nature microbiology

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41564-019-0404-y

In the format provided by the authors and unedited.

Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche

Paul-Adrian Bulzu[1,8], Adrian-Ștefan Andrei[2,8], Michaela M. Salcher[2,3], Maliheh Mehrshad[2], Keiichi Inoue[5], Hideki Kandori[6], Oded Beja[4], Rohit Ghai[2]* and Horia L. Banciu[1,7]

[1]Department of Molecular Biology and Biotechnology, Faculty of Biology and Geology, Babeș-Bolyai University, Cluj-Napoca, Romania. [2]Department of Aquatic Microbial Ecology, Institute of Hydrobiology, Biology Centre of the Academy of Sciences of the Czech Republic, České Budějovice, Czech Republic. [3]Limnological Station, Institute of Plant and Microbial Biology, University of Zurich, Kilchberg, Switzerland. [4]Faculty of Biology, Technion Israel Institute of Technology, Haifa, Israel. [5]The Institute for Solid State Physics, The University of Tokyo, Kashiwa, Japan. [6]Department of Life Science and Applied Chemistry, Nagoya Institute of Technology, Nagoya, Japan. [7]Molecular Biology Center, Institute for Interdisciplinary Research in Bio-Nano-Sciences, Babeș-Bolyai University, Cluj-Napoca, Romania. [8]These authors contributed equally: Paul-Adrian Bulzu, Adrian-Ștefan Andrei. *e-mail: ghai.rohit@gmail.com

Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche

Paul-Adrian Bulzu[1#], Adrian-Ştefan Andrei[2#], Michaela M. Salcher[2,3], Maliheh Mehrshad[2], Keiichi Inoue[5], Hideki Kandori[6], Oded Beja[4], Rohit Ghai[2*], Horia L. Banciu[1,7]

[1]Department of Molecular Biology and Biotechnology, Faculty of Biology and Geology, Babeş-Bolyai University, Cluj-Napoca, Romania

[2] Department of Aquatic Microbial Ecology, Institute of Hydrobiology, Biology Centre of the Academy of Sciences of the Czech Republic, České Budějovice, Czech Republic.

[3]Limnological Station, Institute of Plant and Microbial Biology, University of Zurich, Kilchberg, Switzerland.

[4]Faculty of Biology, Technion Israel Institute of Technology, Haifa, Israel.

[5]The Institute for Solid State Physics, The University of Tokyo, Kashiwa, Japan.

[6]Department of Life Science and Applied Chemistry, Nagoya Institute of Technology, Nagoya, Japan.

[7]Molecular Biology Center, Institute for Interdisciplinary Research in Bio-Nano-Sciences, Babeş-Bolyai University, Cluj-Napoca, Romania.

# These authors contributed equally to this work

*Corresponding author: Rohit Ghai

Institute of Hydrobiology, Department of Aquatic Microbial Ecology, Biology Centre of the Academy of Sciences of the Czech Republic, Na Sádkách 7, 370 05, České Budějovice, Czech Republic.

Phone: +420 387 775 881

Fax: +420 385 310 248

E-mail: ghai.rohit@gmail.com

Supplementary Information

Supplementary Discussion

Sediment chemical analyses: The analysis of leachable salt contents of Tekirghiol (0-40cm) and Amara (0-10cm) sediments indicated that the dominant cations and anions ($g \cdot Kg^{-1}$) were: Na^+ (16.5 and 7.0), K^+ (1.0 and 0.22), Mg^{2+} (1.1 and 4.0), Cl^- (27.7 and 11.2), and SO_4^{2-} (0.25 and 13.2) (Supplementary Table 9).

Environmental distribution of Heimdallarchaeia: From the 279 sequences available in the SILVA database[1] at the time of writing (October 2018), we infer that the most likely habitat of Heimdallarchaeia is represented by marine and brackish sediments (with scarce reports from water column or microbial mats) in both deep (sediments from: hydrothermal vents, subseafloor, cold seeps, etc.) and shallow habitats (sediments from: marshes, mangrove fields and shallow waterbodies).

Phylogenomics: The phylogenomic trees (**Figure 1a, b**) showed that the basal branches of Thorarchaeia were represented by MAGs recovered from the Tekirghiol hypersaline sediment (i.e. TEKIR_14 and TEKIR_12S). The other ones were shown to form a compact cluster (n=8), which appeared to be the outcome of a more recent diversification event (as assessed by short branch lengths) in brackish environments. Thus, the two brackish Amara MAGs (AMARA_2 and AMARA_8) clustered together with the: estuarine Mai Po ones (MP8T_1, MP9T_1 and MP11T_1)[2], Baltic Sea AB_25[3] and the White Oak River Estuary SMTZ1_45[3]; forming a genus-level clade (as assessed by amino acid identity values – **Supplementary Table 1 "AAI_Thorarchaeia"**). Noteworthy, this reduced phylogenomic diversity within Thorarchaeia contrasts with the highly divergent MAGs of Loki- and Heimdallarchaeia, with which it shares common ancestry (**Figure 1a, b**). The presence of the younger RuBisCO type (i.e. IV)[4] (**Supplementary Figure 5**) in Thorarchaeia supports the '(more)recent diversification', as the other Asgardarchaeota (sampled to date) still maintain the ancestral type III (**Supplementary Figure 5**).

The trees generated by Bayesian inference (Figure 1b) and maximum likelihood (ML) (Supplementary Figure 1b) methods recovered the eocyte topology (Eukaryota branching within Asgardarchaeota), even though they failed to confidently resolve the branching point of eukaryotes. We consider that the three-domain topology recovered in a previous study[5] by using ML methods is likely an artifact caused by the long-branch attraction phenomenon, since our phylogenomic trees with SR4 amino acid recoded alignments (used for reducing substitution saturation and compositional heterogeneity) provide high support for the eocyte tree.

Rhodopsins: The recently discovered heliorhodopsins (HeR) were reported to be present in Heimdallarchaeia[6]. HeR share low sequence similarities with all known type-1 rhodopsins, and even more remarkably they are oriented in an opposite topology in the membrane, harbouring a cytoplasmic N-terminus (in contrast, type-1 rhodopsins possess extracellular N-terminus). Heimdallarchaeia appear to possess the most diverse collection of rhodopsins within the Asgardarchaeota superphylum at large: type-1, HeR and schizorhodopsins. In contrast Loki- and Thorarchaeia were found to contain only schizorhodopsins (SzR). Lokiarchaeia were found to harbour sequences that have a helix-turn-helix motif and which are similar to bacterioopsin-activator like proteins. These proteins, which were found to be in the proximity of the SzR (**Supplementary Figure 6**) have been previously characterised in *Halobacterium halobium*, where they are hypothesized to work as low oxygen sensors capable of activating the bacteriorhodopsin gene[7]. However, given the absence of the sensing N-terminal domain (NifL-like), the connection between the bacterioopsin-activator like proteins and SzR is unclear. Although, the presence of Asgardarchaeota in habitats with light exposure possibility (e.g. lake water column, estuarine

sediments, mangrove sediments, microbial mats, etc.)[2,3] fulfils the condition needed for rhodopsin usage, additional experimental data is needed to clarify the functions of these proteins.

Metabolism: We observed that while all Asgardarchaeota clades (i.e. Thor-, Loki- and Heimdallarchaeia) possess transporters for the uptake of organic compounds, their preferences towards their categories show phyla specificity. We found that the genomic repertoire of Lokiarchaeia is highly enriched (up to ten times in comparison to rest of Asgardarchaeota) in genes encoding for the uptake of modified monosaccharides. Thus, Lokiarchaeia had 5.85 transporters/Mb for the glycoside/pentoside/hexuronide symporters, while Thor- and Heimdallarchaeia barely reached 0.66 and respectively 0.52 transporters/Mb. We reason that this high genomic density is linked to the Lokiarchaeia's inferred capacity to degrade cellulose (by employing the synergic action of: endoglucanases, beta-glucosidases and cellobiose phosphorylases), and to import the resulted monosaccharides into the cytosol. These monosaccharides could be used to fuel the cell machinery (by transformation of glucose to acetyl-CoA or lactate with subsequent ATP production), or transformed into glycogen storage for later usage (through glycogen/starch synthases, 1,4-alpha-glucan branching enzymes; starch/glycogen phosphorylases, alpha-amylases, neopullulanases). Peptide/nickel transporters were found to have a higher density in Thorarchaeia (Thor-: 2.48 transporters/Mb; Heimdall-: 0.79 transporters/Mb; Lokiarchaeia: 0.51 transporters/Mb). As previously reported in Thor-[2,8], we found that both Loki- and Heimdallarchaeia have the mechanisms involved in peptide and amino-acid uptake, as well as the enzymatic repertoire needed for their degradation to keto-acids and acetyl-CoA (endopeptidases: PepB, PepD, PepP and PepT; aminotransferases: AspB, ArgD, IlvE, GlmS, HisC and PuuE; glutamate dehydrogenases/ oxidoreductases). Among Asgardarchaeota, Thorarchaeia was the only phylum in which we found the glucarate and dicarboxylate uptake systems, as well as the metabolic pathways needed to catabolize putrescine to succinate. Inorganic phosphate uptake could be achieved by all Asgardarchaeota through the usage of PiT family transporters. Lokiarchaeia was found to harbor the PhoR-PhoB two component system involved in phosphate uptake regulation, while Heimdallarchaeia (i.e. LC_2 and RS678) was found to encode ABC-type transporters for phosphonates: refractory forms of phosphorus found to be highly abundant in marine systems[9].

