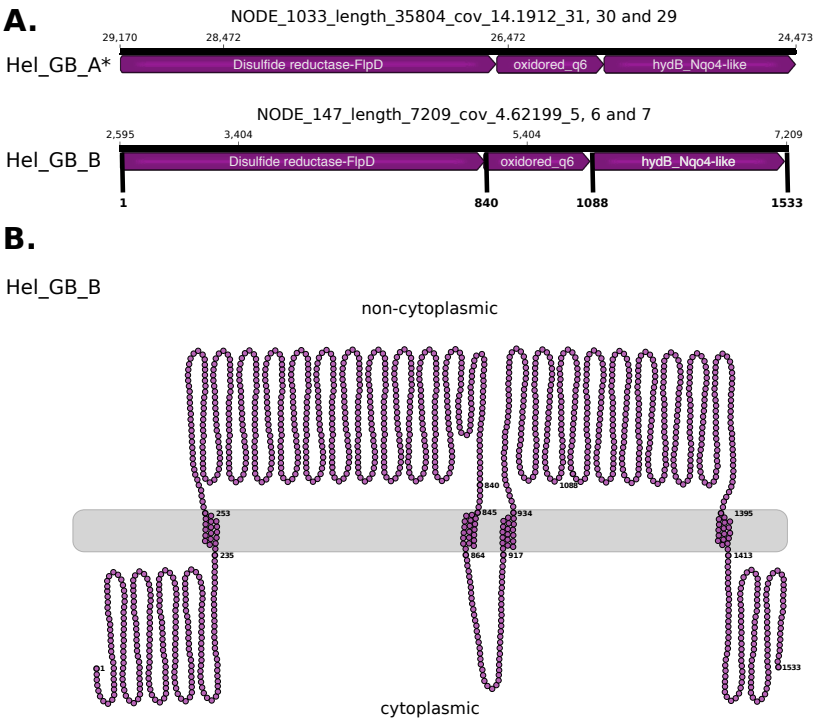


Figure 5. **Depiction of a gene cluster found in both Helarchaeota genomes that consists of genes that encode for a possible energy-transferring complex**. (A) In Hel_GB_A the complex was found on the reverse strand but has been oriented in the forward direction for clarity (asterisk). Arrows indicate the length of the reading frame. Gene names were predicted by various databases (Supplementary Methods). Small numbers located above the arrows refer to the nucleotide position for the full contig. Bold numbers on Hel_GB_B refer to the amino acid number of the whole complex. (B) Figure depicts the membrane motifs identified on NODE_147_length_7209_cov_4.62199_5, 6 and 7 using various programs (Supplementary methods). Each circle represents a single amino acid. Bold circles represent amino acids at the start of the protein, the start and end of the transmembrane sites, and the end of the complex. Numbering corresponds to the amino acid numbers of Hel_GB_B in panel (A). A full loop represents 50 amino acids and does not reflect the secondary structure of the complex.

272 ## Discussion

273 Historically methanogenesis and anaerobic methane oxidation were regarded as the only
274 examples of anaerobic archaeal short-chain alkane metabolism. The enzymes acting in these
275 pathways were considered to be biochemically and phylogenetically unique and limited to
276 lineages within the Euryarchaeota[4]. This study represents the discovery of a novel phylum and
277 the first indications for anaerobic short-chain alkane oxidation using a MCR-like homolog in the
278 Asgard archaea. Since the presence of these *mcr* genes is restricted to Helarchaeota among the
279 known Asgard archaea[19], these genes were likely transferred to Helarchaeota and do not
280 constitute an ancestral trait within the Asgard superphylum. Based on current phylogenetic
281 analysis, the Helarchaeota *mcr* gene cluster may have been horizontally acquired from either
282 Bathyarchaeota or *Ca*. Syntrophoarchaeum (Fig. 1b, Supplementary Figure 3). Due to this close
283 relationship, we based our analysis of Helarchaeota's ability to perform anaerobic short-chain
284 hydrocarbon oxidation on the pathway proposed for Ca. Syntrophoarchaeum. Helarchaeota
285 probably utilize a similar short-chain alkane as a substrate in lieu of methane, but given the low
286 butane concentrations at our site it may not be an exclusive substrate.