Sulfur uptake can be accounted for by the presence of sulphate permease SulP in Loki- and Thorarchaeia, as well as the ABC-type sulfonate/nitrate/taurine transport system, predominantly in the latter. Cysteine may also serve as a source of sulfur which can be mobilized by cysteine desulfurases in all three Asgardarchaeota phyla.

Remarkably, two Heimdallarchaeia MAGs (LC_2 and LC_3) encoded genes for the synthesis and degradation of cyanophycin (**Figure 3**), a non-ribosomally produced polypeptide used as a carbon and nitrogen storage pool in bacteria[10].

While a canonical pentose phosphate pathway (PPP) is lacking in all analyzed Asgardarchaeota, evidence indicates that simple sugar interconversions are carried out by the non-oxidative branch of this pathway. In Thorarchaeia, the xylulose part of the non-oxidative branch was found to be largely complete with the key enzyme transketolase present in multiple MAGs. We also identified two copies of this enzyme in Heimdall_RS678, which points towards similar functional capabilities; however, with low support within the Heimdallarcheia phylum itself. The absence of transketolase enzyme, in Loki- as well as in most Heimdallarchaeia, and the presence of components of the Ribulose Monophosphate Pathway (RuMP) indicates it as a potential alternative for PPP, as previously reported in the case of *Thermococcus kodakaraensis*[11]. The presence of uridine phosphorylase within all three analyzed phyla indicates that the nucleotide degradation pathways could serve as an additional ribose source. While all analyzed Asgardarchaeota phyla encode components of the glycolytic pathway (i.e. type Embden-Meyerhof-Parnas), three Heimdallarchaeia MAGs (LC_3, AB_125 and AMARA_4) were found to employ non-canonical ADP-dependent kinases

and, as previously noted[2], no glucokinase homologue could be identified in Thorarchaeia. We reason that the well represented non-oxidative PPP in this group could either represent an alternative point of entry for sugars in the EMP, or that the function of canonical glucokinase is achieved by yet unidentified archaea-specific sugar kinases[12].

Regarding pyruvate metabolism, in both Loki- and Thorarchaeia phosphoenolpyruvate (PEP) can be converted to pyruvate by phosphoenolpyruvate synthase (pps) as well as pyruvate kinase (pyk), the latter of which we identified in Heimdallarchaeia as well. Also present in all three phyla, malic enzyme (maeA) is probably responsible for catalyzing the oxidative decarboxylation of malate to pyruvate, CO_2 and NADH. Additionally, all groups encode pyruvate phosphate dikinase (PPDK), a PPi-utilizing enzyme which interconverts PEP and pyruvate. Among the anaplerotic reactions for CO_2 fixation, reversible carboxylation of acetyl-CoA to pyruvate may be achieved in all groups by the activity of pyruvate:ferredoxin oxidoreductase (PFOR). PEP can be synthesized by the phosphoenolpyruvate carboxykinases (PEPCK) present in all three groups, starting from oxaloacetate. Therefore, under gluconeogenic conditions, maeA and/or PEPCK, in combination with pps/PPDK, is used for directing C4 carbon intermediates from the TCA cycle, when present, to PEP[13] - the precursor for gluconeogenesis. Noteworthy, as a hint of potential aerobiosis, we identified exclusively in Heimdallarchaeia MAGs (LC2 and LC3) genes encoding for pyruvate oxidase (poxL), which catalyzes the decarboxylation of pyruvate in the presence of phosphate and oxygen, yielding carbon dioxide, hydrogen peroxide and acetyl phosphate[14].

Among all analyzed phyla, the complete TCA cycle was identified in Loki- and Heimdallarchaeia. Inquiringly, genomes from all three phyla, with the exception of Heimdall LC_3, were found to lack the membrane anchoring subunits of succinate dehydrogenase (sdhC, sdhD), while Lokiarchaeia contain key genes (isocitrate dehydrogenase, 2-oxoglutarate-ferredoxin oxidoreductase, ATP-citrate lyase) that are indicative of a reductive TCA cycle, involved in the autotrophic fixation of CO_2. Also, malate dehydrogenase (MDH) was identified in Loki- and Thorarchaeia only by homology search (COG2055) and 3-dimensional structure predictions[15].

Through analyzing Asgardarchaeota genomes assembled in this study, we confirm previous findings[2,16] regarding the presence of a complete Wood-Ljungdahl pathway in both Loki- and Thorarchaeia. This pathway was found absent in the partial TEKIR_3 and AMARA_4 (13 and respectively 30% CheckM[17] estimated completeness) MAGs, as well as in the previously published (AB_125, LC_2, LC_3, RS678) Heimdallarchaeia ones. Importantly, the Wood-Ljungdahl pathway is confined to anoxic niches harboring low reducing substrates such as carbon monoxide (CO) and H_2[18,19]. In spite of the fact that in anaerobic as well as aerobic microbes, CO may be used as both energy and carbon source[20], the ability to utilize CO is conditioned by the presence of enzyme complexes known as carbon monoxide dehydrogenases (CODHs)[21]. In Heimdallarchaeia, we identified all three major subunits of the aerobic type CODH (coxSML). In this case, electrons generated from CO oxidation may be shuttled to oxygen or nitrate, which may serve as final electron acceptors[22–24]. However we observe that Thor- and Lokiarchaeia encode all components of the bifunctional and oxygen-sensitive[25] enzyme complex known as carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS). This complex is part of the Wood-Ljunghdahl pathway and is responsible for catalyzing reactions involved in autotrophic fixation of CO_2.

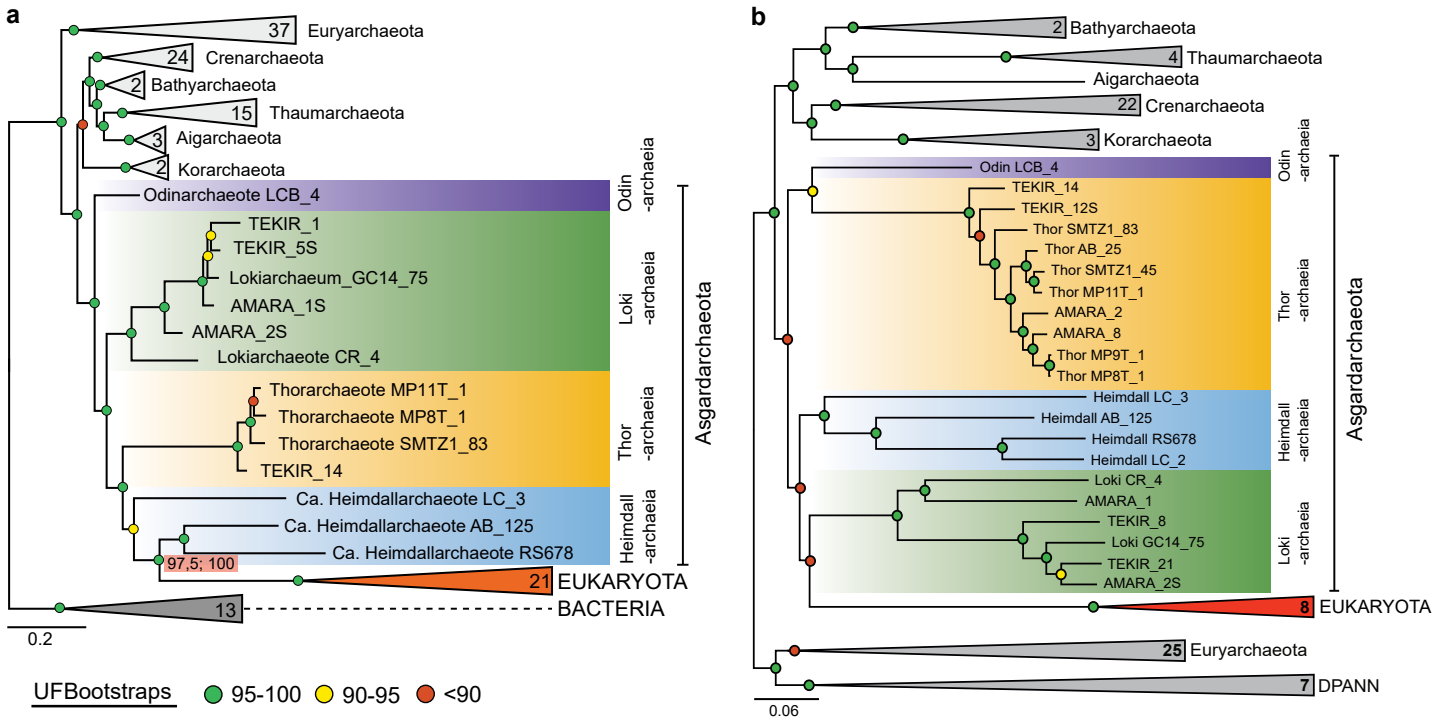
Regarding oxidative phosphorylation, while V/A-type ATPase appears mostly complete in Loki- and Thorarchaeia, the other components involved in oxidative phosphorylation (the electron transport chain), are missing or incomplete, emphasizing anaerobic lifestyles. For Heimdallarchaeia we could identify complete V/A-type ATPase, succinate dehydrogenase, almost complete NADH:quinone oxidoreductase and importantly – cytochrome c oxidase – another hallmark of aerobiosis.

Uniquely among known Archaea, we identify all components of the aerobic tryptophan degradation pathway (i.e. kynurenine pathway[26]) in three published Heimdallarchaeia genomes (LC_2, LC_3,

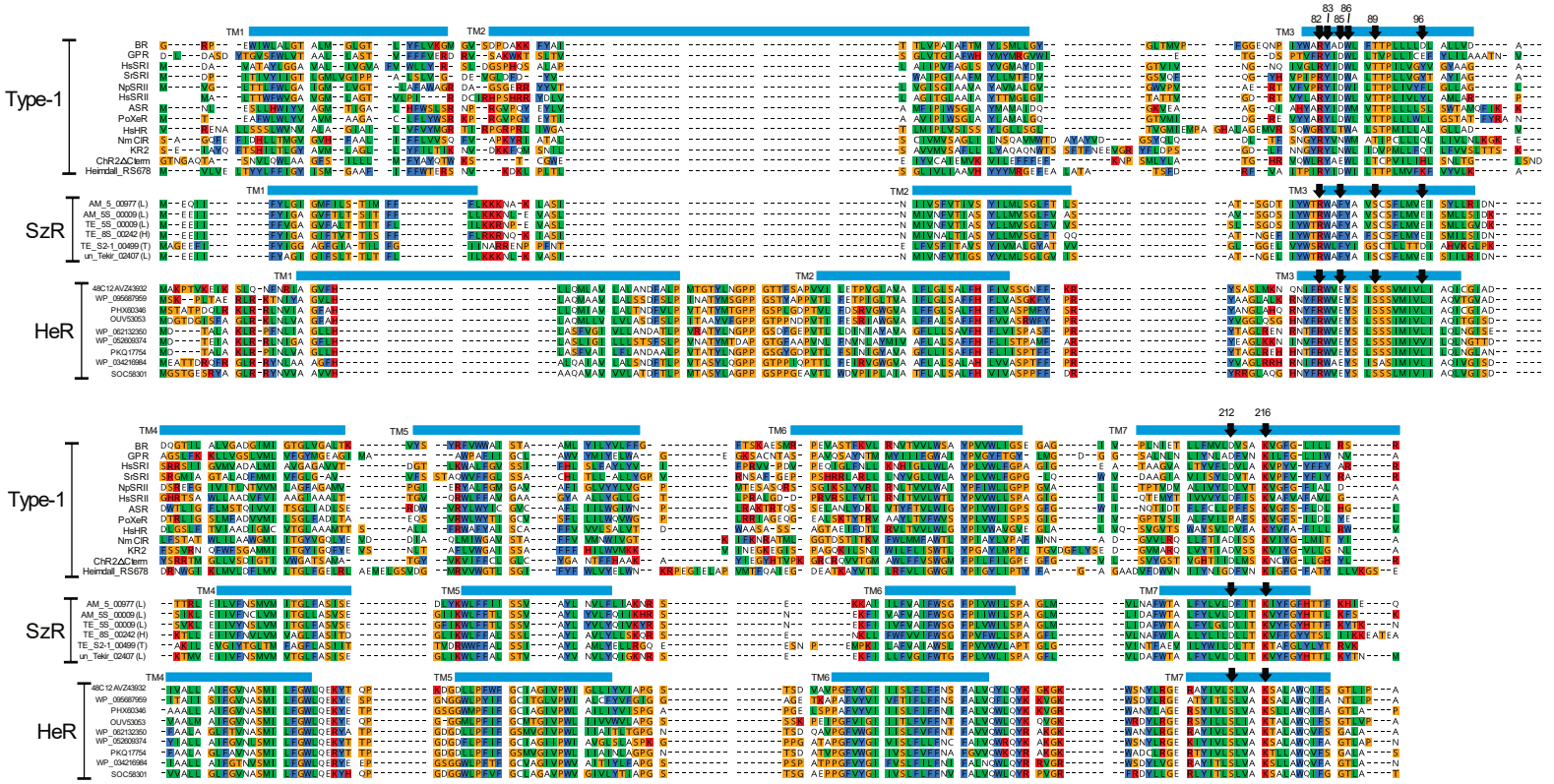
RS678). The performed evolutionary history inferences indicated that the kynurenine pathway was probably acquired by Heimdallarchaeia ancestor through lateral gene transfer from bacteria (**Supplementary Figure 3**). The phylogenetic trees, constructed with key enzymes of the pathway, pointed towards enzyme-specific evolutionary rates (**Supplementary Figure 3**). Thus, while Heimdallarchaeia MAGs formed monophyletic clusters in the dioxygenases trees (TDO, HAAO), in the kynurenine monooxygenase (KMO) one they segregated into two independent clusters. The MAG Heimdall_LC_3 showed an affinity to cluster with a sediment-dwelling Bacteroidetes (with low statistical support).

Among Asgardarchaeota, the Heimdallarchaeia were found to possess genes encoding for sulfide:quinone oxidoreductases, enzymes used in sulfide detoxification and energy generation *via* quinone pool reduction (**Figure 3**). As sulfide binds to cytochrome c oxidase and inhibits aerobic respiration[27], the presence of these enzymes in Heimdallarchaeia could point towards a detoxification role. The fact that one Heimdallarchaeia MAG described in this study (i.e. AMARA_4) had the sulfide:quinone oxidoreductase gene and the other MAGs recovered from the same sample did not (i.e. 7 Loki- and 3 Thorarchaeia MAGs), suggests that the highly lipophilic sulfide does not interfere with the aerobic metabolism, nor is it part of a conserved energy generation strategy in Asgardarchaeota. The superoxide dismutase, catalases, and glutathione peroxidases found in Heimdallarchaeia may act in alleviating the oxidative damage generated by a facultative aerobic metabolism.