287 Our comparison to *Ca.* S. butanivorans shows a consistent presence in genes necessary
288 for this metabolism including a complete Wood-Ljungdahl pathway, acyl oxidation pathway and
289 internal electron transferring systems. These electron-transferring systems are essential
290 housekeeping components that act as electron carriers for oxidation reactions. Interestingly, in
291 the Wood-Ljungdahl pathway identified in *Ca.* S. butanivorans, the bacterial enzyme is 5,10-
292 methylene-tetrahydrofolate reductase (met) is thought to be substituting for the missing 5,10-
293 methylene-tetrahydromethanopterin reductase (mer)[7]. In contrast, Helarchaeota encode the

294 canonical archaeal-type mer. To render anaerobic butane oxidation energetically favorable, it
295 must be coupled to the reduction of an electron acceptor such as nitrate, sulfate or iron[7,26,27]. In
296 ANME archaea that lack genes for internal electron acceptors, methane oxidation is enabled
297 through the transfer of electrons to a syntrophic partner organism. In Syntrophoarchaeum,
298 syntrophic butane oxidation is thought to occur through the exchange of electrons via pili and/or
299 cytochromes with sulfate-reducing bacteria[7]. Helarchaeota do not appear to encode any of the
300 systems traditionally associated with syntrophy and no partner was identified in this study. Thus,
301 further research is needed to identify possible bacterial partners.

302 Furthermore, the hypothesis for Helarchaeota growth through the anaerobic oxidation of
303 short-chain alkanes remains to be confirmed as the genomes of members of this group do not
304 encode canonical routes for electron transfer to a partner bacterium. However, we identified the
305 genetic potential for potential enzymes that may be involved in transfer of electrons. Some
306 methanogenic archaea use formate for syntrophic energy transfer to a syntrophic partner;
307 therefore, the reverse reaction has been speculated to be energetically feasible for methane
308 oxidation[27]. If this is true, the presence of a membrane-bound formate dehydrogenase in the
309 Helarchaeota genomes may support this electron-transferring mechanism, however to our
310 knowledge this has never been shown for an ANME archaea so far. Alternatively, the type 3 NiFe-
311 hydrogenases encoded by Helarchaeota may be involved in transfer of hydrogen to a partner
312 organism. For example, we identified a protein complex distantly related to the *mvh-hdr* of
313 methanogens for electron transfer (Supplementary material). *Mvh-hdr* structures have been
314 proposed to be potentially used by non-obligate hydrogenotrophic methanogens for energy
315 transfer, but the directionality of hydrogen exchange could easily be reversed[2]. These

316 methanogens form syntrophic associations with fermenting, H_2-producing bacteria, lack
317 dedicated cytochromes or pili and use the *mvh-hdr* for electron bifurcation[2]. The detection of a
318 hydrophobic region in the *mvh-hdr* complex led to the suggestion that this complex could be
319 membrane bound and act as mechanism for electron transfer across the membrane; however, a
320 transmembrane association has never been successfully shown[2]. While the membrane
321 association of the disulfide reductase/FlpD needs to be confirmed, we were able to detect several
322 other transmembrane motifs in the associated proteins that could potentially allow electron
323 transfer in form of hydrogen to an external partner. Thus, while we propose that the most likely
324 explanation for anaerobic short-chain alkane oxidation in Helarchaeota is via a syntrophic
325 interaction with a partner, additional experiments are needed to confirm this working hypothesis.
326 The discovery of alkane-oxidizing pathways and possible syntrophic interactions in a new
327 phylum of Asgard archaea indicates a much wider phylogenetic range for hydrocarbon utilization.
328 Based on their phylogenetic distribution, the Helarchaeota *mcr* operon may have been
329 horizontally transferred from either Bathyarchaeota or Syntrophoarchaeum. However, the
330 preservation of a horizontally transferred pathway indicative of a competitive advantage; it
331 follows that gene transfers among different archaeal phyla reflect alkane oxidation as a desirable
332 metabolic trait. The discovery of the alkyl-CoM reductases and alkane-oxidizing pathways among
333 the Asgard archaea indicates ecological roles for these still cryptic organisms, and opens up a
334 wider perspective on the evolution and expansion of hydrocarbon-oxidizing pathways
335 throughout the archaeal domain.
336
337

338 # Methods

339 **Sample collection and processing**. Samples analyzed here are part of a study that aims to characterize
340 the geochemical conditions and microbial community of Guaymas Basin (GB) hydrothermal vent
341 sediments (Gulf of California, Mexico)[31,32]. The two genomic bins discussed in this paper, Hel_GB_A and
342 Hel_GB_B, were obtained from sediment core samples collected in December 2009 on *Alvin* dives 4569_2
343 and 4571_4 respectively[21]. Immediately after the dive, freshly recovered sediment cores were separated
344 into shallow (0-3 cm), intermediate (12-15 cm) and deep (21-24 cm) sections for further molecular and
345 geochemical analysis, and frozen at -80°C on the ship until shore-based DNA extraction. Hel_GB_A was
346 recovered from the intermediate sediment (~28°C) and Hel_GB_B was recovered from shallow sediment
347 (~10°C) from a nearby core (Supplementary Table 1); the sampling context and geochemical gradients of
348 these hydrothermally influenced sediments are published and described in detail[21,31].