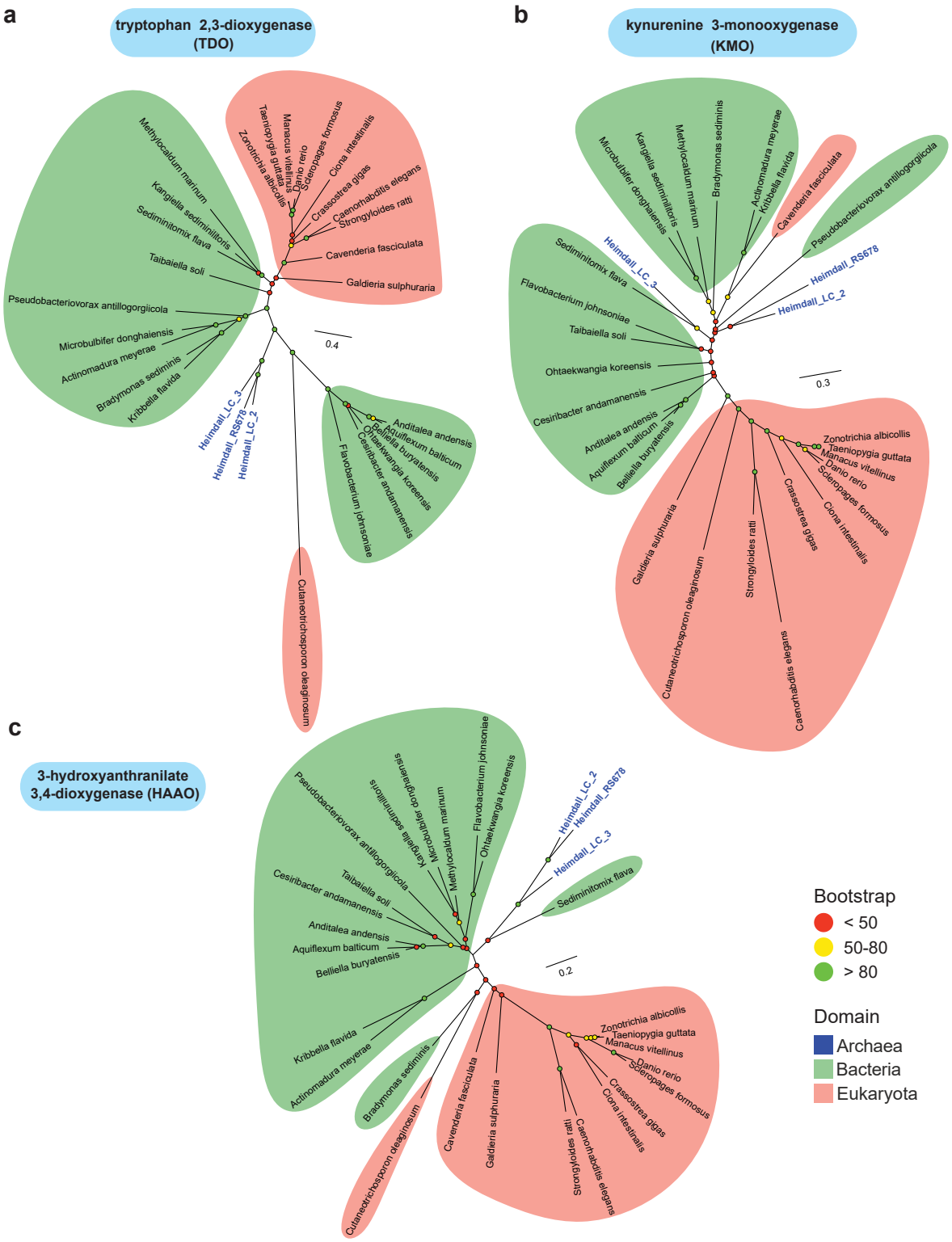
In Heimdallarchaeia the following archaellum[28] components were identified: flaG, flaI, flaJ, flaK, flaH and a homolog of the archaeal flagellin (IPR013373). The presence of sensory methyl-accepting chemotaxis proteins (MCPs), together with a complete chemotaxis signal transduction pathway (CheA, B, C, D, R, W, Y), suggests that Heimdallarchaeia may be motile.



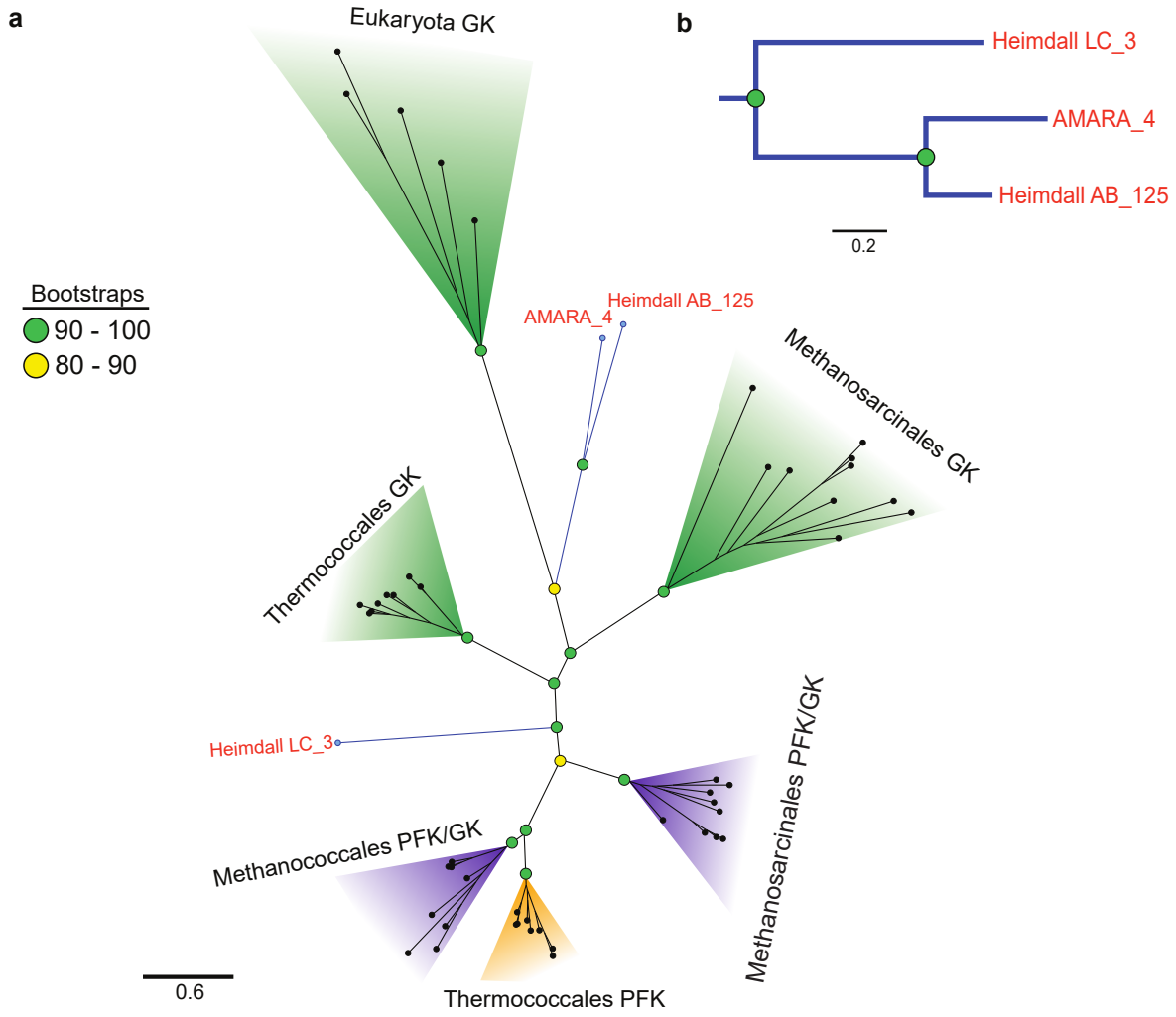
Supplementary Figure 1. Maximum likelihood (ML) inferences of Asgardarchaeota phylogeny. **a**, ML phylogeny (GTR+I+G4, general time reversible model of nucleotide substitution) of concatenated SSU+LSU gene sequences spanning all domains of life (n=97 archaeal, n=21 eukaryotic, n=13 bacterial lineages; details in Supplementary Table 6). The red rectangle indicates the results of the Shimodaira-Hasegawa test and ultrafast bootstrapping for the Eukaryota/Heimdallarchaeia cluster. **b**, ML phylogenomic tree (C60SR4, 60-profile protein models of amino acid substitution with SR4 recoding) based on 48 ribosomal proteins (Supplementary Table 7) including 85 archaeal and 8 eukaryotic lineages (Supplementary Table 6). UFBootstrap values are indicated by colored circles (legend in the lower-left part of the figure). The scale bars indicate number of substitutions per site.