349 DNA was extracted from sediment samples using the MO BIO – PowerMax Soil DNA Isolation kit
350 and sent to the Joint Genome Institute (JGI) for sequencing. A lane of Illumina reads (HiSeq–2500 1TB,
351 read length of 2x151 bp) was generated for both samples. A total of 226,647,966 and 241,605,888 reads
352 were generated for samples from dives for 4569-2 and 4571-4, respectively. Trimmed, screened, paired-
353 end Illumina reads were assembled using the megahit assembler using a range of Kmers (See
354 Supplementary Methods).

355

356 **Genome reconstruction**. The contigs from the JGI assembled data were binned using ESOM[33], MetaBAT[34]
357 and CONCOCT[35] and resulting bins were combined using DAS Tool (version 1.0)[36] (See Supplementary
358 Methods). CheckM lineage_wf (v1.0.5) was run on bins generated from DAS_Tool and 577 bins showed
359 an completeness > 50% and were characterized further[37]. 37 Phylosift[38] identified marker genes were used
360 for preliminary phylogenetic identification of individual bins (Supplementary Table 6). Thereby, we
361 identified two genomes, belonging to a previously uncharacterized phylum within the Asgard archaea,

362 which we named Helarchaeota. To improve the quality of these two Helarchaeota bins (increase the
363 length of the DNA fragments and lower total number), we used Metaspades to reassemble the contigs in
364 each individual bin producing scaffolds. Additionally, we tried to improve the overall assemblies by
365 reassembling the trimmed, screened, paired-end Illumina reads provided by JGI using both IDBA-UD and
366 Metaspades (Supplementary Methods). Binning procedures (using scaffolds longer than 2000 bp) as
367 previously described in Supplementary Methods for the original bins were repeated with these new
368 assembles. All bins were compared to the original Helarchaeota bins using blastn[39] for identification.
369 Mmgenome[40] and CheckM[37] were used to calculate genome statistics (i.e. contig length, genome size,
370 contamination and completeness). The highest quality Helarchaeota bin from each sample was chosen
371 for further analyses. For the 4572-4 dataset, the best bin was generated using the Metaspades reassembly
372 on the trimmed data and for the 4569-2 dataset the best bin was recovered using the Metaspades
373 reassembly on the original Hel bin contigs. The final genomes were further cleaned by GC content, paired-
374 end connections, sequence depth and coverage using Mmgenome[40]. CheckM was rerun on cleaned bins
375 to estimate the Hel_GB_A to be 82% and Hel_GB_B to be 87% complete and both bins were characterized
376 by a low degree of contamination (between 1.4-2.8% with no redundancy) (Table 1)[37]. Genome size was
377 estimated to be 4.6 Mbp for Hel_GB_A and 4.1 for Hel_GB_B and was calculated using percent
378 completeness and bin size to extrapolate the likely size of the complete genome. CompareM[41] was used
379 to analysis differences between Helarchaeota bins and published Asgard bins using the command python
380 comparem aai_wf --tmp_dir tmp/ --file_ext fa -c 8 aai_compair_loki aai_compair_loki_output.
381
382 **16S rRNA gene analysis.** Neither bin possessed a 16S rRNA gene sequence[38], and to uncover potentially
383 unbinned 16S rRNA gene sequences from Helarchaeota, all 16S rRNA gene sequences obtained from
384 samples 4569_2 and 4571_4 were identified using JGI-IMG annotations, regardless of whether or not the
385 contig was successfully binned. These 16S rRNA gene sequences were compared using blastn[39] (blastn -

386 outfmt 6 -query Hel_possible_16s.fasta –db New_Hel_16s -out Hel_possible_16s_blast.txt -evalue 1E-20)
387 to newly acquired 16S rRNA gene sequences from MAGs recovered from preliminary data from new GB
388 sites. A 37 Phylosift[38] marker genes tree was used to assign taxonomy to these MAGs. We were able to
389 identify five MAGs that possessed 16s and that formed a monophyletic group with our Hel_GB bins
390 (Supplementary Table 2; Megxx in Figure 2). Of the unbinned 16S rRNA gene sequences one was identified
391 as likely Helarchaeota sequence. The contig was retrieved from the 4572_4 assembly (designated
392 Ga0180301_10078946) and was 2090 bp long and encoded for an 16S rRNA gene sequence that was 1058
393 bp long. Given the small size of this contig relative to the length of the 16S rRNA gene none of the other
394 genes on the contig could be annotated. Blastn[39] comparison to published Asgard 16S rRNA gene
395 sequences was performed using the following command: blastn -outfmt 6 -query Hel_possible_16s.fasta
396 –db Asgrad_16s -out Hel_possible_16s_blast.txt -evalue 1E-20 (Supplementary Table 2). The GC content
397 of each 16S rRNA gene sequence was calculated using the Geo-omics script length+GC.pl
398 (https://github.com/Geo-omics/scripts/blob/master/AssemblyTools/length%2BGC.pl). For a further
399 phylogenetic placement, the 16S rRNA gene sequences were aligned to the SILVA database (SINA v1.2.11)
400 using the SILVA online server[42] and Geneious (v10.1.3)[43] was used to manually trim sequences. The
401 alignment also contained 16S rRNA gene sequences from the new, preliminary Helarchaeota bins. The
402 cleaned alignment was used to generated a maximum-likelihood tree with RAxML as follows: "/raxmlHPC-
403 PTHREADS-AVX -T 20 -f a -m GTRGAMMA -N autoMRE -p 12345 -x 12345 -s Nucleotide_alignment.phy -n
404 output" (Figure 1b).
405
406 **Phylogenetic analysis of ribosomal proteins**. For a more detailed phylogenetic placement, we used
407 BLASTp[44] to identify orthologs of 56 ribosomal proteins in the two Helarchaeota bins, as well as from a
408 selection of 130 representative taxa of archaeal diversity and 14 eukaryotes. The full list of marker genes
409 selected for phylogenomic analyses is shown in Supplementary Table 7. Individual protein datasets were

410 aligned using mafft-linsi[45] and ambiguously aligned positions were trimmed using BMGE (-m BLOSUM30)[46].
411 Maximum likelihood (ML) individual phylogenies were reconstructed using IQtree v. 1.5.5[47] under the
412 LG+C20+G substitution model with 1000 ultrafast bootstraps that were manually inspected. Trimmed
413 alignments were concatenated into a supermatrix, and two additional datasets were generated by
414 removing eukaryotic and/or DPANN homologues to test the impact of taxon sampling on phylogenetic
415 reconstruction. For each of these concatenated datasets, phylogenies were inferred using ML and
416 Bayesian approaches. ML phylogenies were reconstructed using IQtree under the LG+C60+F+G+PMSF
417 model[48]. Statistical support for branches was calculated using 100 bootstraps replicated under the same
418 model. To test robustness of the phylogenies, the dataset was subjected to several treatments. For the
419 'full dataset' (i.e., with all 146 taxa), we tested the impact of removing the 25% fastest-evolving sites, as
420 within a deep phylogenetic analysis, these sites are often saturated with multiple substitutions and, as a
421 result of model-misspecification can manifest in an artifactual signal[50–52]. The corresponding ML tree was
422 inferred as described above. Bayesian phylogenies were reconstructed with Phylobayes for the dataset
423 "without DPANN" under the LG+GTR model. Four independent Markov chain Monte Carlo chains were
424 run for ~38,000 generations. After a burn-in of 20%, convergence was achieved for three of the chains
425 (maxdiff < 0.29). The initial supermatrix was also recoded into 4 categories, in order to ameliorate effects
426 of model misspecification and saturation[52] and the corresponding phylogeny was reconstructed with
427 Phylobayes, under the CAT+GTR model. Four independent Markov chain Monte Carlo chains were run
428 for ~49,000 generations. After a burn-in of 20 convergence was achieved for all four the chains (maxdiff
429 < 0.19). All phylogenetic analyses performed are summarized in Supplementary Table 8, including maxdiff
430 values and statistical support for the placement of Helarchaeota, and of eukaryotes.
431
432 **Phylogenetic analysis of McrA and concatenated McrA and McrB proteins**. McrA homologs were aligned
433 using mafft-linsi [45], trimmed with trimAL[53] and the final alignment consisting of 528 sites was subjected to