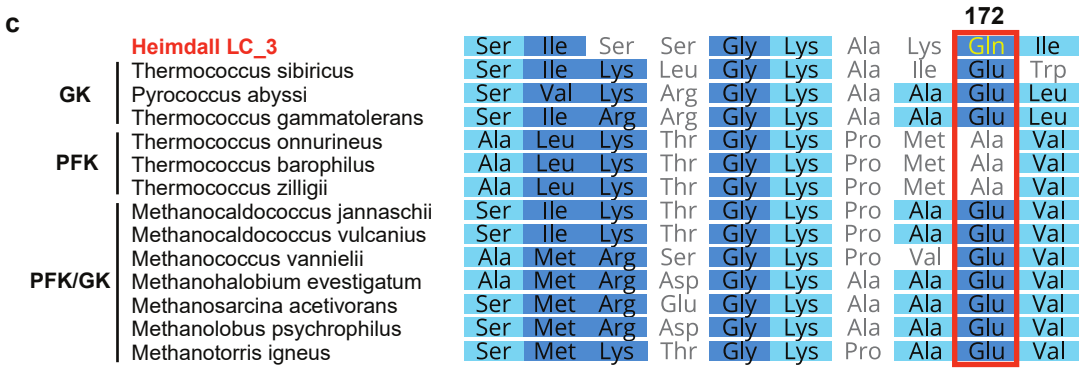


Supplementary Figure 2. Multiple alignment of type-1, schizorhodopsins (SzR) and heliorhodopsins (HeR). Transmembrane helices (labeled as TM1-7) are shown for the first sequence of each group at the top. Sequences of type-1 rhodopsins are - bacteriorhodopsin (BR), green absorbing proteorhodopsin (GPR), sensory rhodopsin I from Halobacterium salinarum (HsSRI) and Salinibacter ruber (SrSRI), sensory rhodopsin II from Natronomonas pharaonis (NpSRII) and H. salinarum (HsSRII), Anabaena sensory rhodopsin (ASR), xenorhodopsin from Parvularcula oceani (PoXeR), halorhodopsin from H. salinarum (HsHR), chloride-pumping rhodopsin from Nonlabens marinus (NmClR), sodium-pumping rhodopsin from Krokinobacter eikastus (KR2), and cation channelrhodopsin-2 from Chlamydomonas reinhardtii (C-terminal side omitted, ChR2 ΔC-term) and putative proton pump from Heimdallarchaeota RS678 MAG assembled from a Red Sea metagenome. Sequences of Asgardarchaeota schizorhodopsins: Six sequences are shown, the number in brackets indicates taxonomic affiliation of the metagenome assembled genome- L – Lokiarchaeia, H – Heimdallarchaeia and T – Thorarchaeia. Sequences of heliorhodopsins are – 48C12: original actinobacterial fosmid clone from Lake Kinneret, WP_095687959: freshwater actinobacterium Ca. Nanopelagicus abundans, PHX60346: Actinobacteria bacterium (Lake Baikal metagenome), OUV53053: Actinomycetales bacterium TMED115 (marine metagenome), WP_062132350 Demequina aestuarii, WP_052609374: freshwater Actinobacteria bacterium IMCC26256, PKQ17754: Actinobacteria bacterium HGW-Actinobacteria-8 (groundwater metagenome) WP_034216984: Actinoplanes subtropicus, SOC58301: Ornithinimicrobium pekingense. In all groups the functionally important positions 82, 83, 85,86, 89,96, 212 and 216 (bacteriorhodopsin numbering) are marked with black arrows.

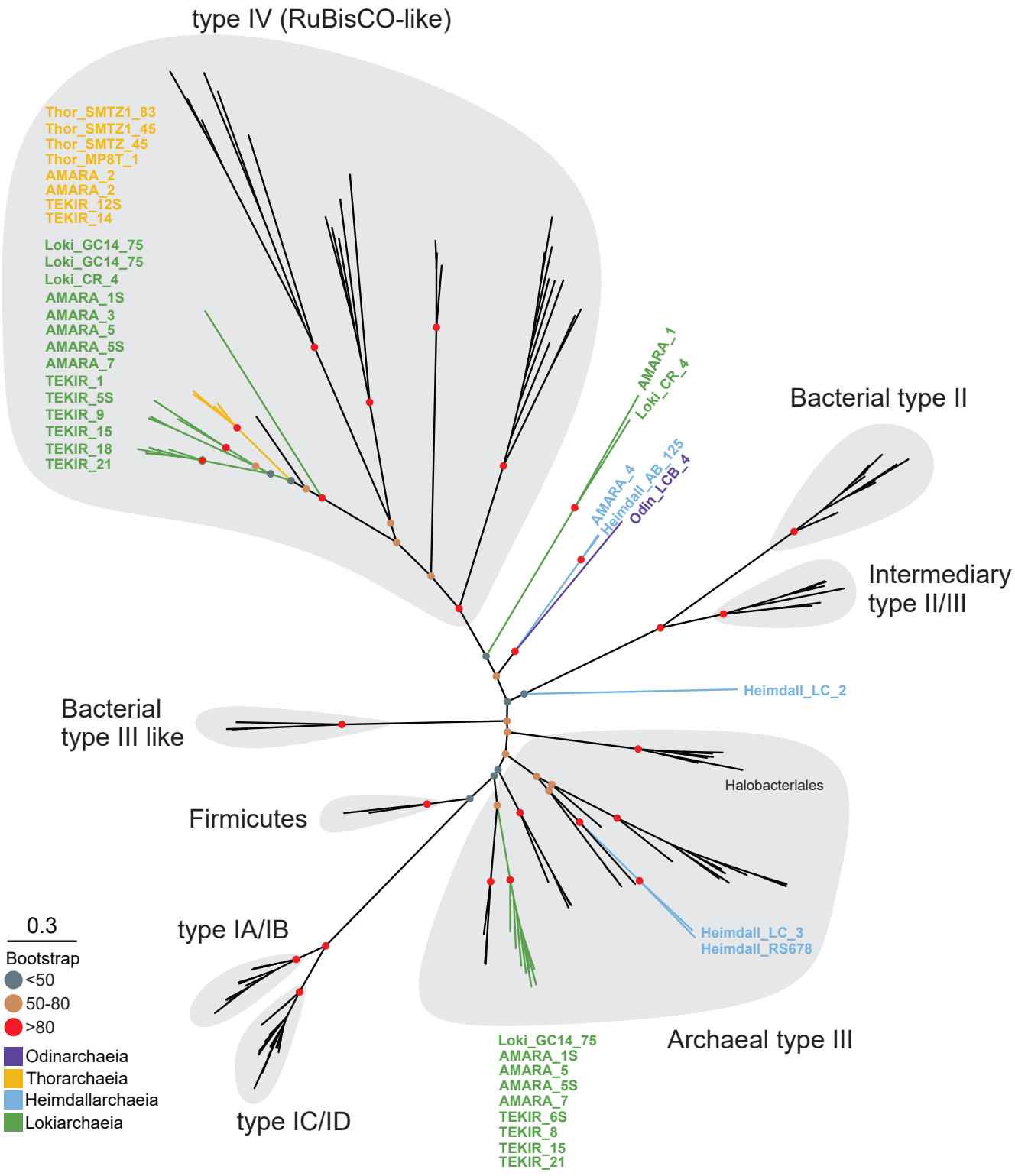


Supplementary Figure 3. **Maximum likelihood phylogenetic trees of key enzymes of the kynurenine pathway. a**, tryptophan 2,3-dioxygenase, **b**, kynurenine 3-monooxygenase and **c**, 3-hydroxyanthranilate 3,4-dioxygenase, respectively. Standard bootstrap values are indicated by colored circles (legend in the lower-right part of the figure). The names of Heimdallarchaeia MAGs are highlighted in blue. Along with sequences recovered from Heimdallarchaeia MAGs (Heimdall_LC_2, Heimdall_LC_3, Heimdall_RS678), enzymes identified in bacterial (n=15) and eukaryotic (n=12) lineages harbouring the complete kynurenine pathway were employed for phylogenetic inferences. Sequence accession numbers and taxonomy for enzymes are presented in Supplementary Table 8.