434 phylogenetic analyses using v. 1.5.5 [47] with the LG+C60+R+F model. Support values were estimated using
435 1000 ultrafast boostraps[54] and SH-like approximate likelihood ratio test[55], respectively. Sequences for
436 McrA and B were aligned separately with mafft-linsi [45] and trimmed using trimAL Subsequently, McrA and
437 McrB encoded in the same gene cluster, were concatenated yielding a total alignment of 972 sites.
438 Bayesian and Maximum likelihood phylogenies were inferred using IQtree v. 1.5.5 [47] with the mixture
439 model LG+C60+R+F and PhyloBayes v. 3.2[56] using the CAT-GTR model. For Maximum likelihood inference,
440 support values were estimated using 1000 ultrafast boostraps[54] and SH-like approximate likelihood ratio
441 test[55], respectively. For Bayesian analyses, four chains were run in parallel, sampling every 50 points until
442 convergence was reached (maximum difference < 0.07; mean difference < 0.002). The first 25% or the
443 respective generations were selected as burn-in. Phylobayes posterior predictive values were mapped
444 onto the IQtree using sumlabels from the DendroPy package[57]. The final trees were rooted artificially
445 between the canonical Mcr and divergent Mcr-like proteins, respectively.
446
447 **Metabolic Analyses**. Gene prediction for the two Helarchaeota bins was performed using prodigal[58]
448 (V2.6.2) with default settings and Prokka[59] (v1.12) with the extension '–kingdom archaea'. Results for both
449 methods were comparable and yielded a total of 3,574-3,769 and 3,164-3,287 genes for Hel_GB_A and
450 Hel_GB_B, respectively, with Prokka consistently identifying fewer genes. Genes were annotated by
451 uploading the protein fasta files from both methods to KAAS (KEGG Automatic Annotation Server) for
452 complete or draft genomes to assign orthologs[60]. Files were run using the following settings: prokaryotic
453 option, GhostX and bi-directional best hit (BBH)[60]. Additionally, genes were annotated by JGI-IMG[61] to
454 confirm hits using two independent databases. Hits of interest were confirmed using blastp on the NCBI
455 webserver[44]. The dbCAN[62] and MEROPS[63] webserver were run using default conditions for identification
456 of carbohydrate degrading enzymes and peptidases respectively. Hits with e-values lower than e^-20 were

457 discarded. In addition to these methods an extended search was used to categorize genes involved in
458 butane metabolism, syntrophy and energy transfer.

459 Identified genes predicted to code for putative alkane oxidation proteins were similar to those
460 described from *Candidatus* Syntrophoarchaeum spp.. Therefore, a blastp[44] database consisting of proteins
461 predicted to be involved in the alkane oxidation pathway of *Ca.* Syntrophoarchaeum was created in order
462 to identify additional proteins in Helarchaeota, which may function in alkane oxidation. Positive hits were
463 confirmed with blastp[44] on the NCBI webserver and compared to the annotations from JGI-IMG[61],
464 Interpro[64], PROKKA[59] and KAAS[60] annotation. Genes for *mcrABG* were further confirmed by a HMMER[65]
465 search to a published database using the designated threshold values[66] and multiple MCR trees (see
466 Methods). To confirm that the contigs with the *mcrA* gene cluster were not missbined, all other genes on
467 these contigs were analyzed for their phylogenetic placement and gene content. The prodigal protein
468 predictions for genes on the contigs with *mcrA* operons were used to determine directionality and length
469 of the potential operon.

470 To identify genes that are involved in electron and hydrogen transfer across the membrane, a
471 database was created of known genes relevant in syntrophy that were download from NCBI. The protein
472 sequences of the two Helarchaeota genomes were blasted against the database to detect relevant hits
473 (E-value ≥ e ^-10). All hits were confirmed using the NCBI webserver, Interpro, JGI-IMG and KEGG.
474 Hydrogenases were identified by a HMMER search to published database using the designated threshold
475 values[67]. Hits were confirmed with comparisons against JGI annotations and NCBI blasts, the HydDB
476 database[68] and a manual database made from published sequences[69,70]. All detected hydrogenases were
477 used to generate two phylogenetic trees, one for proteins identified as small subunits and one for large
478 subunits in order to properly identify the different hydrogenase subgroups. Hydrogenases that are part
479 of the proposed complex were then further analyzed to evaluate if this was a possible operon by looking
480 for possible transcription factors and binding motifs (Supplementary Methods).

481

482 **ESP Identification**. Gene prediction for the two Helarchaeota bins was performed using prodigal[58] (V2.6.2)
483 with default settings. All the hypothetical proteins inferred in both Helarchaeaota were used as seeds
484 against InterPro[64], arCOG[71] and nr using BLAST[44]. The annotation table from Zaremba-Niedzwiedzka, *et al*.
485 2017. was used as a basis for the comparison[12]. The IPRs (or in some cases, the arCOGs) listed in the
486 Zaremba-Niedzwiedzka, et al. 2017 were searched for in the Helarchaeota genomes[12] and the resulting
487 information was used to complete the presence/absence table. When something that had previously been
488 detected in an Asgard bin was not found in a Helarchaeota bin using the InterPro/arCOG annotations,
489 BLASTs were carried out using the closest Asgard seeds to verify the absence. In some cases, specific
490 analyses were used to verify the homology or relevance of particular sequences. The details for each
491 individual ESP are depicted in supplementary materials.