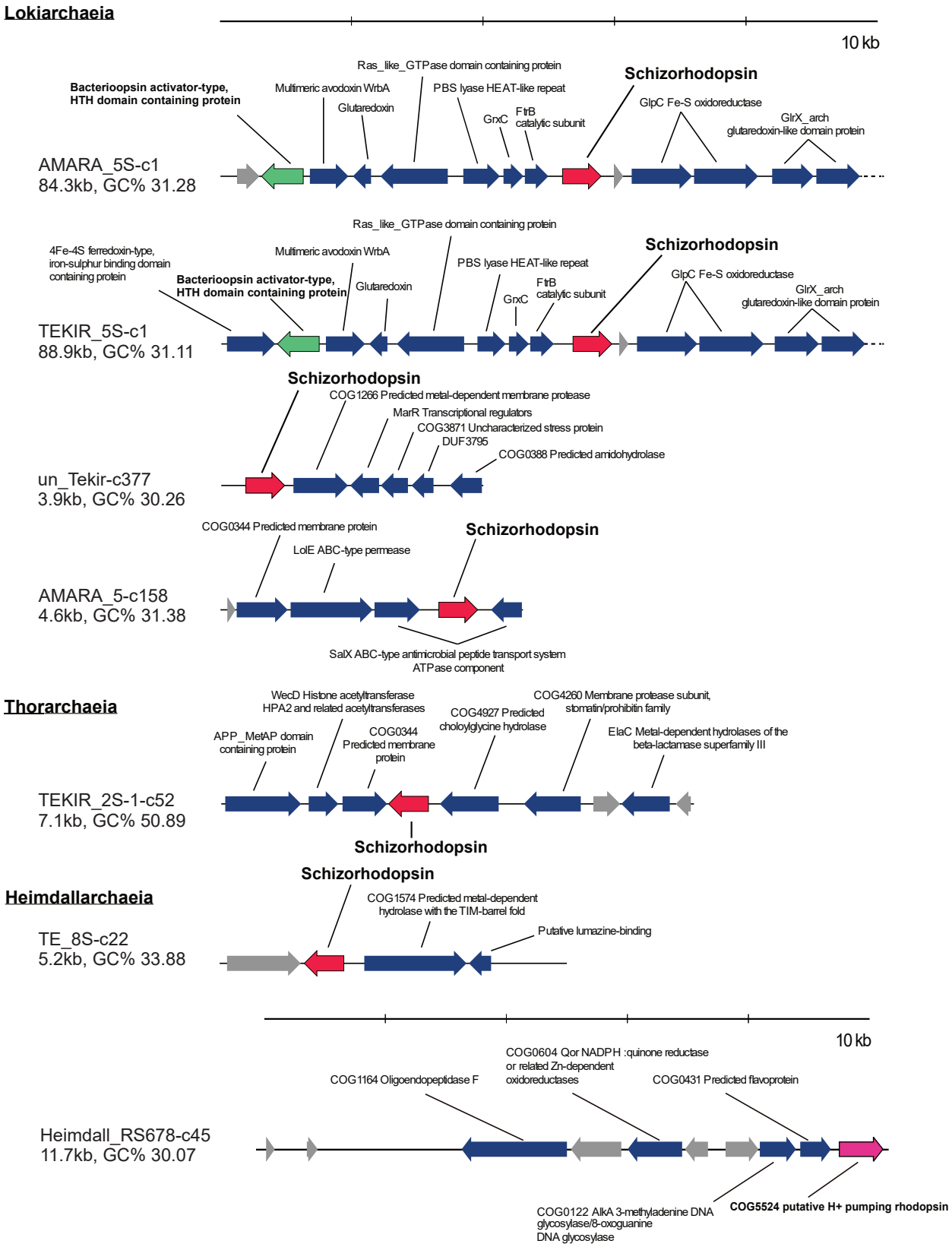




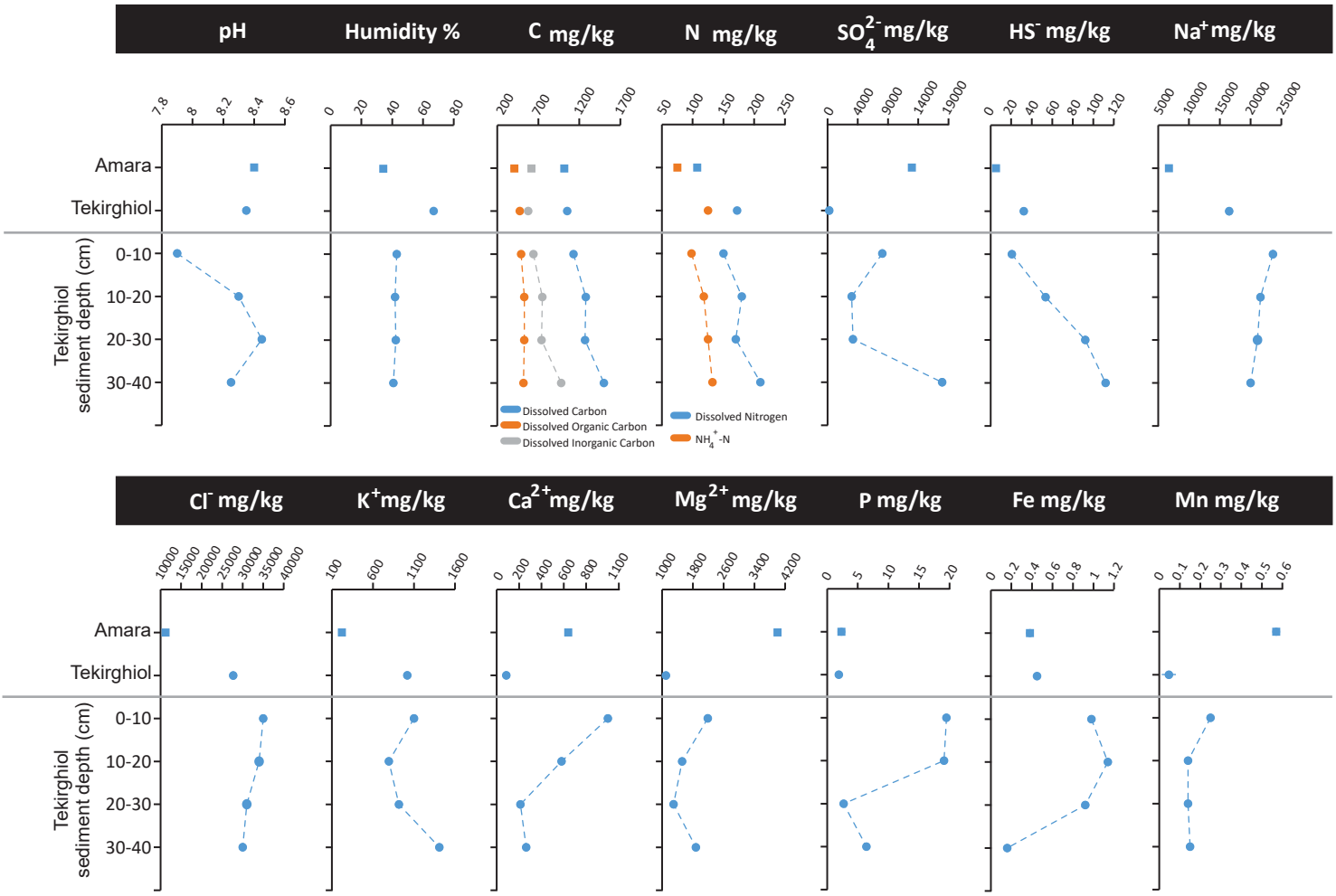
Supplementary Figure 4. **Particular sugar kinases encoded in Heimdallarchaeia**. **a**, ML phylogenetic tree of the ADP-dependent kinases family in Archaea and Eukaryota. The branches belonging to Heimdallarchaeia are colored in blue. The colored panels highlight clades with: glucokinase activity-green, phosphofructokinase activity -orange, and both gluco- and phosphofructokinase activity-purple. Sequences retrieved from the 3 Heimdallarchaeia MAGs (AMARA_4, Heimdall_AB_125, Heimdall_LC_3) were used with other sugar kinase sequences (n=49) listed in a previous study (Castro-Fernandez et al., 2017[29], in Supplementary Table S5). **b**, ML-based phylogenomic subtree showing Heimdall_LC_3 MAG as the oldest within the ones harboring ADP-dependent kinases. The scale bars indicate number of substitutions per site. Standard bootstrap values are indicated by colored circles (legend shown). **c**, Sequence alignment of ADP-dependent kinases highlighting the functionally important position 172, as inferred from Castro-Fernandez et al. (2017)[29]. The similarity-coloring scheme is based on the BLOSUM62 matrix.



Supplementary Figure 5. Maximum likelihood tree of the large subunit of RuBisCO (types I-III) and RuBisCO-like (type IV) (rbcL, K01601) protein sequences (n=146) of bacterial and archaeal taxa. Reference sequences were chosen based on previous trees constructed by Tabita et al. 2007[4] and Wrighton et al. 2016[30]. The branches representing the members of Asgardarchaeota superphylum are colored based on their affiliated phylum (legends on the bottom left). The scale bar indicates the amino acid substitutions per site.



Supplementary Figure 6. Genomic context of rhodopsins in Asgardarchaeota MAGs. Schizorhodopsins are shown in red, bacterioopsin activator-like protein in green, bacteriorhodopsin in purple and the hypothetical proteins in grey. Contigs TE_8S-c22 and un_Tekir-c377 were deposited separately in NCBI GenBank with accession numbers MK463862 and MK463863, respectively. 10 kb scale bars are shown at the top of each category of rhodopsins.



Supplementary Figure 7. Chemical analyses of the Amara and Tekirghiol sediment samples. Mineral concentrations are measured as water-extractable compounds from wet sediments and are reported in mg kg^{-1}. The reported data are the mean values of three sub-samples analysed in parallel from each sediment sample. Relative expanded measurement uncertainty (Ue) was calculated according to the International Organization for Standardization (ISO) - Guide to the Expression of Uncertainty in Measurement (GUM) using a coverage factor (k) of 2 (k=2), equivalent to a confidence of ~ 95%. Ue ranged from 3.6% to 11% depending on the compound as follows: 4.8% for Na^+, 5.1% for K^+, 3.6% for Ca^{2+}, 3.8% for Mg^{2+}, 4.9 % for Fe, 4.4% for Mn, 6.2% for P determined by ICP-OES, 4.7% for sulphate (SO_4^{2-}) determined by ion-chromatography, 7.6% for total sulphides (HS^-) and 6.8% for ammonium (NH_4^+) determined by UV-VIS spectrometry, 11% for Cl^- by titrimetry, 9.0% for total dissolved carbon, 10.4% for dissolved inorganic carbon and 8.7% for dissolved nitrogen determined by TOC/TN Analyser.