492

493 **Data Availability.** The raw reads from the metagenomes described in this study are available at JGI under
494 the IMG genome ID 3300014911 and 3300013103 for samples 4569-2 and 4571-4, respectively. Genome
495 sequences are available at NCBI under the accession numbers SAMN09406154 and SAMN09406174 for
496 Hel_GB_A and Hel _GB_B respectively. Both are associated with BioProject PRJNA362212.

497

498

499

500

501

502

503

504

505 ## Tables

506 **Table 1.** Bin statistics for Helarchaeota Bins. Degree of completeness, contamination and heterogeneity

507 was determined using CheckM[37].

SeqID	Hel_GB_A	Hel_GB_B
Completeness (%)	82.4	86.92
Contamination (%)	2.8	1.40
Strain heterogeneity (%)	0	0
Scaffold number	333	182
GC content (%)	35.40	28.00
N50 (bp)	15,161	28,908
Length total (Mbp)	3.84	3.54
Estimated Genome size (Mbp)	4.6	4.1
Longest contig (bp)	52,512	72,379
Mean contig (bp)	11,531	19,467

508

509

510

511 1. G E Claypool & Kvenvolden, and K. A. Methane and other Hydrocarbon Gases in Marine

512 Sediment. *Annu. Rev. Earth Planet. Sci.* **11**, 299–327 (1983).

513 2. Thauer, R. K., Kaster, A.-K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic archaea:

514 ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* **6**, 579–591

515 (2008).

516 3. Reeburgh, W. S. Oceanic Methane Biogeochemistry. *Chem. Rev.* **107**, 486–513 (2007).

517 4. Spang, A., Caceres, E. F. & Ettema, T. J. G. Genomic exploration of the diversity, ecology, and

518 evolution of the archaeal domain of life. *Science* **357**, eaaf3883 (2017).

519 5. Evans, P. N. *et al.* Methane metabolism in the archaeal phylum Bathyarchaeota revealed by

520 genome-centric metagenomics. *Science* **350**, 434–438 (2015).

521 6. Vanwonterghem, I. *et al.* Methylotrophic methanogenesis discovered in the archaeal phylum
522 Verstraetearchaeota. *Nat. Microbiol.* **1**, 16170 (2016).
523 7. Laso-Pérez, R. *et al.* Thermophilic archaea activate butane via alkyl-coenzyme M formation.
524 *Nature* **539**, 396–401 (2016).
525 8. Dombrowski, N., Seitz, K. W., Teske, A. P. & Baker, B. J. Genomic insights into potential
526 interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments.
527 *Microbiome* **5**, 106 (2017).
528 9. Bazylinski, D. A., Farrington, J. W. & Jannasch, H. W. Hydrocarbons in surface sediments from
529 a Guaymas Basin hydrothermal vent site. *Org. Geochem.* **12**, 547–558 (1988).
530 10. Teske, A., Callaghan, A. V. & LaRowe, D. E. Biosphere frontiers of subsurface life in the
531 sedimented hydrothermal system of Guaymas Basin. *Front. Microbiol.* **5**, (2014).
532 11. Von Damm, K. L., Edmond, J. M., Measures, C. I. & Grant, B. Chemistry of submarine
533 hydrothermal solutions at Guaymas Basin, Gulf of California. *Geochim. Cosmochim. Acta* **49**,
534 2221–2237 (1985).
535 12. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic
536 cellular complexity. *Nature* **541**, 353–358 (2017).
537 13. Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and
538 eukaryotes. *Nature* **521**, 173–179 (2015).
539 14. Seitz, K. W., Lazar, C. S., Hinrichs, K.-U., Teske, A. P. & Baker, B. J. Genomic
540 reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for
541 acetogenesis and sulfur reduction. *ISME J* **10**, 1696–1705 (2016).