Supplementary Table 2. List of Asgardarchaeota genome assemblies downloaded from the NCBI genomes repository (https://www.ncbi.nlm.nih.gov/genome) for this study.

NCBI tax. ID	Organism name	Isolate	GenBank Assembly Accession	Isolation source
2026747	Ca. Heimdallarchaeota	RS678	GCA_002728275.1	Marine water sample
1841596	Ca. Heimdallarchaeota	AB_125	GCA_001940755.1	Marine sediment
1841597	Ca. Heimdallarchaeota	LC_2	GCA_001940725.1	Hydrothermal vent sediment
1841598	Ca. Heimdallarchaeota	LC_3	GCA_001940645.1	Hydrothermal vent sediment
1849166	Ca. Lokiarchaeota	CR_4	GCA_001940655.1	Terrestrial subsurface sediment
1538547	Lokiarchaeum sp.	GC14_75	GCA_000986845.1	Hydrothermal vent sediment
1841599	Ca. Odinarchaeota	LCB_4	GCA_001940665.1	Hot spring
1837170	Ca. Thorarchaeota	AB_25	GCA_001940705.1	Marine sediment
1969372	Ca. Thorarchaeota	MP11T_1	GCA_002825515.1	Mangrove wetland sediments
1969370	Ca. Thorarchaeota	MP8T_1	GCA_002825465.1	
1969371	Ca. Thorarchaeota	MP9T_1	GCA_002825535.1	
1706443	Ca. Thorarchaeota	SMTZ-45	GCA_001563465.1	Sulfate-methane transition zone estuary sediments 16-26 cm
1706444	Ca. Thorarchaeota	SMTZ1-45	GCA_001563335.1	
1706445	Ca. Thorarchaeota	SMTZ1-83	GCA_001563325.1	

Supplementary Table 5. Accession numbers list for INTERPRO (IPR) domains, COGs (Cluster of Orthologous Groups) and UniProtKB protein sequences used to identify potential Eukaryotic Signature Proteins in Asgardarchaeota MAGs. Particularly, the first ESP (1*) was identified by local scanning in MAGs with BLASTP using human sequences downloaded from UNIPROT (https://www.uniprot.org/).

Nr. crt.	Description	IPR domain/COG/UniProtKB
1*	DNA polymerase, ε-like catalytic subunit	Q59EA9, Q9UNF3, Q9Y5S5, Q07864, Q9Y5S4, F5H1D6
2	Topoisomerase IB	arCOG08649,COG3569
3	RNA polymerase, subunit G (rpb8)	IPR031555
4	Ribosomal protein L22e	IPR002671
5	Ribosomal protein L28e/Mak16	IPR029004
6	Tubulins	IPR000217
7	Actin family (divergent)	IPR004000
8	Actin/actin-like conserved site	IPR020902
9	Arp2/3 complex subunit 2/4	IPR008384
10	Gelsolin-domain protein	IPR007122
11	Profilin	IPR036140
12	ESCRT-I:Vps28-like	IPR007143
13	ESCRT-I: steadiness box domain	IPR017916
14	ESCRT-II: EAP30 domain	IPR007286
15	ESCRT-II: Vps25-like	IPR014041 and IPR008570
16	ESCRTIII: Vps2/24/46-like	IPR005024
17	Ubiquitin-domain protein	IPR000626
18	E2-like ubiquitin conjugating protein	IPR000688
19	E3 UFM1-protein ligase 1	IPR018611
20	RAG-type GTPase domain	IPR006762
21	Longin-domain protein	IPR011012
22	Vacuolar fusion protein Mon1	IPR004353
23	RLC7 roadblock domain protein	IPR004942
24	TRAPP-domain protein	IPR007194
25	Zinc finger, Sec23/Sec24-type	IPR006896
26	Arrestin-like proteins (C-terminal)	IPR011021
27	Vesicle coat complex COPII sub. SEC24/SFB2/SFB3	COG5028
28	Folliculin (N-terminal)	IPR037520
29	Ribophorin I	IPR007676
30	Oligosaccharyl transf. OST3/OST6	IPR021149
31	Oligosaccharyl transf. STT32	IPR003674
32	Ezrin/radixin/moesin C-terminal domain	IPR011259
33	active zone protein ELKS	IPR019323

Supplementary Table 7. List of the 48 ribosomal proteins used for phylogenomic analyses.
The number of genomes (out of all 93 used for making trees in Figure 1**b**, and Supplementary Figure 1**b**), in which a particular marker was identified is indicated as “No. of marker hits”. Blue/white cells indicate presence/absence respectively.

				Thorarchaeia				Lokiarchaeia			
arCOG	Class	Description	No. of marker hits	AMARA 2	AMARA 8	TEKIR 12S	TEKIR 14	AMARA 1	AMARA 1S	TEKIR 8	TEKIR 21
arCOG00779	J	Ribosomal protein L15	92								
arCOG00780	J	Ribosomal protein L18E	89								
arCOG00781	J	Ribosomal protein L32E	90								
arCOG00782	J	Ribosomal protein S14	86								
arCOG00785	J	Ribosomal protein L29	90								
arCOG01344	J	Ribosomal protein S19E (S16A)	89								
arCOG01722	J	Ribosomal protein S13	92								
arCOG01758	J	Ribosomal protein S10	89								
arCOG01885	J	Ribosomal protein S17E	89								
arCOG01946	J	Ribosomal protein S6E/S10	91								
arCOG04067	J	Ribosomal protein L2	91								
arCOG04070	J	Ribosomal protein L3	91								
arCOG04071	J	Ribosomal protein L4	90								
arCOG04072	J	Ribosomal protein L23	89								
arCOG04086	J	Ribosomal protein L30	92								
arCOG04087	J	Ribosomal protein S5	92								
arCOG04089	J	Ribosomal protein L19E	90								
arCOG04090	J	Ribosomal protein L6P	89								
arCOG04091	J	Ribosomal protein S8	89								
arCOG04092	J	Ribosomal protein L5	91								
arCOG04093	J	Ribosomal protein S4E	91								
arCOG04094	J	Ribosomal protein L24	91								
arCOG04095	J	Ribosomal protein L14	92								
arCOG04096	J	Ribosomal protein S17	92								
arCOG04097	J	Ribosomal protein S3	93								
arCOG04098	J	Ribosomal protein L22	90								
arCOG04099	J	Ribosomal protein S19	91								
arCOG04108	J	Ribosomal protein S27E	89								
arCOG04109	J	Ribosomal protein L44E	89								
arCOG04113	J	Ribosomal protein L10AE/L16	89								
arCOG04129	J	Ribosomal protein L21E	92								
arCOG04182	J	Ribosomal protein S24E	90								
arCOG04185	J	Ribosomal protein S15P	91								
arCOG04186	J	Ribosomal protein S3AE	89								
arCOG04208	J	Ribosomal protein L37AE/L43A	91								
arCOG04209	J	Ribosomal protein L15E	92								
arCOG04239	J	Ribosomal protein S4 or related protein	92								
arCOG04240	J	Ribosomal protein S11	91								
arCOG04242	J	Ribosomal protein L13	91								
arCOG04243	J	Ribosomal protein S9	90								
arCOG04245	J	Ribosomal protein S2	91								
arCOG04254	J	Ribosomal protein S7	88								
arCOG04255	J	Ribosomal protein S12	87								
arCOG04288	J	Ribosomal protein L10	89								
arCOG04289	J	Ribosomal protein L1	89								
arCOG04314	J	Ribosomal protein S28E/S33	89								
arCOG04372	J	Ribosomal protein L11	91								
arCOG04473	J	Ribosomal protein L31E	90								

Supplementary Table 8. Protein RefSeq accession numbers for amino acid sequences used in phylogenetic inferences for key components of the Tryptophan degradation pathway (kynurenine pathway) in Heimdallarchaeia.

Organism name	NCBI TaxID	Domain	tryptophan 2,3-dioxygenase (TDO)	kynurenine 3-monooxygenase (KMO)	3-hydroxyanthranilate 3,4-dioxygenase (HAAO)
Actinomadura meyerae	240840	Bacteria	WP_089326016.1	WP_089325941.1	WP_089327247.1
Kribbella flavida	479435	Bacteria	WP_012919326.1	WP_012923059.1	WP_012923056.1
Sediminitomix flava	379075	Bacteria	WP_109616529.1	WP_109618774.1	WP_109618766.1
Taibaiella soli	1649169	Bacteria	WP_111000080.1	WP_110999730.1	WP_110997529.1
Belliella buryatensis	1500549	Bacteria	WP_089238736.1	WP_089240074.1	WP_089240078.1
Aquiflexum balticum	758820	Bacteria	WP_084121451.1	WP_084119369.1	WP_084119395.1
Anditalea andensis	1048983	Bacteria	WP_035073830.1	WP_035075278.1	WP_035075279.1
Cesiribacter andamanensis	1279009	Bacteria	WP_009196574.1	WP_009197406.1	WP_009197408.1
Ohtaekwangia koreensis	688867	Bacteria	WP_079686329.1	WP_079684767.1	WP_079684770.1
Flavobacterium johnsoniae	376686	Bacteria	WP_012022820.1	WP_012022586.1	WP_012023303.1
Kangiella sediminilitoris	1144748	Bacteria	WP_068990479.1	WP_068993461.1	WP_068993449.1
Methylocaldum marinum	1432792	Bacteria	BBA35597.1	BBA35436.1	BBA35441.1
Microbulbifer donghaiensis	494016	Bacteria	WP_073275402.1	WP_073275006.1	WP_073275000.1
Pseudobacteriovorax antillogorgiicola	1513793	Bacteria	SMF46342.1	SMF59929.1	SMF59911.1
Bradymonas sediminis	1548548	Bacteria	WP_111337759.1	WP_111332768.1	WP_111335634.1
Cavenderia fasciculata	1054147	Eukaryota	XP_004359190.1	XP_004357271.1	XP_004351239.1
Crassostrea gigas	29159	Eukaryota	EKC35799.1	XP_011452758.1	XP_011415145.1
Cutaneotrichosporon oleaginosum	879819	Eukaryota	XP_018274934.1	XP_018275936.1	XP_018276298.1
Caenorhabditis elegans	6239	Eukaryota	NP_498284.1	NP_506024.3	NP_505450.1
Strongyloides ratti	34506	Eukaryota	XP_024506610.1	XP_024508004.1	XP_024508314.1
Galdieria sulphuraria	130081	Eukaryota	XP_005705420.1	XP_005707814.1	XP_005706413.1
Danio rerio	7955	Eukaryota	NP_001096086.1	NP_001314753.1	NP_001007391.1
Scleropages formosus	113540	Eukaryota	XP_018591176.1	XP_018593530.1	XP_018621588.1
Manacus vitellinus	328815	Eukaryota	XP_017926887.1	XP_008918920.2	XP_008931159.1
Zonotrichia albicollis	44394	Eukaryota	XP_005486392.1	XP_005486991.1	XP_005483133.1
Taeniopygia guttata	59729	Eukaryota	XP_002198679.1	XP_002192779.1	XP_002193125.1
Ciona intestinalis	413601	Eukaryota	XP_002128288.1	XP_002131315.1	XP_002119902.1

Supplementary Table 9. Major water leachable ions and elements from sediments collected in Tekirghiol and Amara Lakes during October 2017. The reported data are the mean values of three sub-samples analysed in parallel from each sediment sample. Relative expanded measurement uncertainties are indicated in the caption of Supplementary Figure 7.

Ions/metals ($mg \cdot Kg^{-1}$)	**Tekirghiol Lake**	**Amara Lake**
pH of pore water	7.38	8.03
Na^{+}	16520	7000
K^{+}	1020	221
Ca^{2+}	86.4	640
Mg^{2+}	1100	4000
Fe*	0.45	0.38
Mn*	<0.1	0.57
Cl^{-}	27700	11200
SO_4^{2-}	250	13240

Note: Values are determined for the water-extracted fraction of the wet sediment. The asterisk (*) indicates element total concentration.

Supplementary References

1. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–7196 (2007).
2. Liu, Y. *et al.* Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. *ISME J.* **12,** 1021–1031 (2018).
3. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541,** 353–358 (2017).
4. Tabita, F. R. *et al.* Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol. Mol. Biol. Rev.* **71,** 576–599 (2007).
5. Da Cunha, V., Gaia, M., Nasir, A. & Forterre, P. Asgard archaea do not close the debate about the universal tree of life topology. *PLOS Genet.* **14,** e1007215 (2018).
6. Pushkarev, A. *et al.* A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **558,** 595–599 (2018).
7. Shand, R. F. & Betlach, M. C. bop gene cluster expression in bacteriorhodopsin-overproducing mutants of Halobacterium halobium. *J. Bacteriol.* **176,** 1655–1660 (1994).
8. Seitz, K. W., Lazar, C. S., Hinrichs, K.-U., Teske, A. P. & Baker, B. J. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *ISME J.* **10,** 1696–1705 (2016).
9. Ilikchyan, I. N., McKay, R. M. L., Zehr, J. P., Dyhrman, S. T. & Bullerjahn, G. S. Detection and expression of the phosphonate transporter gene phnD in marine and freshwater picocyanobacteria. *Environ. Microbiol.* **11,** 1314–1324 (2009).
10. Trautmann, A., Watzer, B., Wilde, A., Forchhammer, K. & Posten, C. Effect of phosphate availability on cyanophycin accumulation in Synechocystis sp. PCC 6803 and the production strain BW86. *Algal Res.* **20,** 189–196 (2016).
11. Orita, I. *et al.* The Ribulose Monophosphate Pathway Substitutes for the Missing Pentose Phosphate Pathway in the Archaeon Thermococcus kodakaraensis. *J. Bacteriol.* **188,** 4698–4704 (2006).
12. Brasen, C., Esser, D., Rauch, B. & Siebers, B. Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiol. Mol. Biol. Rev.* **78,** 89–175 (2014).
13. Uwe, S. & J., E. B. The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol. Rev.* **29,** 765–794 (2006).
14. Tittmann, K., Golbik, R., Ghisla, S. & Hübner, G. Mechanism of Elementary Catalytic Steps of Pyruvate Oxidase from Lactobacillus plantarum. *Biochemistry* **39,** 10747–10754 (2000).
15. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,** 845–858 (2015).
16. Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N. & Martin, W. F. Lokiarchaeon is hydrogen dependent. *Nat. Microbiol.* **1,** 16034 (2016).
17. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25,** 1043–1055 (2015).

18. Berg, I. A. Ecological Aspects of the Distribution of Different Autotrophic CO(2) Fixation Pathways . *Appl. Environ. Microbiol.* **77,** 1925–1936 (2011).

19. Wood, H. G. Life with CO or CO2 and H2 as a source of carbon and energy. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **5,** 156–163 (1991).

20. Techtmann, S. M. *et al.* Regulation of Multiple Carbon Monoxide Consumption Pathways in Anaerobic Bacteria. *Front. Microbiol.* **2,** 147 (2011).

21. Ragsdale, S. W. & Pierce, E. Acetogenesis and the Wood-Ljungdahl pathway of CO(2) fixation. *Biochim. Biophys. Acta* **1784,** 1873–1898 (2008).

22. King, G. M. Uptake of carbon monoxide and hydrogen at environmentally relevant concentrations by mycobacteria. *Appl. Environ. Microbiol.* **69,** 7266–7272 (2003).

23. King, G. M. Nitrate-dependent anaerobic carbon monoxide oxidation by aerobic CO-oxidizing bacteria. *FEMS Microbiol. Ecol.* **56,** 1–7 (2006).

24. King, G. M. & Weber, C. F. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* **5,** 107–118 (2007).

25. Adam, P. S., Borrel, G. & Gribaldo, S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc. Natl. Acad. Sci.* **115,** E1166–E1173 (2018).

26. Ternes, C. M. & Schönknecht, G. Gene Transfers Shaped the Evolution of De Novo NAD(+) Biosynthesis in Eukaryotes. *Genome Biol. Evol.* **6,** 2335–2349 (2014).

27. Nicholls, P., Marshall, D. C., Cooper, C. E. & Wilson, M. T. Sulfide inhibition of and metabolism by cytochrome c oxidase. *Biochem. Soc. Trans.* **41,** 1312–1316 (2013).

28. Jarrell, K. F. & Albers, S.-V. The archaellum: an old motility structure with a new name. *Trends Microbiol.* **20,** 307–312 (2012).

29. Castro-Fernandez, V. *et al.* Reconstructed ancestral enzymes reveal that negative selection drove the evolution of substrate specificity in ADP-dependent kinases. *J. Biol. Chem.* **292,** 21218 (2017).

30. Wrighton, K. C. *et al.* RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. *ISME J.* **10,** 2702–2714 (2016).