542 15. Jørgensen, S. L., Thorseth, I. H., Pedersen, R. B., Baumberger, T. & Schleper, C.
543 Quantitative and phylogenetic study of the Deep Sea Archaeal Group in sediments of the
544 Arctic mid-ocean spreading ridge. *Front. Microbiol.* **4**, (2013).
545 16. Jorgensen, S. L. *et al.* Correlating microbial community profiles with geochemical data in
546 highly stratified sediments from the Arctic Mid-Ocean Ridge. *Proc. Natl. Acad. Sci. U. S. A.*
547 **109**, E2846-2855 (2012).
548 17. Hartman, H. & Fedorov, A. The origin of the eukaryotic cell: a genomic investigation.
549 *Proc. Natl. Acad. Sci.* **99**, 1420–1425 (2002).
550 18. Eme, L., Spang, A., Lombard, J., Stairs, C. & J. G. Ettema, T. *Archaea and the origin of*
551 *eukaryotes*. **15**, (2017).
552 19. Spang, A. *et al.* A renewed syntrophy hypothesis for the origin of the eukaryotic cell
553 based on comparative analysis of Asgard archaeal metabolism. *Nat. Microbiol.* **Submitted**,
554 20. Dombrowski, N., Teske, A. P. & Baker, B. J. Extensive metabolic versatility and
555 redundancy in microbially diverse, dynamic Guaymas Basin hydrothermal sediments. *Nat.*
556 *Commun.* **9:4999**, (2018).
557 21. McKay, L. *et al.* Thermal and geochemical influences on microbial biogeography in the
558 hydrothermal sediments of Guaymas Basin, Gulf of California. *Environ. Microbiol. Rep.* **8**,
559 150–161 (2016).
560 22. Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea
561 using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12**, 635–645 (2014).
562 23. Lazar, C. S. *et al.* Environmental controls on intragroup diversity of the uncultured
563 benthic archaea of the miscellaneous Crenarchaeotal group lineage naturally enriched in

564 anoxic sediments of the White Oak River estuary (North Carolina, USA). *Environ. Microbiol.*
565 **17**, 2228–2238 (2015).
566 24. Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E. & Scott, S. S. Distinct form I, II, III,
567 and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco
568 evolution and structure/function relationships. *J. Exp. Bot.* **59**, 1515–1524 (2007).
569 25. Dowell, F. *et al.* Microbial Communities in Methane- and Short Chain Alkane-Rich
570 Hydrothermal Sediments of Guaymas Basin. *Front. Microbiol.* **7**, (2016).
571 26. Krukenberg, V. *et al.* Candidatus Desulfofervidus auxilii, a hydrogenotrophic sulfate-
572 reducing bacterium involved in the thermophilic anaerobic oxidation of methane. *Environ.*
573 *Microbiol.* **18**, 3073–3091 (2016).
574 27. Stams, A. J. M. & Plugge, C. M. Electron transfer in syntrophic communities of anaerobic
575 bacteria and archaea. *Nat. Rev. Microbiol.* **7**, 568–577 (2009).
576 28. Meuer, J., Kuettner, H. C., Zhang, J. K., Hedderich, R. & Metcalf, W. W. Genetic analysis
577 of the archaeon Methanosarcina barkeri Fusaro reveals a central role for Ech hydrogenase
578 and ferredoxin in methanogenesis and carbon fixation. *Proc. Natl. Acad. Sci.* **99**, 5632–5637
579 (2002).
580 29. Kunow, J., Linder, D., Stetter, K. O. & Thauer, R. K. F420H2: quinone oxidoreductase
581 from Archaeoglobus fulgidus. *Eur. J. Biochem.* **223**, 503–511 (1994).
582 30. Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E. & Boetius, A. Intercellular
583 wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* **526**,
584 587–590 (2015).

585 31. McKay, L. J. *et al.* Spatial heterogeneity and underlying geochemistry of phylogenetically
586 diverse orange and white Beggiatoa mats in Guaymas Basin hydrothermal sediments. *Deep*
587 *Sea Res. Part Oceanogr. Res. Pap.* **67**, 21–31 (2012).
588 32. Meyer, S. *et al.* Microbial habitat connectivity across spatial scales and hydrothermal
589 temperature gradients at Guaymas Basin. *Front. Microbiol.* **4**, (2013).
590 33. Dick, G. J. *et al.* Community-wide analysis of microbial genome sequence signatures.
591 *Genome Biol.* **10**, R85 (2009).
592 34. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately
593 reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
594 35. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat.*
595 *Methods* **11**, nmeth.3103 (2014).
596 36. Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication,
597 aggregation, and scoring strategy. *bioRxiv* 107789 (2017). doi:10.1101/107789
598 37. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
599 assessing the quality of microbial genomes recovered from isolates, single cells, and
600 metagenomes. *Genome Res.* gr.186072.114 (2015). doi:10.1101/gr.186072.114
601 38. Darling, A. E. *et al.* PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ*
602 **2**, e243 (2014).
603 39. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
604 search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
605 40. Karst, S. M., Kirkegaard, R. H. & Albertsen, M. mmgenome: a toolbox for reproducible
606 genome extraction from metagenomes. *bioRxiv* 059121 (2016). doi:10.1101/059121

607 41. https://github.com/dparks1134/CompareM.

608 42. Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: Accurate high-throughput multiple
609 sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829 (2012).

610 43. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software
611 platform for the organization and analysis of sequence data. *Bioinforma. Oxf. Engl.* **28**, 1647–
612 1649 (2012).

613 44. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database
614 search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

615 45. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
616 Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

617 46. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new
618 software for selection of phylogenetic informative regions from multiple sequence
619 alignments. *BMC Evol. Biol.* **10**, 210 (2010).

620 47. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. Iq-tree: A fast and effective
621 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**,
622 268–274 (2015).

623 48. Wang, H.-C., Minh, B. Q., Susko, E. & Roger, A. J. Modeling Site Heterogeneity with
624 Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Syst.*
625 *Biol.* syx068 (2017).

626 49. Jeffroy, O., Brinkmann, H., Delsuc, F. & Philippe, H. Phylogenomics: the beginning of
627 incongruence? *Trends Genet.* **22**, 225–231 (2006).

628 50. Lartillot, N. & Philippe, H. Improvement of molecular phylogenetic inference and the
629 phylogeny of Bilateria. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 1463–72 (2008).
630 51. Brown, M. W. M. *et al.* Phylogenomics demonstrates that breviate flagellates are related
631 to opisthokonts and apusomonads. *Proc. R. Soc. B Biol. Sci.* **280**, 20131755 (2013).
632 52. Susko, E. & Roger, A. J. On reduced amino acid alphabets for phylogenetic inference.
633 *Mol. Biol. Evol.* **24**, 2139–2150 (2007).
634 53. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated
635 alignment trimming in large-scale phylogenetic analyses. *Bioinforma. Oxf. Engl.* **25**, 1972–
636 1973 (2009).
637 54. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast approximation for
638 phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
639 55. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood
640 phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
641 56. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in
642 the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).
643 57. Sukumaran, J. & Holder, M. T. DendroPy: a Python library for phylogenetic computing.
644 *Bioinforma. Oxf. Engl.* **26**, 1569–1571 (2010).
645 58. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site
646 identification. *BMC Bioinformatics* **11**, 119 (2010).
647 59. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinforma. Oxf. Engl.* **30**,
648 2068–2069 (2014).

649 60. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic
650 genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185
651 (2007).
652 61. Markowitz, V. M. *et al.* IMG: the integrated microbial genomes database and
653 comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2012).
654 62. Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme
655 annotation. *Nucleic Acids Res.* **40**, W445–W451 (2012).
656 63. Rawlings, N. D., Barrett, A. J. & Bateman, A. MEROPS: the peptidase database. *Nucleic*
657 *Acids Res.* **38**, D227–D233 (2010).
658 64. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinforma.*
659 *Oxf. Engl.* **30**, 1236–1240 (2014).
660 65. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and
661 iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).
662 66. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected
663 biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
664 67. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected
665 biogeochemical processes in an aquifer system. **7**, (2016).
666 68. Søndergaard, D., Pedersen, C. N. S. & Greening, C. HydDB: A web tool for hydrogenase
667 classification and analysis. *Sci. Rep.* **6**, 34212 (2016).
668 69. Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of
669 hydrogenases: an overview. *Chem. Rev.* **107**, 4206–4272 (2007).

670 70. Vignais, P. M., Billoud, B. & Meyer, J. Classification and phylogeny of hydrogenases1.
671 *FEMS Microbiol. Rev.* **25**, 455–501
672 71. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Archaeal Clusters of Orthologous Genes
673 (arCOGs): An Update and Application for Analysis of Shared Features between
674 Thermococcales, Methanococcales, and Methanobacteriales. *Life* **5**, 818–840 (2015).
675
676
677 **Acknowledgements**

678 This study was supported in part by an Alfred P. Sloan Foundation fellowship (FG-2016-6301) and
679 National Science Foundation Directorate of Biological Sciences (award 1737298) to BJB. Sampling
680 in Guaymas Basin and post-cruise work was supported by NSF Awards OCE-0647633 and OCE-
681 1357238 to APT, respectively. The work conducted by the U.S. Department of Energy Joint
682 Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of
683 the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 provided to ND.
684
685 **Author contributions**
686 KWS, TJGE, ND and BJB conceived the study. KWS, ND, and BJB analyzed the genomic data. APT
687 collected and processed samples. KWS, AS, and LE performed phylogenetic analyses. JL analyzed
688 ESPs. KWS, AS, JRS, APT, BJB handled the metabolic inferences. BJB and KWS wrote the
689 manuscript with inputs from all authors.